BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Eight Number Four November 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 Four November 2022 ISSN: 1350-7265

CONTENTS

(continued)

BERNOULLI Volume 28 Number 4 November 2022 Pages 2123–3078 ISI/BS

Volume 28 Number 4   November 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**(4), 2022, 2123–2150
https://doi.org/10.3150/21-BEJ1406

The complex behaviour of Galton rank-order statistic

EUSTASIO DEL BARRIO[1,a], JUAN A. CUESTA-ALBERTOS[2,c] and CARLOS MATRAN[1,b]

[1]*Department of Statistics and Operational Research and IMUVA, University of Valladolid, Valladolid, Spain.* [a]*eustasio.delbarrio@uva.es,* [b]*carlos.matran@uva.es*
[2]*Departmento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Santander, Spain.* [c]*cuestaj@unican.es*

Galton's rank-order statistic is one of the oldest statistical tools for two-sample comparisons. It is also a very natural index to measure departures from stochastic dominance. Yet, its asymptotic behaviour has been investigated only partially, under restrictive assumptions. This work provides a comprehensive study of this behaviour, based on the analysis of the so-called contact set (a modification of the set in which the quantile functions coincide). We show that a.s. convergence to the population counterpart holds if and only if the contact set has zero Lebesgue measure. When this set is finite we show that the asymptotic behaviour is determined by the local behaviour of a suitable reparameterization of the quantile functions in a neighbourhood of the contact points. Regular crossings result in standard rates and Gaussian limiting distributions, but higher order contacts (in the sense introduced in this work) or contacts at the extremes of the supports may result in different rates and non-Gaussian limits.

Keywords: Relaxed stochastic dominance; asymptotics; consistency; Galton rank-order statistic; comparison of quantile functions; contact points; crossings; tangencies; contact intensity

References

[1] Álvarez-Esteban, P.C., del Barrio, E., Cuesta-Albertos, J.A. and Matrán, C. (2017). Models for the assessment of treatment improvement: The ideal and the feasible. *Statist. Sci.* **32** 469–485. MR3696006 https://doi.org/10.1214/17-STS616

[2] Aly, E.-E.A.A. (1986). Strong approximations of the Q-Q process. *J. Multivariate Anal.* **20** 114–128. MR0862244 https://doi.org/10.1016/0047-259X(86)90022-9

[3] Aly, E.-E.A.A., Csörgő, M. and Horváth, L. (1987). P-P plots, rank processes, and Chernoff-Savage theorems. In *New Perspectives in Theoretical and Applied Statistics (Bilbao, 1986). Wiley Ser. Probab. Math. Statist. Probab. Math. Statist.* 135–156. New York: Wiley. MR0900214

[4] Andersen, E.S. (1953). On the fluctuations of sums of random variables. *Math. Scand.* **1** 263–285. MR0058893 https://doi.org/10.7146/math.scand.a-10385

[5] Behnen, K. and Neuhaus, G. (1983). Galton's test as a linear rank test with estimated scores and its local asymptotic efficiency. *Ann. Statist.* **11** 588–599. MR0696070

[6] Bickel, P.J. and Hodges, J.L. Jr. (1967). The asymptotic theory of Galton's test and a related simple estimate of location. *Ann. Math. Stat.* **38** 73–89. MR0231474 https://doi.org/10.1214/aoms/1177699059

[7] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[8] Bobkov, S. and Ledoux, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** v+126. MR4028181 https://doi.org/10.1090/memo/1259

[9] Chung, K.L. and Feller, W. (1949). On fluctuations in coin-tossing. *Proc. Natl. Acad. Sci. USA* **35** 605–608. MR0033459 https://doi.org/10.1073/pnas.35.10.605

[10] Csáki, E. and Vincze, I. (1961). On some problems connected with the Galton-test. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **6** 97–109. MR0145608

1350-7265

[11] Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Chichester: Wiley. MR1215046

[12] Darwin, C. (1876). *The Effect of Cross- and Self-Fertilization in the Vegetable Kingdom*. London: Murray.

[13] del Barrio, E., Cuesta-Albertos, J.A. Matran, C. (2022). Supplement to "The complex behaviour of Galton rank-order statistic." https://doi.org/10.3150/21-BEJ1406SUPP

[14] Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2** 267–277. MR0356350

[15] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, 3rd ed. New York: Wiley. MR0228020

[16] Gross, S. and Holland, P.W. (1968). The distribution of Galton's statistic. *Ann. Math. Stat.* **39** 2114–2117. MR0232474 https://doi.org/10.1214/aoms/1177698040

[17] Hodges, J.L. Jr. (1955). Galton's rank-order test. *Biometrika* **42** 261–262. MR0068789 https://doi.org/10.2307/2333442

[18] Lehmann, E.L. (1955). Ordered families of distributions. *Ann. Math. Stat.* **26** 399–419. MR0071684 https://doi.org/10.1214/aoms/1177728487

[19] Lévy, P. (1939). Sur certains processus stochastiques homogènes. *Compos. Math.* **7** 283–339. MR0000919

[20] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. *Springer Series in Statistics*. New York: Springer. MR0988164 https://doi.org/10.1007/978-1-4613-9620-8

[21] 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

[22] Zhuang, W.W., Hu, B.Y. and Chen, J. (2019). Semiparametric inference for the dominance index under the density ratio model. *Biometrika* **106** 229–241. MR3912393 https://doi.org/10.1093/biomet/asy068

Bernoulli **28**(4), 2022, 2151–2180
https://doi.org/10.3150/21-BEJ1413

Nonparametric Bayesian volatility estimation for gamma-driven stochastic differential equations

DENIS BELOMESTNY[1,2,a], SHOTA GUGUSHVILI[3,b], MORITZ SCHAUER[4,c] and PETER SPREIJ[5,6,d]

[1] *Faculty of Mathematics, Duisburg-Essen University, Thea-Leymann-Str. 9, D-45127 Essen, Germany*
[2] *Faculty of Computer Sciences, HSE University.* [a] *denis.belomestny@uni-due.de*
[3] *Wageningen University & Research, Postbus 16, 6700 AA Wageningen, The Netherlands.*
[b] *shota@yesdatasolutions.com*
[4] *Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Göteborg, Sweden.* [c] *smoritz@chalmers.se*
[5] *Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands*
[6] *Institute for Mathematics, Astrophysics and Particle Physics, Radboud University, Nijmegen, Netherlands.*
[d] *spreij@uva.nl*

We study a nonparametric Bayesian approach to estimation of the volatility function of a stochastic differential equation driven by a gamma process. The volatility function is modelled a priori as piecewise constant, and we specify a gamma prior on its values. This leads to a straightforward procedure for posterior inference via an MCMC procedure. We give theoretical performance guarantees (minimax optimal contraction rates for the posterior) for the Bayesian estimate in terms of the regularity of the unknown volatility function. We illustrate the method on synthetic and real data examples.

Keywords: Gamma process; nonparametric Bayesian estimation; stochastic differential equation

References

Batir, N. (2008). Inequalities for the gamma function. *Arch. Math. (Basel)* **91** 554–563.

Batz, P., Ruttor, A. and Opper, M. (2018). Approximate Bayes learning of stochastic differential equations. *Phys. Rev. E* **98** 022109.

Belomestny, D., Gugushvili, S., Schauer, M. and Spreij, P. (2019). Nonparametric Bayesian inference for Gamma-type Lévy subordinators. *Commun. Math. Sci.* **17** 781–816.

Belomestny, D., Gugushvili, S., Schauer, M. and Spreij, P. (2021). Weak solutions to gamma-driven stochastic differential equations. Available at arXiv:2108.11891.

Carson, J., Crucifix, M., Preston, S.P. and Wilkinson, R.D. (2019). Quantifying age and model uncertainties in palaeoclimate data and dynamical climate models with a joint inferential analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **475** 20180854.

Chance, D.M., Hillebrand, E. and Hilliard, J.E. (2008). Pricing an option on revenue from an innovation: An application to movie box office revenue. *Manage. Sci.* **54** 1015–1028.

Ditlevsen, P.D. (1999). Observation of α-stable noise induced millennial climate changes from an ice-core record. *Geophys. Res. Lett.* **26** 1441–1444.

Dufresne, F., Gerber, H.U. and Shiu, E.S.W. (1991). Risk theory with the gamma process. *Astin Bull.* **21** 177–192.

Eguchi, S. and Uehara, Y. (2020). Schwartz type model selection for ergodic stochastic differential equation models. Available at arXiv:1904.12398.

Ghosal, S., Ghosh, J.K. and Van Der Vaart, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* 500–531.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge: Cambridge Univ. Press.

Gugushvili, S., Mariucci, E. and van der Meulen, F. (2020). Decompounding discrete distributions: A nonparametric Bayesian approach. *Scand. J. Stat.* **47** 464–492.

Gugushvili, S., van der Meulen, F. and Spreij, P. (2015). Nonparametric Bayesian inference for multidimensional compound Poisson processes. *Mod. Stoch. Theory Appl.* **2** 1–15.

Gugushvili, S., van der Meulen, F. and Spreij, P. (2018). A non-parametric Bayesian approach to decompounding from high frequency data. *Stat. Inference Stoch. Process.* **21** 53–79.

Gugushvili, S., van der Meulen, F., Schauer, M. and Spreij, P. (2019). Nonparametric Bayesian volatility estimation. In *2017 MATRIX Annals* (J. de Gier, C.E. Praeger and T. Tao, eds.) 279–302. Cham: Springer.

Gushchin, A., Pavlyukevich, I. and Ritsch, M. (2019). Drift estimation for a Lévy-driven Ornstein-Uhlenbeck process with heavy tails. Available at arXiv:1911.11202.

Jasra, A., Kamatani, K. and Masuda, H. (2019). Bayesian inference for stable Lévy–driven stochastic differential equations with high-frequency data. *Scand. J. Stat.* **46** 545–574.

Jasra, A., Stephens, D.A., Doucet, A. and Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Stat.* **38** 1–22.

Kabanov, Y.M., Liptser, R.S. and Shiryaev, A. (1986). On the variation distance for probability measures defined on a filtered space. *Probab. Theory Related Fields* **71** 19–35.

Koskela, J., Spanò, D. and Jenkins, P.A. (2019). Consistency of Bayesian nonparametric inference for discretely observed jump diffusions. *Bernoulli* **25** 2183–2205.

Kyprianou, A.E. (2014). *Fluctuations of Lévy Processes with Applications. Introductory Lectures*, 2nd ed. *Universitext.* Heidelberg: Springer.

North Greenland Ice Core Project Members (2007). 50 year means of oxygen isotope data from ice core NGRIP. Supplement to: North Greenland Ice Core Project Members (2004): High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431** 147–151. https://doi.org/10.1038/nature02805

Müller, P. and Mitra, R. (2013). Bayesian nonparametric inference—why and how. *Bayesian Anal.* **8** 269–302.

Müller, P., Quintana, F.A., Jara, A. and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. *Springer Series in Statistics*. Cham: Springer.

Nickl, R. and Söhl, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.* **45** 1664–1693.

Protter, P.E. (2004). *Stochastic Integration and Differential Equations*, 2nd ed. *Applications of Mathematics (New York)* **21**. Berlin: Springer. Stochastic Modelling and Applied Probability.

Reed, W.J. and McKelvey, K.S. (2002). Power-law behaviour and parametric models for the size-distribution of forest fires. *Ecol. Model.* **150** 239–254.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability.* London: CRC Press.

Todorov, V. (2011). Econometric analysis of jump-driven stochastic volatility models. *J. Econometrics* **160** 12–21.

Tsybakov, A.B. (2008). *Introduction to Nonparametric Estimation*. Berlin: Springer.

Uehara, Y. (2019b). Statistical inference for misspecified ergodic Lévy driven stochastic differential equation models. *Stochastic Process. Appl.* **129** 4051–4081. MR3997671 https://doi.org/10.1016/j.spa.2018.11.007

van Noortwijk, J. (2009). A survey of the application of gamma processes in maintenance. *Reliab. Eng. Syst. Saf.* **94** 2–21.

Wenocur, M.L. (1989b). A reliability model based on the gamma process and its analytic theory. *Adv. in Appl. Probab.* **21** 899–918.

Bernoulli **28**(4), 2022, 2181–2208
https://doi.org/10.3150/21-BEJ1415

A Riemann–Stein kernel method

ALESSANDRO BARP[1,2,a], CHRIS. J. OATES[3,2,c], EMILIO PORCU[4,5,d] and MARK GIROLAMI[1,2,b]

[1]*University of Cambridge, Cambridge, UK.* [a]*ab2286@eng.cam.ac.uk,* [b]*mag92@eng.cam.ac.uk*
[2]*Alan Turing Institute, London, UK*
[3]*Newcastle University, Newcastle-upon-Tyne, UK.* [c]*chris.oates@ncl.ac.uk*
[4]*Khalifa University, Abu Dhabi, United Arab Emirates.* [d]*emilio.porcu@ku.ac.ae*
[5]*Trinity College Dublin, Dublin, IE*

This paper proposes and studies a numerical method for approximation of posterior expectations based on interpolation with a Stein reproducing kernel. Finite-sample-size bounds on the approximation error are established for posterior distributions supported on a compact Riemannian manifold, and we relate these to a *kernel Stein discrepancy* (KSD). Moreover, we prove in our setting that the KSD is equivalent to *Sobolev discrepancy* and, in doing so, we completely characterise the convergence-determining properties of KSD. Our contribution is rooted in a novel combination of Stein's method, the theory of reproducing kernels, and existence and regularity results for partial differential equations on a Riemannian manifold.

Keywords: Bayesian computation; kernel Stein discrepancy; reproducing kernel; Stein equation

References

[1] Anastasiou, A., Barp, A., Briol, F.-X., Ebner, B., Gaunt, R.E., Ghaderinezhad, F., Gorham, J., Gretton, A., Ley, C., Liu, Q., Mackey, L., Oates, C.J., Reinert, G. and Swan, Y. (2021). Stein's method meets statistics: A review of some recent developments. Available at arXiv:2105.03481.
[2] Arbel, M., Korba, A., Salim, A. and Gretton, A. (2019). Maximum mean discrepancy gradient flow. In *Advances in Neural Information Processing Systems* 6484–6494.
[3] Arnaudon, A., Barp, A. and Takao, S. (2019). Irreversible Langevin MCMC on Lie Groups. In *International Conference on Geometric Science of Information* 171–179. Springer.
[4] Assaraf, R. and Caffarel, M. (1999). Zero-variance principle for Monte Carlo algorithms. *Phys. Rev. Lett.* **83** 4682.
[5] Bachman, D. (2006). *A Geometric Approach to Differential Forms*. Boston, MA: Birkhäuser, Inc. MR2248379
[6] Baez, J. and Muniain, J.P. (1994). *Gauge Fields, Knots and Gravity*. *Series on Knots and Everything* **4**. River Edge, NJ: World Scientific Co., Inc. MR1313910 https://doi.org/10.1142/2324
[7] Barp, A., Kennedy, A. and Girolami, M. (2019). Hamiltonian Monte Carlo on symmetric and homogeneous spaces via symplectic reduction. Available at arXiv:1903.02699.
[8] Belomestny, D., Iosipoi, L., Moulines, E., Naumov, A. and Samsonov, S. (2020). Variance reduction for Markov chains with application to MCMC. *Stat. Comput.* **30** 973–997. MR4108687 https://doi.org/10.1007/s11222-020-09931-z
[9] Belomestny, D., Iosipoi, L. and Zhivotovskiy, N. (2018). Variance reduction in Monte Carlo estimators via empirical variance minimization. *Dokl. Math.* **98** 494–497.
[10] Berlinet, A. and Thomas-Agnan, C. (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Berlin: Springer.
[11] Bevilacqua, M., Faouzi, T., Furrer, R. and Porcu, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *Ann. Statist.* **47** 828–856. MR3909952 https://doi.org/10.1214/17-AOS1652
[12] Bingham, N.H. (1973). Positive definite functions on spheres. *Proc. Camb. Philos. Soc.* **73** 145–156. MR0339308 https://doi.org/10.1017/s0305004100047551

[13] Brandolini, L., Choirat, C., Colzani, L., Gigante, G., Seri, R. and Travaglini, G. (2014). Quadrature rules and distribution of points on manifolds. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **13** 889–923. MR3362114

[14] Brauchart, J.S. and Dick, J. (2013). A characterization of Sobolev spaces on the sphere and an extension of Stolarsky's invariance principle to arbitrary smoothness. *Constr. Approx.* **38** 397–445. MR3122277 https://doi.org/10.1007/s00365-013-9217-z

[15] Brosse, N., Durmus, A., Meyn, S., Moulines, É. and Radhakrishnan, A. (2018). Diffusion approximations and control variates for MCMC. Available at arXiv:1808.01665.

[16] Byrne, S. and Girolami, M. (2013). Geodesic Monte Carlo on embedded manifolds. *Scand. J. Stat.* **40** 825–845. MR3145120 https://doi.org/10.1111/sjos.12036

[17] Chen, W., Barp, A., Briol, F., Gorham, J., Girolami, M., Mackey, L. and Oates, C. (2019). Stein point Markov chain Monte Carlo. In *Proceedings of the 36th International Conference on Machine Learning*.

[18] Chen, W., Mackey, L., Gorham, J., Briol, F. and Oates, C. (2018). Stein points. In *Proceedings of the 35th International Conference on Machine Learning*.

[19] Chwialkowski, K., Strathmann, H. and Gretton, A. (2016). A kernel test of goodness of fit. In *Proceedings of the 33rd International Conference on Machine Learning*.

[20] Dai, F. and Xu, Y. (2013). *Approximation Theory and Harmonic Analysis on Spheres and Balls*. *Springer Monographs in Mathematics*. New York: Springer. MR3060033 https://doi.org/10.1007/978-1-4614-6660-4

[21] Daley, D.J. and Porcu, E. (2014). Dimension walks and Schoenberg spectral measures. *Proc. Amer. Math. Soc.* **142** 1813–1824. MR3168486 https://doi.org/10.1090/S0002-9939-2014-11894-6

[22] De Vito, E., Mücke, N. and Rosasco, L. (2021). Reproducing kernel Hilbert spaces on manifolds: Sobolev and diffusion spaces. *Anal. Appl. (Singap.)* **19** 363–396. MR4233705 https://doi.org/10.1142/S0219530520400114

[23] Diaconis, P., Holmes, S. and Shahshahani, M. (2013). Sampling from a manifold. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*. *Inst. Math. Stat. (IMS) Collect.* **10** 102–125. Beachwood, OH: IMS. MR3586941

[24] Dick, J., Gantner, R.N., Gia, Q.T.L. and Schwab, C. (2016). Higher order quasi-Monte Carlo integration for Bayesian estimation. Available at arXiv:1602.07363.

[25] Dick, J., Rudolf, D. and Zhu, H. (2019). A weighted discrepancy bound of quasi-Monte Carlo importance sampling. *Statist. Probab. Lett.* **149** 100–106. MR3915359 https://doi.org/10.1016/j.spl.2019.01.014

[26] Ehler, M., Gräf, M. and Oates, C.J. (2019). Optimal Monte Carlo integration on closed manifolds. *Stat. Comput.* **29** 1203–1214. MR4026666 https://doi.org/10.1007/s11222-019-09894-w

[27] Fasshauer, G.E. and Ye, Q. (2011). Reproducing kernels of generalized Sobolev spaces via a Green function approach with distributional operators. *Numer. Math.* **119** 585–611. MR2845629 https://doi.org/10.1007/s00211-011-0391-2

[28] Federer, H. (1969). *Geometric Measure Theory*. *Die Grundlehren der Mathematischen Wissenschaften, Band 153*. New York: Springer New York Inc. MR0257325

[29] Fisher, M., Nolan, T., Graham, M., Prangle, D. and Oates, C. (2021). Measure transport with kernel Stein discrepancy. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*.

[30] Frankel, T. (2011). *The Geometry of Physics: An Introduction*, Cambridge: Cambridge Univ. Press.

[31] Fuselier, E. and Wright, G.B. (2012). Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. *SIAM J. Numer. Anal.* **50** 1753–1776. MR2970764 https://doi.org/10.1137/110821846

[32] Gallot, S., Hulin, D. and Lafontaine, J. (1990). *Riemannian Geometry*, 2nd ed. *Universitext*. Berlin: Springer. MR1083149 https://doi.org/10.1007/978-3-642-97242-3

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

[34] Gilks, W.R., Richardson, S. and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in Practice*. Boca Raton: CRC Press.

[35] Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475 https://doi.org/10.1198/016214502760047113

[36] Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554 https://doi.org/10.3150/12-BEJSP06

[37] Gorham, J. and Mackey, L. (2017). Measuring sample quality with kernels. In *Proceedings of the 34th International Conference on Machine Learning*.

[38] Grigor'yan, A. (2006). Heat kernels on weighted manifolds and applications. *Contemp. Math.* **398** 93–191.

[39] Grosse, N. and Nistor, V. (2017). Neumann and mixed problems on manifolds with boundary and bounded geometry. Available at arXiv:1703.07228.

[40] Große, N. and Schneider, C. (2013). Sobolev spaces on Riemannian manifolds with bounded geometry: General coordinates and traces. *Math. Nachr.* **286** 1586–1613. MR3126616 https://doi.org/10.1002/mana.201300007

[41] Han, J. and Liu, Q. (2018). Stein variational gradient descent without gradient. In *Proceedings of the 35th International Conference on Machine Learning*.

[42] Hebey, E. (2000). *Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities*. Providence: Amer. Math. Soc.

[43] Hodgkinson, L., Salomone, R. and Roosta, F. (2020). The reproducing Stein kernel approach for post-hoc corrected sampling. Available at arXiv:2001.09266.

[44] Holbrook, A., Lan, S., Vandenberg-Rodes, A. and Shahbaba, B. (2018). Geodesic Lagrangian Monte Carlo over the space of positive definite matrices: With application to Bayesian spectral density estimation. *J. Stat. Comput. Simul.* **88** 982–1002. MR3750634 https://doi.org/10.1080/00949655.2017.1416470

[45] Joseph, V.R., Wang, D., Gu, L., Lyu, S. and Tuo, R. (2019). Deterministic sampling of expensive posteriors using minimum energy designs. *Technometrics* **61** 297–308. MR3987621 https://doi.org/10.1080/00401706.2018.1552203

[46] Kingma, D. and Welling, M. (2014). Auto-encoding variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations*.

[47] Lam, H. and Zhang, H. (2019). On the stability of kernelized control functionals on partial and biased stochastic inputs. In *2019 Winter Simulation Conference (WSC)* 344–355. IEEE.

[48] Lan, S., Zhou, B. and Shahbaba, B. (2014). Spherical Hamiltonian Monte Carlo for constrained target distributions. In *Proceedings of the 31st International Conference on Machine Learning* 629–637.

[49] Le, H., Lewis, A., Bharath, K. and Fallaize, C. (2020). A diffusion approach to Stein's method on Riemannian manifolds. Available at arXiv:2003.11497.

[50] Lee, J.M. (2013). *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218**. New York: Springer. MR2954043

[51] Ley, C., Reinert, G. and Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probab. Surv.* **14** 1–52. MR3595350 https://doi.org/10.1214/16-PS278

[52] Lin, L., Niu, M., Cheung, P. and Dunson, D. (2019). Extrinsic Gaussian processes for regression and classification on manifolds. *Bayesian Anal.* **14** 887–906.

[53] Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

[54] Liu, C. and Zhu, J. (2018). Riemannian Stein variational gradient descent for Bayesian inference. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*.

[55] Liu, H., Feng, Y., Mao, Y., Zhou, D., Peng, J. and Liu, Q. (2018). Action-depedent control variates for policy optimization via Stein's identity. In *Proceedings of the 6th International Conference on Learning Representations*.

[56] Liu, Q., Lee, J. and Jordan, M. (2016). A kernelized Stein discrepancy for goodness-of-fit tests and model evaluation. In *Proceedings of the 33rd International Conference on Machine Learning*.

[57] Liu, Q. and Lee, J.D. (2017). Black-box importance sampling. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistic*.

[58] Liu, Q. and Wang, D. (2016). Stein variational gradient descent: A general purpose Bayesian inference algorithm. In *Proceedings of the 30th Annual Conference on Neural Information Processing Systems* 2378–2386.

[59] Mak, S. and Joseph, V.R. (2018). Support points. *Ann. Statist.* **46** 2562–2592. MR3851748 https://doi.org/10.1214/17-AOS1629

[60] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR1828667

[61] Marinucci, D. and Peccati, G. (2011). *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications*. *London Mathematical Society Lecture Note Series* **389**. Cambridge: Cambridge Univ. Press. MR2840154 https://doi.org/10.1017/CBO9780511751677
[62] Marzo, J. and Mas, A. (2021). Discrepancy of Minimal Riesz Energy Points. *Constr. Approx.* **54** 473–506. MR4342622 https://doi.org/10.1007/s00365-021-09534-5
[63] Marzouk, Y., Moselhy, T., Parno, M. and Spantini, A. (2016). Sampling via measure transport: An introduction. In *Handbook of Uncertainty Quantification* Springer.
[64] Meyn, S.P. and Tweedie, R.L. (2012). *Markov Chains and Stochastic Stability*. London: Springer.
[65] Mijatović, A. and Vogrinc, J. (2019). Asymptotic variance for random walk Metropolis chains in high dimensions: Logarithmic growth via the Poisson equation. *Adv. in Appl. Probab.* **51** 994–1026. MR4032170 https://doi.org/10.1017/apr.2019.40
[66] Mijoule, G., Reinert, G. and Swan, Y. (2021). Stein's density method for multivariate continuous distributions. Available at arXiv:2101.05079.
[67] Mira, A., Solgi, R. and Imparato, D. (2013). Zero variance Markov chain Monte Carlo for Bayesian estimators. *Stat. Comput.* **23** 653–662. MR3094805 https://doi.org/10.1007/s11222-012-9344-6
[68] Mroueh, Y. (2018). Regularized finite dimensional kernel Sobolev discrepancy. Available at arXiv:1805.06441.
[69] Mroueh, Y., Li, C.-L., Sercu, T., Raj, A. and Cheng, Y. (2017). Sobolev GAN. In *Proceedings of the 6th International Conference on Learning Representations*.
[70] Mroueh, Y., Sercu, T. and Raj, A. (2019). Sobolev descent. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*.
[71] Narcowich, F.J., Sun, X. and Ward, J.D. (2007). Approximation power of RBFs and their associated SBFs: A connection. *Adv. Comput. Math.* **27** 107–124. MR2317924 https://doi.org/10.1007/s10444-005-7506-1
[72] Narcowich, F.J. and Ward, J.D. (2002). Scattered data interpolation on spheres: Error estimates and locally supported basis functions. *SIAM J. Math. Anal.* **33** 1393–1410. MR1920637 https://doi.org/10.1137/S0036141001395054
[73] Nicolaescu, L.I. (2007). *Lectures on the Geometry of Manifolds*, 2nd ed. Hackensack, NJ: World Scientific Co. Pte. Ltd. MR2363924 https://doi.org/10.1142/9789812770295
[74] Niu, M., Cheung, P., Lin, L., Dai, Z., Lawrence, N. and Dunson, D. (2019). Intrinsic Gaussian processes on complex constrained domains. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 603–627. MR3961500 https://doi.org/10.1111/rssb.12320
[75] Oates, C.J., Cockayne, J., Briol, F.-X. and Girolami, M. (2019). Convergence rates for a class of estimators based on Stein's method. *Bernoulli* **25** 1141–1159. MR3920368 https://doi.org/10.3150/17-bej1016
[76] Oates, C.J., Girolami, M. and Chopin, N. (2017). Control functionals for Monte Carlo integration. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 695–718. MR3641403 https://doi.org/10.1111/rssb.12185
[77] Philippe, A. and Robert, C.P. (2001). Riemann sums for MCMC estimation and convergence monitoring. *Stat. Comput.* **11** 103–115. MR1837130 https://doi.org/10.1023/A:1008926514119
[78] Porcu, E., Bevilacqua, M. and Genton, M.G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Amer. Statist. Assoc.* **111** 888–898. MR3538713 https://doi.org/10.1080/01621459.2015.1072541
[79] Reznikov, A. and Saff, E.B. (2016). The covering radius of randomly distributed points on a manifold. *Int. Math. Res. Not. IMRN* **19** 6065–6094. MR3567267 https://doi.org/10.1093/imrn/rnv342
[80] Riabiz, M., Chen, W., Cockayne, J., Swietach, P., Niederer, S.A., Mackey, L. and Oates, C.J. (2021). Optimal thinning of MCMC output. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*. To appear.
[81] Roussel, J. and Stoltz, G. (2019). A perturbative approach to control variates in molecular dynamics. *Multiscale Model. Simul.* **17** 552–591. MR3932612 https://doi.org/10.1137/18M1171047
[82] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x
[83] Scheuerer, M., Schaback, R. and Schlather, M. (2013). Interpolation of spatial data—a stochastic or a deterministic problem? *European J. Appl. Math.* **24** 601–629. MR3082868 https://doi.org/10.1017/S0956792513000016

[84] Schölkopf, B., Herbrich, R. and Smola, A. (2001). A generalized representer theorem. In *Proceedings of the 14th Annual Conference on Computational Learning Theory*.
[85] Schwab, C. and Stuart, A.M. (2012). Sparse deterministic approximation of Bayesian inverse problems. *Inverse Probl.* **28** 045003. MR2903278 https://doi.org/10.1088/0266-5611/28/4/045003
[86] Semechko, A. (2015). Suite of functions to perform uniform sampling of a sphere. MATLAB File Exchange Server.
[87] Si, S., Oates, C.J., Duncan, A.B., Carin, L. and Briol, F.X. (2020). Scalable control variates for Monte Carlo methods via stochastic optimization. Available at arXiv:2006.07487.
[88] Simon-Gabriel, C.-J., Barp, A. and Mackey, L. (2020). Metrizing weak convergence with maximum mean discrepancies. Available at arXiv:2006.09268.
[89] Simon-Gabriel, C.-J. and Schölkopf, B. (2018). Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. *J. Mach. Learn. Res.* **19** 44. MR3874152
[90] Smola, A., Gretton, A., Song, L. and Schölkopf, B. (2007). A Hilbert space embedding for distributions. In *Proceedings of the 18th International Conference on Algorithmic Learning Theory*.
[91] South, L.F., Karvonen, T., Nemeth, C., Girolami, M. and Oates, C.J. (2021). Semi-exact control functionals from Sard's method. *Biometrika*. To appear.
[92] South, L.F., Oates, C.J., Mira, A. and Drovandi, C. (2018). Regularised zero-variance control variates for high-dimensional variance reduction. Available at arXiv:1811.05073.
[93] South, L.F., Pettitt, A.N., Friel, N. and Drovandi, C.C. (2017). Efficient use of derivative information within SMC methods for static Bayesian models. Technical Report Queensland Univ. Technology.
[94] Sriperumbudur, B.K., Fukumizu, K. and Lanckriet, G.R.G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410. MR2825431
[95] 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 (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 583–602. MR0402873
[96] Stein, M.L. (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.
[97] Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Berlin: Springer.
[98] Stone, M.H. (1948). The generalized Weierstrass approximation theorem. *Math. Mag.* **21** 237–254. MR0027121 https://doi.org/10.2307/3029750
[99] Szekeres, P. (2004). *A Course in Modern Mathematical Physics: Groups, Hilbert Space and Differential Geometry*. Cambridge: Cambridge Univ. Press. MR2111569 https://doi.org/10.1017/CBO9780511607066
[100] Teymur, O., Gorham, J., Riabiz, M., Oates, C. et al. (2021). Optimal quantisation of probability measures using maximum mean discrepancy. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*.
[101] Traub, J.F. (2003). *Information-Based Complexity*. New York: Wiley.
[102] Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **4** 389–396. MR1366510 https://doi.org/10.1007/BF02123482
[103] Wendland, H. (1998). Error estimates for interpolation by compactly supported radial basis functions of minimal degree. *J. Approx. Theory* **93** 258–272. MR1616781 https://doi.org/10.1006/jath.1997.3137
[104] Xu, W. and Matsuda, T. (2020). A Stein goodness-of-fit test for directional distributions. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*.
[105] Zastavnyi, V. (2006). On some properties of Buhmann functions. *Ukrainian Math. J.* **58** 1184–1208.
[106] Zhu, Z., Wan, R. and Zhong, M. (2018). Neural control variates for variance reduction. Available at arXiv:1806.00159.
[107] Zhuo, J., Liu, C., Shi, J., Zhu, J., Chen, N. and Zhang, B. (2018). Message passing Stein variational gradient descent. In *Proceedings of the 35th International Conference on Machine Learning* 6013–6022.

Bernoulli **28**(4), 2022, 2209–2233
https://doi.org/10.3150/21-BEJ1416

Rank-based change-point analysis for long-range dependent time series

ANNIKA BETKEN[1,a] and MARTIN WENDLER[2,b]

[1]*Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Enschede, The Netherlands.* [a]*a.betken@utwente.nl*
[2]*Faculty of Mathematics, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany.* [b]*martin.wendler@ovgu.de*

We consider change-point tests based on rank statistics to test for structural changes in long-range dependent observations. Under the hypothesis of stationary time series and under the assumption of a change with decreasing change-point height, the asymptotic distributions of corresponding test statistics are derived. For this, a uniform reduction principle for the sequential empirical process in a two-parameter Skorohod space equipped with a weighted supremum norm is proved. Moreover, we compare the efficiency of rank tests resulting from the consideration of different score functions. Under Gaussianity, the asymptotic relative efficiency of rank-based tests with respect to the CuSum test is 1, irrespective of the score function. Regarding the practical implementation of rank-based change-point tests, we suggest to combine self-normalized rank statistics with subsampling. The theoretical results are accompanied by simulation studies that, in particular, allow for a comparison of rank tests resulting from different score functions. With respect to the finite sample performance of rank-based change-point tests, the Van der Waerden rank test proves to be favorable in a broad range of situations. Finally, we analyze data sets from economy, hydrology, and network traffic monitoring in view of structural changes and compare our results to previous analysis of the data.

Keywords: Rank statistic; change-point; long memory; self-normalization; subsampling; empirical process; asymptotic relative efficiency

References

[1] Bai, S. and Taqqu, M.S. (2017). On the validity of resampling methods under long memory. *Ann. Statist.* **45** 2365–2399. MR3737895 https://doi.org/10.1214/16-AOS1524
[2] Bai, S. and Taqqu, M.S. (2019). Sensitivity of the Hermite rank. *Stochastic Process. Appl.* **129** 822–840. MR3913269 https://doi.org/10.1016/j.spa.2018.03.020
[3] Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Heidelberg: Springer. MR3075595 https://doi.org/10.1007/978-3-642-35512-7
[4] Beran, J., Möhrle, S. and Ghosh, S. (2016). Testing for Hermite rank in Gaussian subordination processes. *J. Comput. Graph. Statist.* **25** 917–934. MR3533645 https://doi.org/10.1080/10618600.2015.1056345
[5] Betken, A. (2016). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *J. Time Series Anal.* **37** 785–809. MR3564308 https://doi.org/10.1111/jtsa.12187
[6] Betken, A., Buchsteiner, J., Dehling, H., Münker, I., Schnurr, A. and Woerner, J.H.C. (2021). Ordinal patterns in long-range dependent time series. *Scand. J. Stat.* **48** 969–1000. MR4303565 https://doi.org/10.1111/sjos.12478
[7] Betken, A. and Kulik, R. (2019). Testing for change in long-memory stochastic volatility time series. *J. Time Series Anal.* **40** 707–738. MR3995665 https://doi.org/10.1111/jtsa.12449
[8] Betken, A. and Wendler, M. (2018). Subsampling for general statistics under long range dependence with application to change point analysis. *Statist. Sinica* **28** 1199–1224. MR3821001
[9] Betken, A. and Wendler, M. (2022). Supplement to "Rank-based change-point analysis for long-range dependent time series." https://doi.org/10.3150/21-BEJ1416SUPP

[10] Beutner, E. and Zähle, H. (2012). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli* **18** 803–822. MR2948902 https://doi.org/10.3150/11-BEJ358
[11] Bhattacharyya, G.K. and Johnson, R.A. (1968). Nonparametric tests for shift at an unknown time point. *Ann. Math. Stat.* **39** 1731–1743. MR0230425 https://doi.org/10.1214/aoms/1177698156
[12] Buchsteiner, J. (2015). Weak convergence of the weighted sequential empirical process of some long-range dependent data. *Statist. Probab. Lett.* **96** 170–179. MR3281762 https://doi.org/10.1016/j.spl.2014.09.022
[13] Chan, N.H., Ng, W.L. and Yau, C.Y. (2021). A self-normalized approach to sequential change-point detection for time series. *Statist. Sinica* **31** 491–517. MR4270394 https://doi.org/10.5705/ss.20
[14] Darkhovskh, B.S. (1976). A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory Probab. Appl.* **21** 178–183.
[15] Dehling, H., Rooch, A. and Taqqu, M.S. (2013). Non-parametric change-point tests for long-range dependent data. *Scand. J. Stat.* **40** 153–173. MR3024037 https://doi.org/10.1111/j.1467-9469.2012.00799.x
[16] Dehling, H., Rooch, A. and Taqqu, M.S. (2017). Power of change-point tests for long-range dependent data. *Electron. J. Stat.* **11** 2168–2198. MR3654823 https://doi.org/10.1214/17-EJS1283
[17] Dette, H. and Gösmann, J. (2020). A likelihood ratio approach to sequential change point detection for a general class of parameters. *J. Amer. Statist. Assoc.* **115** 1361–1377. MR4143471 https://doi.org/10.1080/01621459.2019.1630562
[18] Dette, H., Kokot, K. and Volgushev, S. (2020). Testing relevant hypotheses in functional time series via self-normalization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 629–660. MR4112779 https://doi.org/10.1111/rssb.12370
[19] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, 3rd ed. New York: Wiley. MR0228020
[20] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Series Anal.* **4** 221–238. MR0738585 https://doi.org/10.1111/j.1467-9892.1983.tb00371.x
[21] Gombay, E. (1994). Testing for change-points with rank and sign statistics. *Statist. Probab. Lett.* **20** 49–55. MR1294803 https://doi.org/10.1016/0167-7152(94)90233-X
[22] Gombay, E. and Hušková, M. (1998). Rank based estimators of the change-point. *J. Statist. Plann. Inference* **67** 137–154. MR1624682 https://doi.org/10.1016/S0378-3758(97)00099-2
[23] Hall, P. and Jing, B. (1996). On sample reuse methods for dependent data. *J. Roy. Statist. Soc. Ser. B* **58** 727–737. MR1410187
[24] Hall, P., Jing, B.-Y. and Lahiri, S.N. (1998). On the sampling window method for long-range dependent data. *Statist. Sinica* **8** 1189–1204. MR1666245
[25] Hodges, J.L. Jr. and Lehmann, E.L. (1961). Comparison of the normal scores and Wilcoxon tests. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 307–317. Berkeley, Calif.: Univ. California Press. MR0132645
[26] Jach, A., McElroy, T. and Politis, D.N. (2012). Subsampling inference for the mean of heavy-tailed long-memory time series. *J. Time Series Anal.* **33** 96–111. MR2877610 https://doi.org/10.1111/j.1467-9892.2011.00742.x
[27] Jach, A., McElroy, T.S. and Politis, D.N. (2016). Corrigendum to 'Subsampling inference for the mean of heavy-tailed long-memory time series' [MR2877610]. *J. Time Series Anal.* **37** 713–720. MR3536227 https://doi.org/10.1111/jtsa.12191
[28] Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the 1st World Congress of the Bernoulli Society, Vol. 1 (Tashkent, 1986)* 67–74. Utrecht: VNU Sci. Press. MR1092336
[29] Lobato, I.N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.* **96** 1066–1076. MR1947254 https://doi.org/10.1198/016214501753208726
[30] Lombard, F. (1987). Rank tests for changepoint problems. *Biometrika* **74** 615–624. MR0909366 https://doi.org/10.1093/biomet/74.3.615
[31] Mandelbrot, B.B. (1974/75). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Z. Wahrsch. Verw. Gebiete* **31** 271–285. MR0423481 https://doi.org/10.1007/BF00532867
[32] Mandelbrot, B.B. and Taqqu, M.S. (1979). Robust R/S analysis of long-run serial correlation. *Bull. Int. Stat. Inst.* **48** 69–99. MR0731558

[33] Mandelbrot, B.B. and Wallis, J.R. (1969). Computer experiments with fractional Gaussian noises: Part 1, averages and variances. *Water Resour. Res.* **5** 228–241.
[34] Nordman, D.J. and Lahiri, S.N. (2005). Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory* **21** 1087–1111. MR2200986 https://doi.org/10.1017/S0266466605050541
[35] Pešta, M. and Wendler, M. (2020). Nuisance-parameter-free changepoint detection in non-stationary series. *TEST* **29** 379–408. MR4095034 https://doi.org/10.1007/s11749-019-00659-1
[36] Pettitt, A.N. (1979). Two-sample Cramér-von Mises type rank statistics. *J. Roy. Statist. Soc. Ser. B* **41** 46–53. MR0535544
[37] Pipiras, V. and Taqqu, M.S. (2017). *Long-Range Dependence and Self-Similarity*. *Cambridge Series in Statistical and Probabilistic Mathematics*, **45**. Cambridge: Cambridge Univ. Press. MR3729426
[38] Politis, D.N. and Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181 https://doi.org/10.1214/aos/1176325770
[39] Praagman, J. (1988). Bahadur efficiency of rank tests for the change-point problem. *Ann. Statist.* **16** 198–217. MR0924866 https://doi.org/10.1214/aos/1176350700
[40] Pyke, R. and Shorack, G.R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Stat.* **39** 755–771. MR0226770 https://doi.org/10.1214/aoms/1177698309
[41] Rea, W., Oxley, L., Reale, M. and Brown, J. (2009). Estimators for long range dependence: An empirical study. Preprint. Available at arXiv:0901.0762.
[42] Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301 https://doi.org/10.1214/aos/1176324317
[43] Sen, P.K. (1978). Invariance principles for linear rank statistics revisited. *Sankhyā Ser. A* **40** 215–236. MR0589279
[44] Shao, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *J. Time Series Anal.* **32** 598–606. MR2846559 https://doi.org/10.1111/j.1467-9892.2010.00717.x
[45] Shao, X. (2015). Self-normalization for time series: A review of recent developments. *J. Amer. Statist. Assoc.* **110** 1797–1817. MR3449074 https://doi.org/10.1080/01621459.2015.1050493
[46] Shao, X. and Zhang, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. MR2752617 https://doi.org/10.1198/jasa.2010.tm10103
[47] Sherman, M. and Carlstein, E. (1996). Replicate histograms. *J. Amer. Statist. Assoc.* **91** 566–576. MR1395726 https://doi.org/10.2307/2291653
[48] Sinn, M. and Keller, K. (2011). Estimation of ordinal pattern probabilities in Gaussian processes with stationary increments. *Comput. Statist. Data Anal.* **55** 1781–1790. MR2748679 https://doi.org/10.1016/j.csda.2010.11.009
[49] Szyszkowicz, B. (1994). Weak convergence of weighted empirical type processes under contiguous and changepoint alternatives. *Stochastic Process. Appl.* **50** 281–313. MR1273777 https://doi.org/10.1016/0304-4149(94)90125-2
[50] Taqqu, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* **50** 53–83. MR0550123 https://doi.org/10.1007/BF00535674
[51] Taqqu, M.S., Teverovsky, V. and Willinger, W. (1995). Estimators for long-range dependence: An empirical study. *Fractals* **3** 785–798.
[52] Wang, L. (2008). Change-point detection with rank statistics in long-memory time-series models. *Aust. N. Z. J. Stat.* **50** 241–256. MR2455431 https://doi.org/10.1111/j.1467-842X.2008.00515.x
[53] Wolfe, D.A. and Schechtman, E. (1984). Nonparametric statistical procedures for the changepoint problem. *J. Statist. Plann. Inference* **9** 389–396. MR0748267 https://doi.org/10.1016/0378-3758(84)90013-2
[54] Zhang, T., Ho, H.-C., Wendler, M. and Wu, W.B. (2013). Block sampling under strong dependence. *Stochastic Process. Appl.* **123** 2323–2339. MR3038507 https://doi.org/10.1016/j.spa.2013.02.006
[55] Zhang, T. and Lavitas, L. (2018). Unsupervised self-normalized change-point testing for time series. *J. Amer. Statist. Assoc.* **113** 637–648. MR3832215 https://doi.org/10.1080/01621459.2016.1270214
[56] Zhao, Z. and Li, X. (2013). Inference for modulated stationary processes. *Bernoulli* **19** 205–227. MR3019492 https://doi.org/10.3150/11-BEJ399

Bernoulli **28**(4), 2022, 2234–2259
https://doi.org/10.3150/21-BEJ1417

Splitting the sample at the largest uncensored observation

ROSS MALLER[1,a], SIDNEY RESNICK[2,b] and SOUDABEH SHEMEHSAVAR[3,c]

[1]*Research School of Finance, Actuarial Studies & Statistics, The Australian National University, Canberra, Australia.* [a]*Ross.Maller@anu.edu.au*
[2]*School of Operations Research & Information Engineering, Cornell University, Ithaca, New York, USA.* [b]*sir1@cornell.edu*
[3]*School of Mathematics, Statistics and Computer Science, University of Tehran, Tehran, Iran.* [c]*shemehsavar@ut.ac.ir*

We calculate finite sample and asymptotic distributions for the largest censored and uncensored survival times, and some related statistics, from a sample of survival data generated according to an iid censoring model. These statistics are important for assessing whether there is sufficient follow-up in the sample to be confident of the presence of immune or cured individuals in the population. A key structural result obtained is that, conditional on the value of the largest uncensored survival time, and knowing the number of censored observations exceeding this time, the sample partitions into two independent subsamples, each subsample having the distribution of an iid sample of censored survival times, of reduced size, from truncated random variables. This result provides valuable insight into the construction of censored survival data, and facilitates the calculation of explicit finite sample formulae. We illustrate by calculating distributions of statistics useful for testing for sufficient follow-up in a sample, and apply extreme value methods to derive asymptotic distributions for some of those.

Keywords: Survival data; largest censored and uncensored survival times; sufficient follow-up; immune or cured individuals; cure model; extreme value methods

References

[1] Amico, M. and Van Keilegom, I. (2018). Cure models in survival analysis. *Ann. Rev. Stat. Appl.* **5** 311–345. MR3774750 https://doi.org/10.1146/annurev-statistics-031017-100101
[2] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156 https://doi.org/10.1007/0-387-34471-3
[3] Escobar-Bach, M., Maller, R.A., Van Keilegom, I. and Zhao, M. (2021). Estimation of the cure rate for distributions in the Gumbel maximum domain of attraction under insufficient follow-up. *Biometrika*. https://doi.org/10.1093/biomet/asaa106.
[4] Gill, R.D. (1980). *Censoring and Stochastic Integrals*. *Mathematical Centre Tracts* **124**. Amsterdam: Mathematisch Centrum. MR0596815
[5] Goldman, A. (1984). Survivorship analysis when cure is a possibility: A Monte Carlo study. *Stat. Med.* **3** 153–163.
[6] Goldman, A. (1991). The cure model and time confounded risk in the analysis of survival and other timed events. *J. Clin. Epidemiol.* **44** 1327–1340.
[7] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481. MR0093867
[8] Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., Omberg, L., Wolf, D.M., Shriver, C.D., Thorsson, V. and Hu, H. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173** 400–416.

[9] Maller, R.A. and Resnick, S.I. (2021). Extremes of censored and uncensored lifetimes in survival data. *Extremes*. https://doi.org/10.1007/s10687-021-00426-2
[10] Maller, R.A., Resnick, S.I. and Shemehsavar, S. (2020). Splitting the sample at the largest uncensored observation. Available at arXiv:2103.01337.
[11] Maller, R.A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. *Wiley Series in Probability and Statistics: Applied Probability and Statistics*. Chichester: Wiley. MR1453117
[12] Othus, M., Bansal, A., Erba, H. and Ramsey, S. (2020). Bias in mean survival from fitting cure models with limited follow-up. *Value Health* **23** 1034–1039. https://doi.org/10.1016/j.jval.2020.02.015
[13] Othus, M., Barlogie, B., LeBlanc, M.L. and Crowley, J.J. (2012). Cure models as a useful statistical tool for analyzing survival. *Clin. Cancer Res.* **18** 311–342.
[14] Peng, Y. and Taylor, J.M.G. (2014). Cure models. In *Handbook of Survival Analysis* (J. Klein, H. van Houwelingen, J.G. Ibrahim and T.H. Scheike, eds.). *Handbooks of Modern Statistical Methods Series* 113–134. Boca Raton, FL, USA: CRC Press.
[15] Resnick, S.I. (2008). *Extreme Values, Regular Variation and Point Processes*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. Reprint of the 1987 original. MR2364939
[16] Shen, P. (2000). Testing for sufficient follow-up in survival data. *Statist. Probab. Lett.* **49** 313–322. MR1796475 https://doi.org/10.1016/S0167-7152(00)00063-8
[17] Taweab, F. and Ibrahim, N.A. (2014). Cure rate models: A review of recent progress with a study of change-point cure models when cured is partially known. *J. Appl. Sci.* **14** 609–616.

Bernoulli **28**(4), 2022, 2260–2293
https://doi.org/10.3150/21-BEJ1418

Pivotal tests for relevant differences in the second order dynamics of functional time series

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

[1]*Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA.*
[a]*anne.vandelft@columbia.edu*
[2]*Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany.* [b]*holger.dette@rub.de*

Motivated by the need to statistically quantify differences between modern (complex) data-sets which commonly result as high-resolution measurements of stochastic processes varying over a continuum, we propose novel testing procedures to detect relevant differences between the second order dynamics of two functional time series. In order to take the between-function dynamics into account that characterize this type of functional data, a frequency domain approach is taken. Test statistics are developed to compare differences in the spectral density operators and in the primary modes of variation as encoded in the associated eigenelements. Under mild moment conditions, we show convergence of the underlying statistics to Brownian motions and construct pivotal test statistics. The latter is essential because the nuisance parameters can be unwieldy and their robust estimation infeasible, especially if the two functional time series are dependent. In addition to these novel features, the properties of the tests are robust to any choice of frequency band enabling also to compare energy contents at a single frequency. The finite sample performance of the tests are verified through a simulation study and are illustrated with an application to fMRI data.

Keywords: Functional data; time series; spectral analysis; relevant tests; self-normalization; martingale theory

References

[1] Antoniadis, A., Paparoditis, E. and Sapatinas, T. (2006). A functional wavelet-kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 837–857. MR2301297 https://doi.org/10.1111/j.1467-9868.2006.00569.x

[2] Aston, J.A.D. and Kirch, C. (2012). Detecting and estimating changes in dependent functional data. *J. Multivariate Anal.* **109** 204–220. MR2922864 https://doi.org/10.1016/j.jmva.2012.03.006

[3] Aue, A., Dette, H. and Rice, G. (2019). Two-sample tests for relevant differences in the eigenfunctions of covariance operators. Available at arXiv:1909.06098.

[4] Benko, M., Härdle, W. and Kneip, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343 https://doi.org/10.1214/07-AOS516

[5] Berkson, J. (1940). A note on the chi-square test, the Poisson and the binomial. *J. Amer. Statist. Assoc.* **35** 362–367. MR0002074

[6] Deheuvels, P. and Martynov, G. (2003). Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. In *High Dimensional Probability, III (Sandjberg, 2002). Progress in Probability* **55** 57–93. Basel: Birkhäuser. MR2033881

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

[8] Fiecas, M. and Ombao, H. (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *J. Amer. Statist. Assoc.* **111** 1440–1453. MR3601700 https://doi.org/10.1080/01621459.2016.1165683

[9] Fogarty, C.B. and Small, D.S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Ann. Appl. Stat.* **8** 2002–2026. MR3292487 https://doi.org/10.1214/14-AOAS763

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

[11] Guo, J., Zhou, B. and Zhang, J.-T. (2018). Testing the equality of several covariance functions for functional data: A supremum-norm based test. *Comput. Statist. Data Anal.* **124** 15–26. MR3787611 https://doi.org/10.1016/j.csda.2018.02.002

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

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

[14] Hörmann, S., Kokoszka, P. and Nisol, G. (2018). Testing for periodicity in functional time series. *Ann. Statist.* **46** 2960–2984. MR3851761 https://doi.org/10.1214/17-AOS1645

[15] Kadison, R.V. and Ringrose, J.R. (1997). *Fundamentals of the Theory of Operator Algebras. Vol. I. Graduate Studies in Mathematics* **15**. Providence, RI: Amer. Math. Soc. MR1468229 https://doi.org/10.1090/gsm/015

[16] Kowal, D.R., Matteson, D.S. and Ruppert, D. (2019). Functional autoregression for sparsely sampled data. *J. Bus. Econom. Statist.* **37** 97–109. MR3910228 https://doi.org/10.1080/07350015.2017.1279058

[17] Leucht, A., Paporoditis, E. and Sapatinas, T. (2018). Testing equality of spectral density operators for functional linear processes. Available at arXiv:1804.03366.

[18] McLeish, D.L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628. MR0358933 https://doi.org/10.1214/aop/1176996608

[19] Móricz, F. (1976). Moment inequalities and the strong laws of large numbers. *Z. Wahrsch. Verw. Gebiete* **35** 299–314. MR0407950 https://doi.org/10.1007/BF00532956

[20] Panaretos, V.M., Kraus, D. and Maddocks, J.H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Amer. Statist. Assoc.* **105** 670–682. MR2724851 https://doi.org/10.1198/jasa.2010.tm09239

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

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

[23] Paparoditis, E. and Sapatinas, T. (2016). Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika* **103** 727–733. MR3551795 https://doi.org/10.1093/biomet/asw033

[24] Pilavakis, D., Paparoditis, E. and Sapatinas, T. (2020). Testing equality of autocovariance operators for functional time series. *J. Time Series Anal.* **41** 571–589. MR4136903 https://doi.org/10.1111/jtsa.12523

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

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

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

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

[29] Tavakoli, S. and Panaretos, V.M. (2016). Detecting and localizing differences in functional time series dynamics: A case study in molecular biophysics. *J. Amer. Statist. Assoc.* **111** 1020–1035. MR3561926 https://doi.org/10.1080/01621459.2016.1147355

[30] van Delft, A. (2020). A note on quadratic forms of stationary functional time series under mild conditions. *Stochastic Process. Appl.* **130** 4206–4251. MR4102264 https://doi.org/10.1016/j.spa.2019.12.002. arXiv:1905.13186

[31] van Delft, A., Dette, H. (2022). Supplement to “Pivotal tests for relevant differences in the second order dynamics of functional time series.” https://doi.org/10.3150/21-BEJ1418SUPP

[32] van Delft, A. and Dette, H. (2021). A similarity measure for second order properties of non-stationary functional time series with applications to clustering and testing. *Bernoulli* **27** 469–501. MR4177377 https://doi.org/10.3150/20-BEJ1246
[33] van Delft, A. and Eichler, M. (2018). Locally stationary functional time series. *Electron. J. Stat.* **12** 107–170. MR3746979 https://doi.org/10.1214/17-EJS1384
[34] van Delft, A. and Eichler, M. (2020). A note on Herglotz's theorem for time series on function spaces. *Stochastic Process. Appl.* **130** 3687–3710. MR4092417 https://doi.org/10.1016/j.spa.2019.10.006
[35] Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd ed. Boca Raton, FL: CRC Press. MR2676002 https://doi.org/10.1201/EBK1439808184
[36] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102
[37] Yuen, N.H., Osachoff, N. and Chen, J.J. (2019). Intrinsic frequencies of the resting-state fMRI signal: The frequency dependence of functional connectivity and the effect of mode mixing. *Front. Neurosci.* **13** 900–917.
[38] Zhang, X. and Shao, X. (2015). Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli* **21** 909–929. MR3338651 https://doi.org/10.3150/13-BEJ592
[39] Zhang, X., Shao, X., Hayhoe, K. and Wuebbles, D.J. (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Stat.* **5** 1765–1796. MR2870150 https://doi.org/10.1214/11-EJS655

Bernoulli **28**(4), 2022, 2294–2321
https://doi.org/10.3150/21-BEJ1419

Functional inequalities for perturbed measures with applications to log-concave measures and to some Bayesian problems

PATRICK CATTIAUX[1,a] and ARNAUD GUILLIN[2,b]

[1]*Institut de Mathématiques de Toulouse, CNRS UMR 5219, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 09, France.* [a]*patrick.cattiaux@math.univ-toulouse.fr*
[2]*Université Clermont Auvergne, CNRS, LMBP, F-63000 Clermont-Ferrand, France.* [b]*arnaud.guillin@uca.fr*

We study functional inequalities (Poincaré, Cheeger, log-Sobolev) for probability measures obtained as perturbations. Several explicit results for general measures as well as log-concave distributions are given. The initial goal of this work was to obtain explicit bounds on the constants in view of statistical applications. These results are then applied to the Langevin Monte-Carlo method used in statistics in order to compute Bayesian estimators.

Keywords: Logconcave measure; Poincaré inequality; Cheeger inequality; logarithmic Sobolev inequality; perturbation; bayesian statistic; sparse learning

References

[1] Aida, S. and Shigekawa, I. (1994). Logarithmic Sobolev inequalities and spectral gaps: Perturbation theory. *J. Funct. Anal.* **126** 448–475. MR1305076 https://doi.org/10.1006/jfan.1994.1154
[2] Alonso-Gutiérrez, D. and Bastero, J. (2015). *Approaching the Kannan-Lovász-Simonovits and Variance Conjectures*. *Lecture Notes in Math.* **2131**. Cham: Springer. MR3308635 https://doi.org/10.1007/978-3-319-13263-1
[3] 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
[4] Bakry, D., Cattiaux, P. and Guillin, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. MR2381160 https://doi.org/10.1016/j.jfa.2007.11.002
[5] 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
[6] Bakry, D., Ledoux, M. and Wang, F.-Y. (2007). Perturbations of functional inequalities using growth conditions. *J. Math. Pures Appl. (9)* **87** 394–407. MR2317340 https://doi.org/10.1016/j.matpur.2007.01.006
[7] Bardet, J.-B., Gozlan, N., Malrieu, F. and Zitt, P.-A. (2018). Functional inequalities for Gaussian convolutions of compactly supported measures: Explicit bounds and dimension dependence. *Bernoulli* **24** 333–353. MR3706760 https://doi.org/10.3150/16-BEJ879
[8] Barthe, F. and Klartag, B. (2020). Spectral gaps, symmetries and log-concave perturbations. *Bull. Hellenic Math. Soc.* **64** 1–31. MR4075908
[9] Barthe, F. and Milman, E. (2013). Transference principles for log-Sobolev and spectral-gap with applications to conservative spin systems. *Comm. Math. Phys.* **323** 575–625. MR3096532 https://doi.org/10.1007/s00220-013-1782-2
[10] Berestycki, N. and Nickl, R. (2009). Concentration of measure. Available at http://www.statslab.cam.ac.uk/~beresty/teach/cm10.pdf.
[11] Bobkov, S.G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27** 1903–1921. MR1742893 https://doi.org/10.1214/aop/1022874820

[12] Bonnefont, M., Joulin, A. and Ma, Y. (2016). A note on spectral gap and weighted Poincaré inequalities for some one-dimensional diffusions. *ESAIM Probab. Stat.* **20** 18–29. MR3519678 https://doi.org/10.1051/ps/2015019
[13] Brascamp, H.J. and Lieb, E.H. (1976). On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480 https://doi.org/10.1016/0022-1236(76)90004-5
[14] Cattiaux, P. (2004). A pathwise approach of some classical inequalities. *Potential Anal.* **20** 361–394. MR2032116 https://doi.org/10.1023/B:POTA.0000009847.84908.6f
[15] Cattiaux, P. (2005). Hypercontractivity for perturbed diffusion semigroups. *Ann. Fac. Sci. Toulouse Math. (6)* **14** 609–628. MR2188585
[16] Cattiaux, P., Fathi, M. and Guillin, A. (2020). Self-improvement of the Bary-Emery criterion for Poincaré inequalities and Wasserstein contraction using variable curvature bounds. Available on arXiv:2002.09221 [math FA].
[17] Cattiaux, P. and Guillin, A. (2014). Semi log-concave Markov diffusions. In *Séminaire de Probabilités XLVI*. *Lecture Notes in Math.* **2123** 231–292. Cham: Springer. MR3330820 https://doi.org/10.1007/978-3-319-11970-0_9
[18] Cattiaux, P. and Guillin, A. (2017). Hitting times, functional inequalities, Lyapunov conditions and uniform ergodicity. *J. Funct. Anal.* **272** 2361–2391. MR3603301 https://doi.org/10.1016/j.jfa.2016.10.003
[19] Cattiaux, P. and Guillin, A. (2020). On the Poincaré constant of log-concave measures. In *Geometric Aspects of Functional Analysis. Vol. I*. *Lecture Notes in Math.* **2256** 171–217. Cham: Springer. MR4175748 https://doi.org/10.1007/978-3-030-36020-7_9
[20] Cattiaux, P. and Guillin, A. (2022). Supplement to "Functional inequalities for perturbed measures with applications to log-concave measures and to some Bayesian problems." https://doi.org/10.3150/21-BEJ1419SUPP
[21] Cattiaux, P., Guillin, A., Monmarché, P. and Zhang, C. (2019). Entropic multipliers method for Langevin diffusion and weighted log Sobolev inequalities. *J. Funct. Anal.* **277** 108288. MR4013832 https://doi.org/10.1016/j.jfa.2019.108288
[22] Cattiaux, P., Guillin, A. and Wu, L.-M. (2010). A note on Talagrand's transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields* **148** 285–304. MR2653230 https://doi.org/10.1007/s00440-009-0231-9
[23] Cattiaux, P., Guillin, A. and Zitt, P.A. (2013). Poincaré inequalities and hitting times. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 95–118. MR3060149 https://doi.org/10.1214/11-AIHP447
[24] Chen, Y. (2021). An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geom. Funct. Anal.* **31** 34–61. MR4244847 https://doi.org/10.1007/s00039-021-00558-4
[25] Chewi, S., Le Gouic, T., Lu, C., Maunu, T., Rigollet, P. and Stromme, A. (2020). *Exponential Ergodicity of Mirror-Langevin Diffusions*. **33**.
[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. 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
[28] Dalalyan, A.S. and Tsybakov, A.B. (2012). Sparse regression learning by aggregation and Langevin Monte-Carlo. *J. Comput. System Sci.* **78** 1423–1443. MR2926142 https://doi.org/10.1016/j.jcss.2011.12.023
[29] 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
[30] Down, D., Meyn, S.P. and Tweedie, R.L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** 1671–1691. MR1379163
[31] 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
[32] Gadat, S., Panloup, F. and Pellegrini, C. (2020). On the cost of Bayesian posterior mean strategy for log-concave models. Available on arXiv:2010.06420 [math.PR].
[33] Gong, F. and Wu, L. (2000). Spectral gap of positive operators and applications. *C. R. Acad. Sci. Paris Sér. I Math.* **331** 983–988. MR1809440 https://doi.org/10.1016/S0764-4442(00)01739-0

[34] Guédon, O., Nayar, P. and Tkocz, T. (2014). Concentration inequalities and geometry of convex bodies. In *Analytical and Probabilistic Methods in the Geometry of Convex Bodies*. *IMPAN Lect. Notes* **2** 9–86. Warsaw: Polish Acad. Sci. Inst. Math. MR3329056

[35] 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

[36] Klartag, B. (2009). A Berry-Esseen type inequality for convex bodies with an unconditional basis. *Probab. Theory Related Fields* **145** 1–33. MR2520120 https://doi.org/10.1007/s00440-008-0158-6

[37] Latała, R. (2017). On some problems concerning log-concave random vectors. In *Convexity and Concentration*. *IMA Vol. Math. Appl.* **161** 525–539. New York: Springer. MR3837281

[38] Ledoux, M. (2004). Spectral gap, logarithmic Sobolev constant, and geometric bounds. In *Surveys in Differential Geometry. Vol. IX*. *Surv. Differ. Geom.* **9** 219–240. Somerville, MA: Int. Press. MR2195409 https://doi.org/10.4310/SDG.2004.v9.n1.a6

[39] Ledoux, M. (2011). From concentration to isoperimetry: Semigroup proofs. In *Concentration, Functional Inequalities and Isoperimetry*. *Contemp. Math.* **545** 155–166. Providence, RI: Amer. Math. Soc. MR2858471 https://doi.org/10.1090/conm/545/10770

[40] Lee, Y.T. and Vempala, S.S. (2017). Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017* 998–1007. Los Alamitos, CA: IEEE Computer Soc. MR3734299 https://doi.org/10.1109/FOCS.2017.96

[41] 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

[42] Milman, E. (2009). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** 1–43. MR2507637 https://doi.org/10.1007/s00222-009-0175-9

[43] Milman, E. (2010). Isoperimetric and concentration inequalities: Equivalence under curvature lower bound. *Duke Math. J.* **154** 207–239. MR2682183 https://doi.org/10.1215/00127094-2010-038

[44] Milman, E. (2011). Isoperimetric bounds on convex manifolds. In *Concentration, Functional Inequalities and Isoperimetry*. *Contemp. Math.* **545** 195–208. Providence, RI: Amer. Math. Soc. MR2858533 https://doi.org/10.1090/conm/545/10772

[45] Payne, L.E. and Weinberger, H.F. (1960). An optimal Poincaré inequality for convex domains. *Arch. Ration. Mech. Anal.* **5** 286–292. MR0117419 https://doi.org/10.1007/BF00252910

[46] Saumard, A. and Wellner, J.A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

[47] Troyanov, M. (2001). Concentration et inégalité de Poincaré. Available at https://infoscience.epfl.ch/record/118471/files/concentration2001.pdf.

[48] Welling, M. and Yee, W.T. (2011). *Bayesian Learning via Stochastic Gradient Langevin Dynamics*.

[49] Zimmermann, D. (2013). Logarithmic Sobolev inequalities for mollified compactly supported measures. *J. Funct. Anal.* **265** 1064–1083. MR3067796 https://doi.org/10.1016/j.jfa.2013.05.029

Bernoulli **28**(4), 2022, 2322–2344
https://doi.org/10.3150/21-BEJ1420

On the characterization of Brownian bridge measure on the pinned path space over a compact Riemannian manifold

FUZHOU GONG[1,2,a] and XIAOXIA SUN[3,b]

[1] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*
[2] *School of Mathematical Science, University of the Chinese Academy of Sciences, Beijing, China.*
[a] *fzgong@amt.ac.cn*
[3] *School of Data Science and Artificial Intelligence, Dongbei University of Finance and Economics, Dalian, China.* [b] *xiaoxiasun@dufe.edu.cn*

In this paper, we focus on the characterization of a Brownian bridge measure on the pinned path space over a compact Riemannian manifold. In the case when the Riemannian manifold is simply connected, we prove that the integration by parts formula can characterize the Brownian bridge measure. Otherwise, we show that it is not always true by constructing an illustrating example.

Keywords: Brownian bridge; pull back formula; integration by parts formula; characterization

References

[1] Aida, S. (1998). Differential calculus on path and loop spaces. II. Irreducibility of Dirichlet forms on loop spaces. *Bull. Sci. Math.* **122** 635–666. MR1668546 https://doi.org/10.1016/S0007-4497(99)80007-4
[2] Aida, S. (2000). On the irreducibility of Dirichlet forms on domains in infinite-dimensional spaces. *Osaka J. Math.* **37** 953–966. MR1809914
[3] Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen-Stein method: Rejoinder. *Statist. Sci.* **5** 432–434.
[4] Barbour, A.D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. MR1035659 https://doi.org/10.1007/BF01197887
[5] Bismut, J.-M. (1984). *Large Deviations and the Malliavin Calculus*. *Progress in Mathematics* **45**. Boston, MA: Birkhäuser, Inc. MR0755001
[6] Chatterjee, S., Fulman, J. and Röllin, A. (2011). Exponential approximation by Stein's method and spectral graph theory. *ALEA Lat. Am. J. Probab. Math. Stat.* **8** 197–223. MR2802856
[7] Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. MR0428387 https://doi.org/10.1214/aop/1176996359
[8] Chen, X., Li, X.-M. and Wu, B. (2011). A concrete estimate for the weak Poincaré inequality on loop space. *Probab. Theory Related Fields* **151** 559–590. MR2851693 https://doi.org/10.1007/s00440-010-0308-5
[9] Driver, B.K. (1992). A Cameron-Martin type quasi-invariance theorem for Brownian motion on a compact Riemannian manifold. *J. Funct. Anal.* **110** 272–376. MR1194990 https://doi.org/10.1016/0022-1236(92)90035-H
[10] Driver, B.K. (1994). A Cameron-Martin type quasi-invariance theorem for pinned Brownian motion on a compact Riemannian manifold. *Trans. Amer. Math. Soc.* **342** 375–395. MR1154540 https://doi.org/10.2307/2154699
[11] Enchev, O. and Stroock, D. (1995). Integration by parts for pinned Brownian motion. *Math. Res. Lett.* **2** 161–169. MR1324699 https://doi.org/10.4310/MRL.1995.v2.n2.a5
[12] Enchev, O. and Stroock, D.W. (1995). Towards a Riemannian geometry on the path space over a Riemannian manifold. *J. Funct. Anal.* **134** 392–416. MR1363806 https://doi.org/10.1006/jfan.1995.1151

[13] Enchev, O. and Stroock, D.W. (1996). Pinned Brownian motion and its perturbations. *Adv. Math.* **119** 127–154. MR1390796 https://doi.org/10.1006/aima.1996.0029
[14] Fang, S. and Malliavin, P. (1993). Stochastic analysis on the path space of a Riemannian manifold. I. Markovian stochastic calculus. *J. Funct. Anal.* **118** 249–274. MR1245604 https://doi.org/10.1006/jfan.1993.1145
[15] Gaunt, R.E. (2014). Variance-gamma approximation via Stein's method. *Electron. J. Probab.* **19** 38. MR3194737 https://doi.org/10.1214/EJP.v19-3020
[16] Gaunt, R.E., Pickett, A.M. and Reinert, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *Ann. Appl. Probab.* **27** 720–756. MR3655852 https://doi.org/10.1214/16-AAP1213
[17] Gong, F., Röckner, M. and Wu, L. (2001). Poincaré inequality for weighted first order Sobolev spaces on loop spaces. *J. Funct. Anal.* **185** 527–563. MR1856276 https://doi.org/10.1006/jfan.2001.3775
[18] Gong, F.-Z. and Ma, Z.-M. (1998). The log-Sobolev inequality on loop space over a compact Riemannian manifold. *J. Funct. Anal.* **157** 599–623. MR1638269 https://doi.org/10.1006/jfan.1998.3263
[19] Hsu, E.P. (1995). Quasi-invariance of the Wiener measure on the path space over a compact Riemannian manifold. *J. Funct. Anal.* **134** 417–450. MR1363807 https://doi.org/10.1006/jfan.1995.1152
[20] Hsu, E.P. (1997). Integration by parts in loop spaces. *Math. Ann.* **309** 331–339. MR1474195 https://doi.org/10.1007/s002080050115
[21] Hsu, E.P. (2005). Characterization of Brownian motion on manifolds through integration by parts. In *Stein's Method and Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 195–208. Singapore: Singapore Univ. Press. MR2205337 https://doi.org/10.1142/9789812567673_0013
[22] Loh, W.-L. (1992). Stein's method and multinomial approximation. *Ann. Appl. Probab.* **2** 536–554. MR1177898
[23] Luk, H.M. (1994). *Stein's Method for the Gamma Distribution and Related Statistical Applications*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.)–University of Southern California. MR2693204
[24] Ma, Z.M. and Röckner, M. (1992). *Introduction to the Theory of (Nonsymmetric) Dirichlet Forms*. *Universitext*. Berlin: Springer. MR1214375 https://doi.org/10.1007/978-3-642-77739-4
[25] Norris, J. (1986). Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*. *Lecture Notes in Math.* **1204** 101–130. Berlin: Springer. MR0942019 https://doi.org/10.1007/BFb0075716
[26] Peköz, E.A. and Röllin, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.* **39** 587–608. MR2789507 https://doi.org/10.1214/10-AOP559
[27] Röckner, M. and Wang, F.-Y. (2001). Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups. *J. Funct. Anal.* **185** 564–603. MR1856277 https://doi.org/10.1006/jfan.2001.3776
[28] Sheu, S.J. (1991). Some estimates of the transition density of a nondegenerate diffusion Markov process. *Ann. Probab.* **19** 538–561. MR1106275
[29] Shih, H.-H. (2011). On Stein's method for infinite-dimensional Gaussian approximation in abstract Wiener spaces. *J. Funct. Anal.* **261** 1236–1283. MR2807099 https://doi.org/10.1016/j.jfa.2011.04.016
[30] 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 (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 583–602. MR0402873

Bernoulli **28**(4), 2022, 2345–2374
https://doi.org/10.3150/21-BEJ1421

Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics

FABRICE GAMBOA[1,a], PIERRE GREMAUD[2,b], THIERRY KLEIN[3,c] and AGNÈS LAGNOUX[4,d]

[1]*Institut de Mathématiques de Toulouse and ANITI; UMR5219, Université de Toulouse; CNRS, UT3, F-31062 Toulouse, France.* [a]*fabrice.gamboa@math.univ-toulouse.fr*
[2]*Department of Mathematics, NC State University, Raleigh, North Carolina 27695, USA.* [b]*gremaud@ncsu.edu*
[3]*Institut de Mathématiques de Toulouse; UMR5219, Université de Toulouse; ENAC - Ecole Nationale de l'Aviation Civile, Université de Toulouse, France.* [c]*thierry01.klein@enac.fr*
[4]*Institut de Mathématiques de Toulouse; UMR5219, Université de Toulouse; CNRS, UT2J, F-31058 Toulouse, France.* [d]*lagnoux@univ-tlse2.fr*

We propose a new statistical estimation framework for a large family of global sensitivity analysis indices. Our approach is based on rank statistics and uses an empirical correlation coefficient recently introduced by Chatterjee (*Calcutta Statist. Assoc. Bull.* **33** (1984) 1–2). We show how to apply this approach to compute not only the Cramér-von-Mises indices, directly related to Chatterjee's notion of correlation, but also first-order Sobol' indices, general metric space indices and higher-order moment indices. We establish consistency of the resulting estimators and demonstrate their numerical efficiency, especially for small sample sizes. In addition, we prove a central limit theorem for the estimators of the first-order Sobol' indices.

Keywords: Global sensitivity analysis; Cramér-von-Mises distance; Pick-Freeze method; Chatterjee's coefficient of correlation; Sobol' indices estimation

References

[1] Antoniadis, A. (1984). Analysis of variance on function spaces. *Math. Operationsforsch. Statist. Ser. Statist.* **15** 59–71. MR0729613 https://doi.org/10.1080/02331888408801747
[2] Auddy, A., Deb, N. and Nandy, S. (2021). Exact detection thresholds for chatterjee's correlation. ArXiv preprint. Available at arXiv:2104.15140.
[3] Azadkia, M. and Chatterjee, S. (2019). A simple measure of conditional dependence. ArXiv preprint. Available at arXiv:1910.12327.
[4] Borgonovo, E. (2007). A new uncertainty importance measure. *Reliab. Eng. Syst. Saf.* **92** 771–784.
[5] Borgonovo, E., Castaings, W. and Tarantola, S. (2011). Moment independent importance measures: New results and analytical test cases. *Risk Analysis* **31** 404–428.
[6] Boucheron, S., Lugosi, G. and Massart, P. (2009). On concentration of self-bounding functions. *Electron. J. Probab.* **14** 1884–1899. MR2540852 https://doi.org/10.1214/EJP.v14-690
[7] 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
[8] Broto, B., Bachoc, F. and Depecker, M. (2020). Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. *SIAM/ASA J. Uncertain. Quantificat.* **8** 693–716. MR4096129 https://doi.org/10.1137/18M1234631
[9] Chatterjee, S.K. (1984). Professor K. B. Madhava: 1895–1984. *Calcutta Statist. Assoc. Bull.* **33** 1–2. MR0772649 https://doi.org/10.1177/0008068319840101
[10] Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *J. Stat. Comput. Simul.* **85** 1283–1305. MR3306798 https://doi.org/10.1080/00949655.2014.945932

[11] Da Veiga, S. and Gamboa, F. (2013). Efficient estimation of sensitivity indices. *J. Nonparametr. Stat.* **25** 573–595. MR3174285 https://doi.org/10.1080/10485252.2013.784762
[12] De Rocquigny, E., Devictor, N. and Tarantola, S. (2008). *Uncertainty in Industrial Practice*. Wiley Online Library.
[13] Dette, H., Siburg, K.F. and Stoimenov, P.A. (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.* **40** 21–41. MR3024030 https://doi.org/10.1111/j.1467-9469.2011.00767.x
[14] Fort, J.-C., Klein, T. and Lagnoux, A. (2021). Global sensitivity analysis and Wasserstein spaces. *SIAM/ASA J. Uncertain. Quantificat.* **9** 880–921. MR4276978 https://doi.org/10.1137/20M1354957
[15] Fort, J.-C., Klein, T. and Rachdi, N. (2016). New sensitivity analysis subordinated to a contrast. *Comm. Statist. Theory Methods* **45** 4349–4364. MR3509524 https://doi.org/10.1080/03610926.2014.901369
[16] Fraiman, R., Gamboa, F. and Moreno, L. (2020). Sensitivity indices for output on a Riemannian manifold. *Int. J. Uncertain. Quantif.* **10** 297–314. MR4200147 https://doi.org/10.1615/Int.J.UncertaintyQuantification.2020029614
[17] Gamboa, F., Gremaud, P., Klein, T. and Lagnoux, A. (2022). Supplement to "Global Sensitivity Analysis: a novel generation of mighty estimators based on rank statistics." https://doi.org/10.3150/21-BEJ1421SUPP
[18] Gamboa, F., Janon, A., Klein, T., Lagnoux, A. and Prieur, C. (2016). Statistical inference for Sobol pick-freeze Monte Carlo method. *Statistics* **50** 881–902. MR3508283 https://doi.org/10.1080/02331888.2015.1105803
[19] Gamboa, F., Klein, T. and Lagnoux, A. (2018). Sensitivity analysis based on Cramér–von Mises distance. *SIAM/ASA J. Uncertain. Quantificat.* **6** 522–548. MR3784121 https://doi.org/10.1137/15M1025621
[20] Gamboa, F., Klein, T. and Lagnoux, A. (2021). A central limit theorem for generalized L-statistics. Preprint.
[21] Gamboa, F., Klein, T., Lagnoux, A. and Moreno, L. (2021). Sensitivity analysis in general metric spaces. *Reliab. Eng. Syst. Saf.* **212** 107611.
[22] Gamboa, F., Klein, T., Lagnoux, A., Prieur, C. and da Veiga, S. (2021). New estimations of sensitivity indices using kernels. Work in progress.
[23] Hart, J.L., Gremaud, P.A. and David, T. (2019). Global sensitivity analysis of high-dimensional neuroscience models: An example of neurovascular coupling. *Bull. Math. Biol.* **81** 1805–1828. MR3945911 https://doi.org/10.1007/s11538-019-00578-0
[24] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196
[25] Janon, A., Klein, T., Lagnoux, A., Nodet, M. and Prieur, C. (2014). Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM Probab. Stat.* **18** 342–364. MR3333994 https://doi.org/10.1051/ps/2013040
[26] Kucherenko, S. and Song, S. (2017). Different numerical estimators for main effect global sensitivity indices. *Reliab. Eng. Syst. Saf.* **165** 222–238.
[27] McDiarmid, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics, 1989 (Norwich, 1989). London Mathematical Society Lecture Note Series* **141** 148–188. Cambridge: Cambridge Univ. Press. MR1036755
[28] Orey, S. (1958). A central limit theorem for m-dependent random variables. *Duke Math. J.* **25** 543–546. MR0097841
[29] Owen, A.B. (2013). Better estimation of small Sobol' sensitivity indices. *ACM Trans. Model. Comput. Simul.* **23** Art. 11, 17. MR3065606 https://doi.org/10.1145/2457459.2457460
[30] Owen, A.B. (2013). Variance components and generalized Sobol' indices. *SIAM/ASA J. Uncertain. Quantificat.* **1** 19–41. MR3283876 https://doi.org/10.1137/120876782
[31] Owen, A.B., Dick, J. and Chen, S. (2014). Higher order Sobol' indices. *Inf. Inference* **3** 59–81. MR3311449 https://doi.org/10.1093/imaiai/iau001
[32] Saltelli, A., Chan, K. and Scott, E.M., eds. (2000). *Sensitivity Analysis*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR1886391
[33] Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*. *Springer Series in Statistics*. New York: Springer. MR2160708 https://doi.org/10.1007/978-1-4757-3799-8
[34] Shi, H., Drton, M. and Han, F. (2020). On the power of chatterjee rank correlation.
[35] Sobol, I.M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment* **1** 407–414. MR1335161

[36] Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates *Math. Comput. Simulation* **55** 271–280. MR1823119 https://doi.org/10.1016/S0378-4754(00)00270-6

[37] Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.* **93** 964–979.

[38] Trutschnig, W. (2011). On a strong metric on the space of copulas and its induced dependence measure. *J. Math. Anal. Appl.* **384** 690–705. MR2825218 https://doi.org/10.1016/j.jmaa.2011.06.013

[39] 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

Bernoulli **28**(4), 2022, 2375–2403
https://doi.org/10.3150/21-BEJ1422

Asymptotics of AIC, BIC and C_p model selection rules in high-dimensional regression

ZHIDONG BAI[1,a], KWOK PUI CHOI[2,c], YASUNORI FUJIKOSHI[3,d] and JIANG HU[1,b]

[1]*KLASMOE and School of Mathematics & Statistics, Northeast Normal University, China.* [a]*baizd@nenu.edu.cn,* [b]*huj156@nenu.edu.cn*
[2]*Department of Statistics and Data Science, National University of Singapore, Singapore.* [c]*stackp@nus.edu.sg*
[3]*Department of Mathematics, Graduate School of Science, Hiroshima University, Japan.*
[d]*fujikoshi_y@yahoo.co.jp*

Variable selection in multivariate linear regression is essential for the interpretation, subsequent statistical inferences and predictions of the statistical problem at hand. It has a long history of being studied, and many regressor selection criteria have been proposed. Most commonly used criteria include the Akaike information criterion (AIC), Bayesian information criterion (BIC), and Mallow's C_p and their modifications. It is well-known that if the true model is among the candidate models, then BIC is strongly consistent while AIC is not when only the sample size tends to infinity and the numbers of response variables and regressors remain fixed; a setting often described as large-sample. Increasingly, more and more datasets are viewed as high-dimensional in the sense that the number of response variables (p), the number of regressors (k) and the sample size (n) tend to infinity such that $p/n \to c \in (0,1)$ and $k/n \to \alpha \in [0,1)$ with $\alpha + c < 1$. A few recent works reported that, under high dimension, the asymptotic properties of AIC, BIC and C_p selection rules in the large-sample setting do not necessarily carry over in the high-dimensional setting. In this paper, we clarify their asymptotic properties and provide sufficient conditions for which a selection rule is strongly consistent, almost surely under specify and over specify a true model. We do not assume normality in the errors, and we only require finite fourth moment. The main tool employed is random matrix theory techniques. A consequence of this work states that, under certain mild high-dimensional conditions, if the BIC selection rule is strongly consistent then the AIC selection rule is also strongly consistent, but not vice versa. This result is in stark contrast to the large-sample result.

Keywords: AIC; BIC; C_p; strong consistency; high-dimensional criteria; multi-response regression; variable selection; RMT

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)* 267–281. MR0483125

Anzanello, M.J. and Fogliatto, F.S. (2014). A review of recent variable selection methods in industrial and chemometrics applications. *European J. of Industrial Engineering* **8** 619.

Bai, Z.D., Choi, K.P. and Fujikoshi, Y. (2018). Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis. *Ann. Statist.* **46** 1050–1076. MR3797996 https://doi.org/10.1214/17-AOS1577

Bai, Z.D., Miao, B. and Pan, G. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35** 1532–1572. MR2330979 https://doi.org/10.1214/009117906000001079

Bai, Z.D. 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

Bai, Z.D., Choi, K.P., Fujikoshi, Y. and Hu, J. (2022). Supplement to "Asymptotics of AIC, BIC and C_p model selection rules in high-dimensional regression." https://doi.org/10.3150/21-BEJ1422SUPP

Bao, Z., Hu, J., Pan, G. and Zhou, W. (2019). Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.* **47** 612–640. MR3909944 https://doi.org/10.1214/18-AOS1704

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **52** 345–370. MR0914460 https://doi.org/10.1007/BF02294361

Chen, L.S., Paul, D., Prentice, R.L. and Wang, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *J. Amer. Statist. Assoc.* **106** 1345–1360. MR2896840 https://doi.org/10.1198/jasa.2011.ap10599

Enomoto, R., Sakurai, T. and Fujikoshi, Y. (2015). Consistency properties of AIC, BIC, Cp and their modifications in the growth curve model under a large-(q,n) framework. *SUT J. Math.* **51** 59–81. MR3409058

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

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica* **20** 101–148. MR2640659

Fan, Y. and Tang, C.Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 531–552. MR3065478 https://doi.org/10.1111/rssb.12001

Fujikoshi, Y. (1983). A criterion for variable selection in multiple discriminant analysis. *Hiroshima Math. J.* **13** 203–214. MR0693557

Fujikoshi, Y. (1985). Selection of variables in two-group discriminant analysis by error rate and Akaike's information criteria. *J. Multivariate Anal.* **17** 27–37. MR0797518 https://doi.org/10.1016/0047-259X(85)90092-2

Fujikoshi, Y., Enomoto, R. and Sakurai, T. (2013). High-dimensional AIC in the growth curve model. *J. Multivariate Anal.* **122** 239–250. MR3189321 https://doi.org/10.1016/j.jmva.2013.07.006

Fujikoshi, Y. and Sakurai, T. (2016a). High-dimensional consistency of rank estimation criteria in multivariate linear model. *J. Multivariate Anal.* **149** 199–212. MR3507324 https://doi.org/10.1016/j.jmva.2016.04.005

Fujikoshi, Y. and Sakurai, T. (2016b). Some properties of estimation criteria for dimensionality in principal component analysis. *Amer. J. Math. Management Sci.* **35** 133–142.

Fujikoshi, Y., Sakurai, T. and Yanagihara, H. (2014). Consistency of high-dimensional AIC-type and C_p-type criteria in multivariate linear regression. *J. Multivariate Anal.* **123** 184–200. MR3130429 https://doi.org/10.1016/j.jmva.2013.09.006

Fujikoshi, Y. and Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression. *Biometrika* **84** 707–716. MR1603952 https://doi.org/10.1093/biomet/84.3.707

Fujikoshi, Y. and Veitch, L.G. (1979). Estimation of dimensionality in canonical correlation analysis. *Biometrika* **66** 345–351. MR0548204 https://doi.org/10.1093/biomet/66.2.345

Heinze, G., Wallisch, C. and Dunkler, D. (2018). Variable selection—a review and recommendations for the practicing statistician. *Biom. J.* **60** 431–449. MR3802914 https://doi.org/10.1002/bimj.201700067

Kong, Y., Li, D., Fan, Y. and Lv, J. (2017). Interaction pursuit in high-dimensional multi-response regression via distance correlation. *Ann. Statist.* **45** 897–922. MR3650404 https://doi.org/10.1214/16-AOS1474

Li, Y., Nan, B. and Zhu, J. (2015). Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure. *Biometrics* **71** 354–363. MR3366240 https://doi.org/10.1111/biom.12292

Li, R., Zhong, W. and Zhu, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. MR3010900 https://doi.org/10.1080/01621459.2012.695654

Mallows, C.L. (1973). Some comments on Cp. *Technometrics* **15** 661.

Nishii, R., Bai, Z.D. and Krishnaiah, P.R. (1988). Strong consistency of the information criterion for model selection in multivariate analysis. *Hiroshima Math. J.* **18** 451–462. MR0991240

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

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

Silverstein, J.W. and Choi, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. MR1345541 https://doi.org/10.1006/jmva.1995.1058

Sparks, R.S., Coutsourides, D. and Troskie, L. (1983). The multivariate C_p. *Comm. Statist. Theory Methods* **12** 1775–1793. MR0704853 https://doi.org/10.1080/03610928308828569

Yamamura, M., Yanagihara, H. and Srivastava, M.S. (2010). Variable selection in multivariate linear regression models with fewer observations than the dimension. *Japanese Journal of Applied Statistics* **39** 1–19.

Yanagihara, H. (2015). Conditions for consistency of a log-likelihood-based information criterion in normal multivariate linear regression models under the violation of the normality assumption. *J. Japan Statist. Soc.* **45** 21–56. MR3444402 https://doi.org/10.14490/jjss.45.21

Yanagihara, H., Wakaki, H. and Fujikoshi, Y. (2015). A consistency property of the AIC for multivariate linear models when the dimension and the sample size are large. *Electron. J. Stat.* **9** 869–897. MR3338666 https://doi.org/10.1214/15-EJS1022

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x

Bernoulli **28**(4), 2022, 2404–2430
https://doi.org/10.3150/21-BEJ1423

Adaptive schemes for piecewise deterministic Monte Carlo algorithms

ANDREA BERTAZZI[a] and JORIS BIERKENS[b]

Delft Institute of Applied Mathematics, TU Delft, Mekelweg 4, 2628 CD Delft, The Netherlands.
[a]*a.bertazzi@tudelft.nl,* [b]*joris.bierkens@tudelft.nl*

The Bouncy Particle sampler (BPS) and the Zig-Zag sampler (ZZS) are continuous time, non-reversible Monte Carlo methods based on piecewise deterministic Markov processes. Experiments show that the speed of convergence of these samplers can be affected by the shape of the target distribution, as for instance in the case of anisotropic targets. We propose an adaptive scheme that iteratively learns all or part of the covariance matrix of the target and takes advantage of the obtained information to modify the underlying process with the aim of increasing the speed of convergence. Moreover, we define an adaptive scheme that automatically tunes the refreshment rate of the BPS or ZZS. We prove ergodicity and a law of large numbers for all the proposed adaptive algorithms. Finally, we show the benefits of the adaptive samplers with several numerical simulations.

Keywords: Adaptive Markov process Monte Carlo; piecewise deterministic Markov processes; bouncy particle sampler; zig-zag sampler; ergodicity

References

[1] Andrieu, C., Durmus, A., Nüsken, N. and Roussel, J. (2021). Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* **31** 2478–2517. MR4332703 https://doi.org/10.1214/20-aap1653
[2] Andrieu, C. and Livingstone, S. (2021). Peskun-Tierney ordering for Markovian Monte Carlo: Beyond the reversible scenario. *Ann. Statist.* **49** 1958–1981. MR4319237 https://doi.org/10.1214/20-aos2008
[3] 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
[4] Bai, Y., Roberts, G.O. and Rosenthal, J.S. (2011). On the containment condition for adaptive Markov chain Monte Carlo algorithms. *Adv. Appl. Stat.* **21** 1–54. MR2849670
[5] Bardenet, R., Doucet, A. and Holmes, C. (2017). On Markov chain Monte Carlo methods for tall data. *J. Mach. Learn. Res.* **18** 47. MR3670492
[6] Bertazzi, A. Bierkens, J. (2022). Supplement to “Adaptive schemes for piecewise deterministic Monte Carlo algorithms.” https://doi.org/10.3150/21-BEJ1423SUPP
[7] Bierkens, J. and Duncan, A. (2017). Limit theorems for the zig-zag process. *Adv. in Appl. Probab.* **49** 791–825. MR3694318 https://doi.org/10.1017/apr.2017.22
[8] Bierkens, J., Fearnhead, P. and Roberts, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715
[9] Bierkens, J., Grazzi, S., Kamatani, K. and Roberts, G. (2020). The boomerang sampler. In *Proceedings of the 37th International Conference on Machine Learning* (H.D. III and A. Singh, eds.). *Proceedings of Machine Learning Research* **119** 908–918. PMLR.
[10] Bierkens, J., Kamatani, K. and Roberts, G.O. (2018). High-dimensional scaling limits of piecewise deterministic sampling algorithms. Available at arXiv:1807.11358.
[11] Bierkens, J., Roberts, G.O. and Zitt, P.-A. (2019). Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** 2266–2301. MR3983339 https://doi.org/10.1214/18-AAP1453
[12] Bouchard-Côté, A., Vollmer, S.J. and Doucet, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

[13] Chimisov, C., Latuszynski, K. and Roberts, G. (2018). Adapting The Gibbs Sampler. Available at arXiv:1801.09299.

[14] 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

[15] Davis, M.H.A. (1993). *Markov Models and Optimization*. *Monographs on Statistics and Applied Probability* **49**. London: CRC Press. MR1283589 https://doi.org/10.1007/978-1-4899-4483-2

[16] Deligiannidis, G., Bouchard-Côté, A. and Doucet, A. (2019). Exponential ergodicity of the bouncy particle sampler. *Ann. Statist.* **47** 1268–1287. MR3911112 https://doi.org/10.1214/18-AOS1714

[17] Durmus, A., Guillin, A. and Monmarché, P. (2020). Geometric ergodicity of the bouncy particle sampler. *Ann. Appl. Probab.* **30** 2069–2098. MR4149523 https://doi.org/10.1214/19-AAP1552

[18] Fearnhead, P., Bierkens, J., Pollock, M. and Roberts, G.O. (2018). Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.* **33** 386–412. MR3843382 https://doi.org/10.1214/18-STS648

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

[20] Fort, G., Moulines, E., Priouret, P. and Vandekerkhove, P. (2014). A central limit theorem for adaptive and interacting Markov chains. *Bernoulli* **20** 457–485. MR3178506 https://doi.org/10.3150/12-BEJ493

[21] 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

[22] Johndrow, J.E., Pillai, N.S. and Smith, A. (2020). No free lunch for approximate MCMC. ArXiv preprint. Available at arXiv:2010.12514.

[23] Lu, J. and Wang, L. (2020). On explicit L^2-convergence rate estimate for piecewise deterministic Markov processes. Available at arXiv:2007.14927.

[24] Michel, M., Kapfer, S.C. and Krauth, W. (2014). Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *J. Chem. Phys.* **140** 054116. https://doi.org/10.1063/1.4863991

[25] Monmarché, P. (2016). Piecewise deterministic simulated annealing. *ALEA Lat. Am. J. Probab. Math. Stat.* **13** 357–398. MR3487077

[26] Neal, R.M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. Boca Raton, FL: CRC Press. MR2858447

[27] Peters, E.A.J.F. and De With, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E, Stat. Nonlin. Soft Matter Phys.* **85** 1–5. https://doi.org/10.1103/PhysRevE.85.026703

[28] Pompe, E., Holmes, C. and Łatuszyński, K. (2020). A framework for adaptive MCMC targeting multimodal distributions. *Ann. Statist.* **48** 2930–2952. MR4152629 https://doi.org/10.1214/19-AOS1916

[29] 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

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

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

[32] 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

[33] Vanetti, P., Bouchard-Côté, A., Deligiannidis, G. and Doucet, A. (2017). Piecewise-Deterministic Markov Chain Monte Carlo. Available at arXiv:1707.05296.

[34] Vialaret, M. and Maire, F. (2020). On the convergence time of some non-reversible Markov chain Monte Carlo methods. *Methodol. Comput. Appl. Probab.* **22** 1349–1387. MR4129137 https://doi.org/10.1007/s11009-019-09766-w

[35] Wallin, J. and Bolin, D. (2018). Efficient adaptive MCMC through precision estimation. *J. Comput. Graph. Statist.* **27** 887–897. MR3890878 https://doi.org/10.1080/10618600.2018.1459303

Bernoulli **28**(4), 2022, 2431–2458
https://doi.org/10.3150/21-BEJ1424

Convergence rates of Gibbs measures with degenerate minimum

PIERRE BRAS[a]

Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, case 188, Sorbonne Université, 4 pl. Jussieu, F-75252 Paris Cedex 5, France. [a]*pierre.bras@sorbonne-universite.fr*

We study convergence rates of Gibbs measures, with density proportional to $\exp(-f(x)/t)$, as $t \to 0$ where $f : \mathbb{R}^d \to \mathbb{R}$ admits a unique global minimum at $x^\star$. We focus on the case where the Hessian is not definite at $x^\star$. We assume instead that the minimum is strictly polynomial and give a higher order nested expansion of f at $x^\star$, which depends on every coordinate. We give an algorithm yielding such an expansion if the polynomial order of $x^\star$ is no more than 8, in connection with Hilbert's 17^{th} problem. However, we prove that the case where the order is 10 or higher is fundamentally different and that further assumptions are needed. We then give the rate of convergence of Gibbs measures using this expansion. Finally we adapt our results to the multiple well case.

Keywords: Gibbs measures; stochastic optimization; central limit theorem; Hilbert's 17^{th} problem

References

[1] Athreya, K.B. and Hwang, C.-R. (2010). Gibbs measures asymptotics. *Sankhya A* **72** 191–207. MR2658170 https://doi.org/10.1007/s13171-010-0006-5
[2] Bajbar, T. and Stein, O. (2015). Coercive polynomials and their Newton polytopes. *SIAM J. Optim.* **25** 1542–1570. MR3376789 https://doi.org/10.1137/140980624
[3] Bajbar, T. and Stein, O. (2019). Coercive polynomials: Stability, order of growth, and Newton polytopes. *Optimization* **68** 99–124. MR3902159 https://doi.org/10.1080/02331934.2018.1426585
[4] Barrera, G. (2020). Limit behavior of the invariant measure for Langevin dynamics. arXiv:2006.06808.
[5] Bras, P. and Pagès, G. (2021). Convergence of Langevin-Simulated Annealing algorithms with multiplicative noise. arXiv:2109.11669.
[6] Comon, P., Golub, G., Lim, L.-H. and Mourrain, B. (2008). Symmetric tensors and symmetric tensor rank. *SIAM J. Matrix Anal. Appl.* **30** 1254–1279. MR2447451 https://doi.org/10.1137/060661569
[7] 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
[8] Dauphin, Y.N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S. and Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In *Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2. NIPS'14* 2933–2941. Cambridge, MA, USA: MIT Press.
[9] Fort, J.-C. and Pagès, G. (1999). Asymptotic behavior of a Markovian stochastic algorithm with constant step. *SIAM J. Control Optim.* **37** 1456–1482. MR1710228 https://doi.org/10.1137/S0363012997328610
[10] Gelfand, S.B. and Mitter, S.K. (1991). Recursive stochastic algorithms for global optimization in $\mathbf{R}^d$. *SIAM J. Control Optim.* **29** 999–1018. MR1110084 https://doi.org/10.1137/0329055
[11] Hilbert, D. (1888). Ueber die Darstellung definiter Formen als Summe von Formenquadraten. *Math. Ann.* **32** 342–350. MR1510517 https://doi.org/10.1007/BF01443605
[12] Hwang, C.-R. (1980). Laplace's method revisited: Weak convergence of probability measures. *Ann. Probab.* **8** 1177–1182. MR0602391
[13] Hwang, C.R. (1981). A generalization of Laplace's method. *Proc. Amer. Math. Soc.* **82** 446–451. MR0612737 https://doi.org/10.2307/2043958

1350-7265

[14] Lazarev, V.A. (1992). Convergence of stochastic approximation procedures in the case of several roots of a regression equation. *Problemy Peredachi Informatsii* **28** 75–88. MR1163142

[15] Li, C., Chen, C., Carlson, D. and Carin, L. (2016). Preconditioned stochastic gradient Langevin dynamics for deep neural networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. *AAAI'16* 1788–1794. AAAI Press.

[16] Motzkin, T.S. (1967). The arithmetic-geometric inequality. In *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)* 205–224. New York: Academic Press. MR0223521

[17] Sagun, L., Bottou, L. and LeCun, Y. (2016). Eigenvalues of the Hessian in deep learning: Singularity and beyond. arXiv:1611.07476.

[18] Sagun, L., Evci, U., Ugur Guney, V., Dauphin, Y. and Bottou, L. (2017). Empirical analysis of the Hessian of over-parametrized neural networks. arXiv:1706.04454.

[19] van Laarhoven, P.J.M. and Aarts, E.H.L. (1987). *Simulated Annealing: Theory and Applications*. *Mathematics and Its Applications* **37**. Dordrecht: D. Reidel Publishing Co. MR0904050 https://doi.org/10.1007/978-94-015-7744-1

[20] Welling, M. and Teh, Y.W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*. *ICML'11* 681–688. Omnipress.

Bernoulli **28**(4), 2022, 2459–2491
https://doi.org/10.3150/21-BEJ1425

Sharp detection boundaries on testing dense subhypergraph

MINGAO YUAN[1,a] and ZUOFENG SHANG[2,b]

[1]*Department of Statistics, North Dakota State University, 1340 Administration Ave, Fargo, ND 58105, USA.* [a]*mingao.yuan@ndsu.edu*
[2]*Department of Mathematical Sciences, New Jersey Institute of Technology, 323 Dr Martin Luther King Jr Blvd, Newark, NJ 07102, USA.* [b]*zshang@njit.edu*

We study the problem of testing the existence of a dense subhypergraph. The null hypothesis corresponds to an Erdős-Rényi uniform random hypergraph and the alternative hypothesis corresponds to a uniform random hypergraph that contains a dense subhypergraph. We establish sharp detection boundaries in both scenarios: (1) the edge probabilities are known; (2) the edge probabilities are unknown. In both scenarios, sharp detection boundaries are characterized by the appropriate model parameters. Asymptotically powerful tests are provided when the model parameters fall in the detectable regions. Our results indicate that the detectable regions for general hypergraph models are dramatically different from their graph counterparts.

Keywords: Sharp detection boundary; uniform hypergraph; dense subhypergraph detection; asymptotically powerful test

References

[1] Ahn, K., Lee, K. and Suh, C. (2016). *Community Recovery in Hypergraphs, 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. https://doi.org/10.1109/ALLERTON.2016.7852294
[2] Ahn, K., Lee, K. and Suh, C. (2018). Hypergraph spectral clustering in the weighted stochastic block model. *IEEE J. Sel. Top. Signal Process.* **12**.
[3] Angelini, M., Caltagirone, F., Krzakala, F. and Zdeborova, L. (2015). Spectral detection on sparse hypergraphs. In *Allerton Conference on Communication, Control, and Computing* 66–73.
[4] Arias-Castro, E. and Verzelen, N. (2014). Community detection in dense random networks. *Ann. Statist.* **42** 940–969. MR3210992 https://doi.org/10.1214/14-AOS1208
[5] Bach, F., Ahipasaoglu and D'Aspremont, A. (2010). Convex relaxations for subset selection. https://arxiv.org/pdf/1006.3601.pdf.
[6] Berthet, Q. and Rigollet, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127
[7] Bhangale, A., Gandhi, R. and Kortsarz, G. (2017). Improved approximation algorithm for the dense-3-subhypergraph problem. https://arxiv.org/abs/1704.08620.
[8] Bi, X., Tang, X., Yuan, Y., Zhang, Y. and Qu, A. (2021). Tensors in statistics. *Annu. Rev. Stat. Appl.* **8** 345–368. MR4243551 https://doi.org/10.1146/annurev-statistics-042720-020816
[9] Bolla, M. (1993). Spectra, Euclidean representations and clusterings of hypergraphs. *Discrete Math.* **117** 19–39. MR1226129 https://doi.org/10.1016/0012-365X(93)90322-K
[10] Bollobás, B. and Erdős, P. (1976). Cliques in random graphs. *Math. Proc. Cambridge Philos. Soc.* **80** 419–427. MR0498256 https://doi.org/10.1017/S0305004100053056
[11] Bulo, S. and Pelillo, M. (2009). New bounds on the clique number of graphs based on spectral hypergraph theory. In *Learning and Intelligent Optimization. LION 2009* 45–58.
[12] Butucea, C. and Ingster, Y.I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

1350-7265

[13] Chertok, M. and Keller, Y. (2010). Efficient high order matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 2205–2215.

[14] Chlamtáč, E., Dinitz, M., Konrad, C., Kortsarz, G. and Rabanca, G. (2018). The densest k-subhypergraph problem. *SIAM J. Discrete Math.* **32** 1458–1477. MR3819151 https://doi.org/10.1137/16M1096402

[15] d'Aspremont, A., El Ghaoui, L., Jordan, M.I. and Lanckriet, G.R.G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49** 434–448. MR2353806 https://doi.org/10.1137/050645506

[16] Florescu, L. and Perkins, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *29th Annual Conference on Learning Theory* **49** 943–959.

[17] Gao, C. and Lafferty, J. (2017). Testing Network Structure Using Relations Between Small Subgraph Probabilities. Preprint.

[18] Ghoshdastidar, D. (2016). Consistency of spectral algorithms for hypergraphs under planted partition model PhD Thesis.

[19] Ghoshdastidar, D. and Dukkipati, A. (2014). Consistency of spectral partitioning of uniform hypergraphs under planted partition model. *Adv. Neural Inf. Process. Syst.* 397–405.

[20] Ghoshdastidar, D. and Dukkipati, A. (2017). Consistency of spectral hypergraph partitioning under planted partition model. *Ann. Statist.* **45** 289–315. MR3611493 https://doi.org/10.1214/16-AOS1453

[21] Hall, P. and Heyde, C.C. (2014). *Martingale Limit Theory and Its Application*. New York–London: Academic Press. MR0624435

[22] Hu, S., Wu, X. and Chan, T.-H. (2017). Maintaining densest subsets efficiently in evolving hypergraphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* 929–938. Singapore.

[23] Jin, J., Ke, Z.T. and Luo, S. (2019). Optimal adaptivity of signed-polygon statistics for network testing. *Ann. Statist.* To appear.

[24] Ke, Z., Shi, F. and Xia, D. (2020). Community Detection for Hypergraph Networks via Regularized Tensor Power Iteration. https://arxiv.org/pdf/1909.06503.pdf.

[25] Kim, C., Bandeira, A. and Goemans, M. (2017). Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In *2017 International Conference on Sampling Theory and Applications (SampTA)* 124–128.

[26] Kim, S. (2011). Higher-order correlation clustering for image segmentation. In: *Advances in Neural Information Processing Systems* 1530–8.

[27] Lin, C., Chien, I. and Wang, I. (2017). On the fundamental statistical limit of community detection in random hypergraphs. In *Information Theory (ISIT), 2017 IEEE International Symposium* 2178–2182.

[28] Liu, H., Jan, L. and Yan, S. (2015). Dense subgraph partition of positive hypergraphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 541–554.

[29] Luo, Y. and Zhang, A. (2020). Open problem: Average-case hardness of hypergraphic planted clique detection. *Proc. Mach. Learn. Res.* **125** 3852–3856.

[30] Luo, Y. and Zhang, A. (2020). Tensor clustering with planted structures: Statistical optimality and computational limits. https://arxiv.org/pdf/2005.10743.pdf.

[31] Peng, X. (2016). The Ramsey number of generalized loose paths in hypergraphs. *Discrete Math.* **339** 539–546. MR3431366 https://doi.org/10.1016/j.disc.2015.09.023

[32] Rodríguez, J.A. (2009). Laplacian eigenvalues and partition problems in hypergraphs. *Appl. Math. Lett.* **22** 916–921. MR2523606 https://doi.org/10.1016/j.aml.2008.07.020

[33] Rota Bulo, S. and Pelillo, M. (2013). A game-theoretic approach to hypergraph clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1312–1327.

[34] Verzelen, N. and Arias-Castro, E. (2015). Community detection in sparse random networks. *Ann. Appl. Probab.* **25** 3465–3510. MR3404642 https://doi.org/10.1214/14-AAP1080

[35] Yuan, M., Liu, R., Feng, Y. and Shang, Z. (2018). Testing community structures for hypergraphs. *Ann. Statist.* To appear.

[36] Yuan, M. and Nan, Y. (2021). Test dense subgraphs in sparse uniform hypergraph. *Comm. Statist. Theory Methods* **50** 4743–4762. MR4308707 https://doi.org/10.1080/03610926.2020.1723637

[37] Yuan, M., Shang, Z. (2022). Supplement to "Sharp Detection Boundaries on Testing Dense Subhypergraph." https://doi.org/10.3150/21-BEJ1425SUPP

[38] Zhou, D., Huang, J. and Scholkopf, B. (2006). Learning with hyper graphs: Clustering, classification, and embedding. *Adv. Neural Inf. Process. Syst.* **6** 1601–8.

Bernoulli **28**(4), 2022, 2492–2517
https://doi.org/10.3150/21-BEJ1426

Smoothing and adaptation of shifted Pólya tree ensembles

THIBAULT RANDRIANARISOA[a]

LPSM, Sorbonne Université, Paris, France. [a]*thibault.randrianarisoa@sorbonne-universite.fr*

Recently, S. Arlot and R. Genuer have shown that a random forest model outperforms its single-tree counterpart in estimating α-Hölder functions, $1 \leq \alpha \leq 2$. This backs up the idea that ensembles of tree estimators are smoother estimators than single trees. On the other hand, most positive optimality results on Bayesian tree-based methods assume that $\alpha \leq 1$. Naturally, one wonders whether Bayesian counterparts of forest estimators are optimal on smoother classes, just as observed with frequentist estimators for $\alpha \leq 2$. We focus on density estimation and introduce an ensemble estimator from the classical (truncated) Pólya tree construction in Bayesian nonparametrics. Inspired by the work mentioned above, the resulting Bayesian forest estimator is shown to lead to optimal posterior contraction rates, up to logarithmic terms, for the Hellinger and L^1 distances on probability density functions on $[0;1)$ for arbitrary Hölder regularity $\alpha > 0$. This improves upon previous results for constructions related to the Pólya tree prior, whose optimality was only proven when $\alpha \leq 1$. Also, by adding a hyperprior on the trees' depth, we obtain an adaptive version of the prior that does not require α to be specified to attain optimality.

Keywords: Bayesian nonparametrics; ensemble; forest; Pólya Tree

References

[1] Arlot, S. and Genuer, R. (2014). Analysis of purely random forests bias. ArXiv E-prints arXiv:1407.3939.
[2] Aubin, J.-P. (2000). *Applied Functional Analysis*, 2nd ed. *Pure and Applied Mathematics (New York)*. New York: Wiley Interscience. With exercises by Bernard Cornet and Jean-Michel Lasry, Translated from the French by Carole Labrousse.
[3] Balestriero, R. and Baraniuk, R. (2018). A spline theory of deep learning. In *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 374–383. PMLR.
[4] Biau, G., Devroye, L. and Lugosi, G. (2008). Consistency of random forests and other averaging classifiers. *J. Mach. Learn. Res.* **9** 2015–2033.
[5] Biau, G. and Scornet, E. (2016). A random forest guided tour. *TEST* **25** 197–227.
[6] Biau, G., Scornet, E. and Welbl, J. (2019). Neural random forests. *Sankhya A* **81** 347–386.
[7] Blanchard, G., Schäfer, C. and Rozenholc, Y. (2004). Oracle bounds and exact algorithm for dyadic classification trees. In *Learning Theory*. *Lecture Notes in Computer Science* **3120** 378–392. Springer, Berlin.
[8] Bleich, J. and Kapelner, A. (2014). Bayesian Additive Regression Trees with parametric models of heteroskedasticity. ArXiv E-prints arXiv:1402.5397.
[9] Breiman, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.
[10] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees*. *Wadsworth Statistics/Probability Series*. Belmont, CA: Wadsworth Advanced Books and Software.
[11] Bühlmann, P. and Yu, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927–961.
[12] Castillo, I. (2017). Pólya tree posterior distributions on densities. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2074–2102.
[13] Castillo, I. and Mismer, R. (2021). Spike and slab Pólya tree posterior densities: Adaptive inference. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1521–1548.
[14] Castillo, I. and Rockova, V. (2019). Uncertainty quantification for Bayesian CART. ArXiv E-prints arXiv:1910.07635.

[15] Chipman, H.A., George, E.I. and McCulloch, R.E. (1998). Bayesian CART model search. *J. Amer. Statist. Assoc.* **93** 935–948.

[16] Chipman, H.A., George, E.I. and McCulloch, R.E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298.

[17] Chipman, H.A., George, E.I., McCulloch, R.E. and Shively, T.S. (2016). High-dimensional nonparametric monotone function estimation using BART. ArXiv E-prints arXiv:1612.01619.

[18] Denil, M., Matheson, D. and Freitas, N.D. (2014). Narrowing the gap: Random forests in theory and in practice. In *Proceedings of the 31st International Conference on Machine Learning* (E.P. Xing and T. Jebara, eds.). *Proceedings of Machine Learning Research* **32** 665–673. Bejing, China: PMLR.

[19] Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998). A Bayesian CART algorithm. *Biometrika* **85** 363–377.

[20] Deshpande, S.K., Bai, R., Balocchi, C., Starling, J.E. and Weiss, J. (2020). VCBART: Bayesian trees for varying coefficients. ArXiv E-prints arXiv:2003.06416.

[21] De Boor, C. (1986). B(asic)-Spline Basics. Technical Report Wisconsin Univ. – Madison mathematics research center.

[22] Ding, J., Bashashati, A., Roth, A., Oloumi, A., Tse, K., Zeng, T., Haffari, G., Hirst, M., Marra, M.A., Condon, A., Aparicio, S. and Shah, S.P. (2012). Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* **28** 167–175.

[23] Donoho, D.L. (1997). CART and best-ortho-basis: A connection. *Ann. Statist.* **25** 1870–1911.

[24] Eckle, K. and Schmidt-Hieber, J. (2019). A comparison of deep networks with ReLU activation function and linear spline-type methods. *Neural Netw.* **110** 232–242.

[25] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

[26] Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629.

[27] Freedman, D.A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *Ann. Math. Stat.* **34** 1386–1403.

[28] Freedman, D.A. (1965). On the asymptotic behavior of Bayes estimates in the discrete case. II. *Ann. Math. Stat.* **36** 454–456.

[29] Friedman, J.H. (1991). Multivariate adaptive regression splines. *Ann. Statist.* **19** 1–141. With discussion and a rejoinder by the author.

[30] Gey, S. and Nedelec, E. (2005). Model selection for CART regression trees. *IEEE Trans. Inf. Theory* **51** 658–670.

[31] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge: Cambridge Univ. Press.

[32] He, J. and Hahn, P.R. (2020). Stochastic tree ensembles for regularized nonlinear regression. ArXiv E-prints arXiv:2002.03375.

[33] He, J., Yalov, S. and Hahn, P.R. (2018). XBART: Accelerated Bayesian Additive Regression Trees. ArXiv E-prints arXiv:1810.02215.

[34] Hill, J., Linero, A. and Murray, J. (2020). Bayesian additive regression trees: A review and look forward. *Annu. Rev. Stat. Appl.* **7** 251–278.

[35] Hill, J.L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. Supplementary material available online.

[36] Hjort, N.L. and Walker, S.G. (2009). Quantile pyramids for Bayesian nonparametrics. *Ann. Statist.* **37** 105–131.

[37] Ioannou, Y., Robertson, D., Zikic, D., Kontschieder, P., Shotton, J., Brown, M. and Criminisi, A. (2015). Decision Forests, Convolutional Networks and the Models in-Between techreport No. MSR-TR-2015-58 Microsoft Research.

[38] Ishwaran, H. and Kogalur, U.B. (2010). Consistency of random survival forests. *Statist. Probab. Lett.* **80** 1056–1064.

[39] Klusowski, J.M. (2018). Sharp analysis of a simple model for random forests. ArXiv E-prints arXiv:1805.02587.

[40] Klusowski, J.M. (2019). Analyzing CART. ArXiv E-prints arXiv:1906.10086.

[41] Kontschieder, P., Fiterau, M., Criminisi, A. and Bulò, S.R. (2015). Deep neural decision forests. In *2015 IEEE International Conference on Computer Vision (ICCV)* 1467–1475.

[42] Kruijer, W., Rousseau, J. and van der Vaart, A. (2010). Adaptive Bayesian density estimation with location-scale mixtures. *Electron. J. Stat.* **4** 1225–1257.
[43] Lakshminarayanan, B., Roy, D.M. and Whye Teh, Y. (2014). Mondrian forests: Efficient online random forests. ArXiv E-prints arXiv:1406.2673.
[44] Lamprinakou, S., McCoy, E., Barahona, M., Gandy, A., Flaxman, S. and Filippi, S. (2020). BART-based inference for Poisson processes. ArXiv E-prints arXiv:2005.07927.
[45] Lavine, M. (1992). Some aspects of Pólya tree distributions for statistical modelling. *Ann. Statist.* **20** 1222–1235.
[46] Linero, A.R. (2017). A review of tree-based Bayesian methods. *Commun. Stat. Appl. Methods* **24** 543–559.
[47] Linero, A.R. (2018). Bayesian regression trees for high-dimensional prediction and variable selection. *J. Amer. Statist. Assoc.* **113** 626–636.
[48] Linero, A.R. and Yang, Y. (2018). Bayesian regression tree ensembles that adapt to smoothness and sparsity. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 1087–1110.
[49] Liu, Y., Shao, Z. and Yuan, G.-C. (2010). Prediction of polycomb target genes in mouse embryonic stem cells. *Genomics* **96** 17–26.
[50] Mauldin, R.D., Sudderth, W.D. and Williams, S.C. (1992). Pólya trees and random distributions. *Ann. Statist.* **20** 1203–1221.
[51] Mentch, L. and Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J. Mach. Learn. Res.* **17** 26.
[52] Mourtada, J., Gaïffas, S. and Scornet, E. (2020). Minimax optimal rates for Mondrian trees and forests. *Ann. Statist.* **48** 2253–2276.
[53] Nieto-Barajas, L.E. and Müller, P. (2012). Rubbery Polya tree. *Scand. J. Stat.* **39** 166–184.
[54] O'Reilly, E. and Tran, N. (2020). Stochastic geometry to generalize the Mondrian process. ArXiv E-prints arXiv:2002.00797.
[55] Pratola, M.T. (2016). Efficient Metropolis-Hastings proposal mechanisms for Bayesian regression tree models. *Bayesian Anal.* **11** 885–911.
[56] Pratola, M.T., Chipman, H.A., Gattiker, J.R., Higdon, D.M., McCulloch, R. and Rust, W.N. (2014). Parallel Bayesian Additive Regression Trees. *J. Comput. Graph. Statist.* **23** 830–852.
[57] Pratola, M.T., Chipman, H.A., George, E.I. and McCulloch, R.E. (2020). Heteroscedastic BART via multiplicative regression trees. *J. Comput. Graph. Statist.* **29** 405–417.
[58] Randrianarisoa, T. (2022). Supplement to "Smoothing and adaptation of shifted Pólya tree ensembles." https://doi.org/10.3150/21-BEJ1426SUPP
[59] Rockova, V. (2019). On semi-parametric Bernstein-von Mises theorems for BART. ArXiv E-prints arXiv:1905.03735.
[60] Ročková, V. and Saha, E. (2019). On theory for BART. In *Proceedings of Machine Learning Research* (K. Chaudhuri and M. Sugiyama, eds.) **89** 2839–2848. PMLR.
[61] Ročková, V. and van der Pas, S. (2020). Posterior concentration for Bayesian regression trees and forests. *Ann. Statist.* **48** 2108–2131.
[62] Roy, D.M. and Teh, Y.W. (2009). The Mondrian process. In *Advances in Neural Information Processing Systems 21* (D. Koller, D. Schuurmans, Y. Bengio and L. Bottou, eds.) 1377–1384. Curran Associates.
[63] Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.* **48** 1875–1897.
[64] Scornet, E. (2016). On the asymptotics of random forests. *J. Multivariate Anal.* **146** 72–83.
[65] Scornet, E. (2016). Random forests and kernel methods. *IEEE Trans. Inf. Theory* **62** 1485–1500.
[66] Scornet, E., Biau, G. and Vert, J.-P. (2015). Consistency of random forests. *Ann. Statist.* **43** 1716–1741.
[67] Sethi, I.K. (1990). Entropy nets: From decision trees to neural networks. Technical Report Dept. of Computer Science, Wayne State Univ., Detroit, MI.
[68] Tan, Y.V. and Roy, J. (2019). Bayesian additive regression trees and the general BART model. *Stat. Med.* **38** 5048–5069.
[69] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[70] van der Pas, S. and Ročková, V. (2017). Bayesian dyadic trees and histograms for regression. In *Advances in Neural Information Processing Systems 30* (I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) 2089–2099. Curran Associates.
[71] Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242.
[72] Wager, S. and Walther, G. (2015). Adaptive concentration of regression trees, with application to random forests. ArXiv E-prints arXiv:1503.06388.
[73] Welbl, J. (2014). Casting random forests as artificial neural networks (and profiting from it). In *Pattern Recognition*. *Lecture Notes in Computer Science* **8753** 765–771. Springer, Cham.
[74] Wong, W.H. and Ma, L. (2010). Optional Pólya tree and Bayesian inference. *Ann. Statist.* **38** 1433–1459.
[75] Yang, Y., Morillo, I.G. and Hospedales, T.M. (2018). Deep neural decision trees. In *ICML Workshop on Human Interpretability in Machine Learning (WHI)*.
[76] Zhang, J.L. and Härdle, W.K. (2010). The Bayesian additive classification tree applied to credit risk modelling. *Comput. Statist. Data Anal.* **54** 1197–1205.

Bernoulli **28**(4), 2022, 2518–2545
https://doi.org/10.3150/21-BEJ1427

Asymptotically equivalent prediction in multivariate geostatistics

FRANÇOIS BACHOC[1,a], EMILIO PORCU[2,b], MORENO BEVILACQUA[3,c], REINHARD FURRER[4,d] and TARIK FAOUZI[5,e]

[1]*Institut de mathématique, UMR5219; Université de Toulouse; CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France.* [a]*francois.bachoc@math.univ-toulouse.fr*
[2]*Department of Mathematics, Khalifa University, Abu Dhabi, The United Arab Emirates; School of Computer Science and Statistics, Trinity College, Dublin.* [b]*porcue@tcd.ie*
[3]*Facultad de Ingenieria y Ciencias, Universidad Adolfo Ibáñez, Viña del Mar, Chile; ANID-Millennium Science Initiative Program-Millennium Nucleus Center for the Discovery of Structures in Complex Data.* [c]*moreno.bevilacqua@uai.cl*
[4]*Department of Mathematics and Department of Computational Science, University of Zurich, 8057, Zurich, Switzerland.* [d]*reinhard.furrer@math.uzh.ch*
[5]*Department of Statistics, Universidad del Biobio, 4081112, Conception, Chile.* [e]*tfaouzi@ubiobio.cl*

Cokriging is the common method of spatial interpolation (best linear unbiased prediction) in multivariate geostatistics. While best linear prediction has been well understood in univariate spatial statistics, the literature for the multivariate case has been elusive so far. The new challenges provided by modern spatial datasets, being typically multivariate, call for a deeper study of cokriging. In particular, we deal with the problem of misspecified cokriging prediction within the framework of fixed domain asymptotics. Specifically, we provide conditions for equivalence of measures associated with multivariate Gaussian random fields, with index set in a compact set of a d-dimensional Euclidean space. Such conditions have been elusive for over about 50 years of spatial statistics.

We then focus on the multivariate Matérn and Generalized Wendland classes of matrix valued covariance functions, that have been very popular for having parameters that are crucial to spatial interpolation, and that control the mean square differentiability of the associated Gaussian process. We provide sufficient conditions, for equivalence of Gaussian measures, relying on the covariance parameters of these two classes. This enables to identify the parameters that are crucial to asymptotically equivalent interpolation in multivariate geostatistics. Our findings are then illustrated through simulation studies.

Keywords: Cokriging; equivalence of Gaussian measures; fixed domain asymptotics; functional analysis; Generalized Wendland; Matérn; spectral analysis

References

Abramowitz, M. and Stegun, I.A., eds. (1966). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. New York: Dover. MR0208797

Alegría, A. and Porcu, E. (2017). The dimple problem related to space-time modeling under the Lagrangian framework. *J. Multivariate Anal.* **162** 110–121. MR3719338 https://doi.org/10.1016/j.jmva.2017.09.001

Alegría, A., Porcu, E., Furrer, R. and Mateu, J. (2019). Covariance functions for multivariate Gaussian fields. *Stoch. Environ. Res. Risk Assess.* **33** 1593–1608.

Anderes, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38** 870–893. MR2604700 https://doi.org/10.1214/09-AOS725

Apanasovich, T. and Genton, M.G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97** 15–30. MR2594414 https://doi.org/10.1093/biomet/asp078

Apanasovich, T.V., Genton, M.G. and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Amer. Statist. Assoc.* **107** 180–193. MR2949350 https://doi.org/10.1080/01621459.2011.643197

Askey, R. (1973). Radial characteristic functions Technical report, Research Center, Univ. Wisconsin.

Azaïs, J.-M., Bachoc, F., Lagnoux, A. and Nguyen, T.M.N. (2020). Semi-parametric estimation of the variogram scale parameter of a Gaussian process with stationary increments. *ESAIM Probab. Stat.* **24** 842–882. MR4178368 https://doi.org/10.1051/ps/2020021

Bachoc, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *J. Multivariate Anal.* **125** 1–35. MR3163828 https://doi.org/10.1016/j.jmva.2013.11.015

Bachoc, F., Porcu, E., Bevilacqua, M., Furrer, R. and Faouzi, T. (2022). Supplement to "Asymptotically equivalent prediction in multivariate geostatistics." https://doi.org/10.3150/21-BEJ1427SUPP

Bevilacqua, M., Hering, A.S. and Porcu, E. (2015). On the flexibility of multivariate covariance models: Comment on the paper by Genton and Kleiber. *Statist. Sci.* **30** 167–169. MR3353098 https://doi.org/10.1214/15-STS516

Bevilacqua, M., Faouzi, T., Furrer, R. and Porcu, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *Ann. Statist.* **47** 828–856. MR3909952 https://doi.org/10.1214/17-AOS1652

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *Ann. Math. Stat.* **33** 882–886. MR0149577 https://doi.org/10.1214/aoms/1177704456

Da Prato, G. and Zabczyk, J. (2014). *Stochastic Equations in Infinite Dimensions*, 2nd ed. *Encyclopedia of Mathematics and Its Applications* **152**. Cambridge: Cambridge Univ. Press. MR3236753 https://doi.org/10.1017/CBO9781107295513

Daley, D., Porcu, E. and Bevilacqua, M. (2015). Classes of compactly supported covariance functions for multivariate random fields. *Stoch. Environ. Res. Risk Assess.* **29** 1249–1263.

Furrer, R., Genton, M.G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523. MR2291261 https://doi.org/10.1198/106186006X132178

Genton, M.G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statist. Sci.* **30** 147–163. MR3353096 https://doi.org/10.1214/14-STS487

Gneiting, T. (2002a). Compactly supported correlation functions. *J. Multivariate Anal.* **83** 493–508. MR1945966 https://doi.org/10.1006/jmva.2001.2056

Gneiting, T. (2002b). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475 https://doi.org/10.1198/016214502760047113

Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105** 1167–1177. MR2752612 https://doi.org/10.1198/jasa.2010.tm09420

Golubov, B.I. (1981). On Abel-Poisson type and Riesz means. *Anal. Math.* **7** 161–184. MR0635483 https://doi.org/10.1007/BF01908520

Ibragimov, I.D.A. and Rozanov, Y.A. (1978). *Gaussian Random Processes*. *Applications of Mathematics* **9**. New York-Berlin: Springer. Translated from the Russian by A. B. Aries. MR0543837

Istas, J. and Lang, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. Henri Poincaré Probab. Stat.* **33** 407–436. MR1465796 https://doi.org/10.1016/S0246-0203(97)80099-4

Kaufman, C.G., Schervish, M.J. and Nychka, D.W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555. MR2504203 https://doi.org/10.1198/016214508000000959

Lang, G. and Roueff, F. (2001). Semi-parametric estimation of the Hölder exponent of a stationary Gaussian process with minimax rates. *Stat. Inference Stoch. Process.* **4** 283–306. MR1868724 https://doi.org/10.1023/A:1012227325436

Loh, W.-L. (2015). Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. *Ann. Statist.* **43** 2766–2794. MR3405611 https://doi.org/10.1214/15-AOS1365

Maniglia, S. and Rhandi, A. (2004). Gaussian measures on separable Hilbert spaces and applications. *Quad. Mat.* **2004**.

Mardia, K.V. and Marshall, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146. MR0738334 https://doi.org/10.1093/biomet/71.1.135

Porcu, E., Bevilacqua, M. and Genton, M.G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Amer. Statist. Assoc.* **111** 888–898. MR3538713 https://doi.org/10.1080/01621459.2015.1072541

Porcu, E., Gregori, P. and Mateu, J. (2006). Nonseparable stationary anisotropic space-time covariance functions. *Stoch. Environ. Res. Risk Assess.* **21** 113–122. MR2307631 https://doi.org/10.1007/s00477-006-0048-3

Porcu, E. and Zastavnyi, V. (2011). Characterization theorems for some classes of covariance functions associated to vector valued random fields. *J. Multivariate Anal.* **102** 1293–1301. MR2811618 https://doi.org/10.1016/j.jmva.2011.04.013

Putter, H. and Young, G.A. (2001). On the effect of covariance function estimation on the accuracy of Kriging predictors. *Bernoulli* **7** 421–438. MR1836738 https://doi.org/10.2307/3318494

Schaback, R. (2011). The missing Wendland functions. *Adv. Comput. Math.* **34** 67–81. MR2783302 https://doi.org/10.1007/s10444-009-9142-7

Skorohod, A.V. and Jadrenko, M.Ĭ. (1973). Absolute continuity of measures that correspond to homogeneous Gaussian fields. *Theory Probab. Appl.* **18** 27–40.

Stein, M.L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63. MR0924856 https://doi.org/10.1214/aos/1176350690

Stein, M. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Statist.* **18** 850–872. MR1056340 https://doi.org/10.1214/aos/1176347629

Stein, M.L. (1993). A simple condition for asymptotic optimality of linear predictions of random fields. *Statist. Probab. Lett.* **17** 399–404. MR1237787 https://doi.org/10.1016/0167-7152(93)90261-G

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

Stein, M.L. (1999b). Predicting random fields with increasing dense observations. *Ann. Appl. Probab.* **9** 242–273. MR1682572 https://doi.org/10.1214/aoap/1029962604

Stein, M.L. (2004). Equivalence of Gaussian measures for some nonstationary random fields. *J. Statist. Plann. Inference* **123** 1–11. MR2058118 https://doi.org/10.1016/S0378-3758(03)00144-7

Vallejos, R., Osorio, F. and Bevilacqua, M. (2020). *Spatial Relationships Between Two Georeferenced Variables—with Applications in R*. Cham: Springer. MR4248932 https://doi.org/10.1007/978-3-030-56681-4

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*, 3rd ed. New York: Springer.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **4** 389–396. MR1366510 https://doi.org/10.1007/BF02123482

Zastavnyĭ, V.P. (2006). On some properties of the Buhmann functions. *Ukraïn. Mat. Zh.* **58** 1045–1067. MR2345078 https://doi.org/10.1007/s11253-006-0128-z

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241

Zhang, H. and Cai, W. (2015). When doesn't cokriging outperform kriging? *Statist. Sci.* **30** 176–180. MR3353100 https://doi.org/10.1214/15-STS518

Bernoulli **28**(4), 2022, 2546–2562
https://doi.org/10.3150/21-BEJ1428

Phase transitions in non-linear urns with interacting types

MARCELO COSTA[1,a] and JONATHAN JORDAN[2,b]

[1] *Universidad de Buenos Aires, Buenos Aires, Argentina.* [a] *mrocha@dm.uba.ar*
[2] *School of Mathematics and Statistics, University of Sheffield, Sheffield, U.K.* [b] *jonathan.jordan@sheffield.ac.uk*

We investigate reinforced non-linear urns with interacting types, and show that where there are three interacting types there are phenomena which do not occur with two types. In a model with three types where the interactions between the types are symmetric, we show the existence of a double phase transition with three phases: as well as a phase with an almost sure limit where each of the three colours is equally represented and a phase with almost sure convergence to an asymmetric limit, which both occur with two types, there is also an intermediate phase where both symmetric and asymmetric limits are possible. In a model with anti-symmetric interactions between the types, we show the existence of a phase where the proportions of the three colours cycle and do not converge to a limit, alongside a phase where the proportions of the three colours can converge to a limit where each of the three is equally represented.

Keywords: Non-linear urn; phase transition

References

[1] Benaïm, M. (1999). Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*. *Lecture Notes in Math.* **1709** 1–68. Berlin: Springer. MR1767993 https://doi.org/10.1007/BFb0096509
[2] Benaïm, M., Benjamini, I., Chen, J. and Lima, Y. (2015). A generalized Pólya's urn with graph based interactions. *Random Structures Algorithms* **46** 614–634. MR3346459 https://doi.org/10.1002/rsa.20523
[3] Costa, M. (2018). Cooperative Models of Stochastic Growth Ph.D. thesis Durham Univ.
[4] Davis, B. (1990). Reinforced random walk. *Probab. Theory Related Fields* **84** 203–229. MR1030727 https://doi.org/10.1007/BF01197845
[5] Khanin, K. and Khanin, R. (2001). A probabilistic model for the establishment of neuron polarity. *J. Math. Biol.* **42** 26–40. MR1820779 https://doi.org/10.1007/PL00000071
[6] Laruelle, S. and Pagès, G. (2019). Nonlinear randomized urn models: A stochastic approximation viewpoint. *Electron. J. Probab.* **24** 98. MR4017116 https://doi.org/10.1214/19-ejp312
[7] Launay, M. and Limic, V. (2012). Generalized Interacting Urn Models. Available at https://arxiv.org/abs/1207.5635.
[8] Oliveira, R.I. (2009). The onset of dominance in balls-in-bins processes with feedback. *Random Structures Algorithms* **34** 454–477. MR2531779 https://doi.org/10.1002/rsa.20261
[9] Pemantle, R. (1990). Nonconvergence to unstable points in urn models and stochastic approximations. *Ann. Probab.* **18** 698–712. MR1055428
[10] Pemantle, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4** 1–79. MR2282181 https://doi.org/10.1214/07-PS094
[11] van der Hofstad, R., Holmes, M., Kuznetsov, A. and Ruszel, W. (2016). Strongly reinforced Pólya urns with graph-based competition. *Ann. Appl. Probab.* **26** 2494–2539. MR3543903 https://doi.org/10.1214/16-AAP1153

Bernoulli **28**(4), 2022, 2563–2577
https://doi.org/10.3150/21-BEJ1429

Continuous-time digital search tree and a border aggregation model

SVANTE JANSON[1,a] and DEBLEENA THACKER[2,b]

[1] *Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden.*
[a] *svante.janson@math.uu.se*
[2] *Department of Mathematics, NYU, Shanghai, 1555 Century Avenue, Pudong New District, Shanghai, China 200122.* [b] *thackerdebleena@gmail.com*

We consider the continuous-time version of the random digital search tree, and construct a coupling with a border aggregation model as studied in Thacker and Volkov (*Ann. Appl. Probab.* **28** (2018) 1604–1633), showing a relation between the height of the tree and the time required for aggregation. This relation carries over to the corresponding discrete-time models. As a consequence we find a very precise asymptotic result for the time to aggregation, using recent results by Drmota et al. (*Random Structures Algorithms* **58** (2021) 430–467) for the digital search tree.

Keywords: Random walk; digital search trees; border aggregation model

References

[1] Aldous, D. and Shields, P. (1988). A diffusion limit for a class of randomly-growing binary trees. *Probab. Theory Related Fields* **79** 509–542. MR0966174 https://doi.org/10.1007/BF00318784
[2] Barlow, M.T., Pemantle, R. and Perkins, E.A. (1997). Diffusion-limited aggregation on a tree. *Probab. Theory Related Fields* **107** 1–60. MR1427716 https://doi.org/10.1007/s004400050076
[3] Drmota, M. (2002). The variance of the height of digital search trees. *Acta Inform.* **38** 261–276. MR1884411 https://doi.org/10.1007/s236-002-8034-5
[4] Drmota, M. (2009). *Random Trees: An Interplay Between Combinatorics and Probability*. Vienna: Springer. MR2484382 https://doi.org/10.1007/978-3-211-75357-6
[5] Drmota, M., Fuchs, M., Hwang, H.-K. and Neininger, R. (2021). Node profiles of symmetric digital search trees: Concentration properties. *Random Structures Algorithms* **58** 430–467. MR4234992 https://doi.org/10.1002/rsa.20979
[6] Drmota, M., Janson, S. and Neininger, R. (2008). A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.* **18** 288–333. MR2380900 https://doi.org/10.1214/07-AAP457
[7] Gut, A. (2005). *Probability: A Graduate Course*. *Springer Texts in Statistics*. New York: Springer. MR2125120
[8] Jacquet, P. and Szpankowski, W. (2015). *Analytic Pattern Matching: From DNA to Twitter*. Cambridge: Cambridge Univ. Press. MR3524836 https://doi.org/10.1017/CBO9780511843204
[9] Kesten, H. (1987). How long are the arms in DLA? *J. Phys. A* **20** L29–L33. MR0873177 https://doi.org/10.1088/0305-4470/20/1/007
[10] Knessl, C. and Szpankowski, W. (2000). Asymptotic behavior of the height in a digital search tree and the longest phrase of the Lempel-Ziv scheme. *SIAM J. Comput.* **30** 923–964. MR1778293 https://doi.org/10.1137/S0097539799356812
[11] Knuth, D.E. (1998). *The Art of Computer Programming. Vol. 3: Sorting and Searching*, 2nd ed. Reading, MA: Addison-Wesley. MR3077154
[12] Lawler, G.F., Bramson, M. and Griffeath, D. (1992). Internal diffusion limited aggregation. *Ann. Probab.* **20** 2117–2140. MR1188055
[13] Levine, L. and Peres, Y. (2007). Internal erosion and the exponent $3/4$. Preprint, 2007. Available at http://pi.math.cornell.edu/~levine/erosion.pdf.

1350-7265

[14] Levine, L. and Silvestri, V. (2019). How long does it take for internal DLA to forget its initial profile? *Probab. Theory Related Fields* **174** 1219–1271. MR3980315 https://doi.org/10.1007/s00440-018-0880-7
[15] Mahmoud, H.M. (1992). *Evolution of Random Search Trees*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. New York: Wiley. MR1140708
[16] Silvestri, V. (2020). Internal DLA on cylinder graphs: Fluctuations and mixing. *Electron. Commun. Probab.* **25** Paper No. 61. MR4137946 https://doi.org/10.3390/mca25010013
[17] Thacker, D. and Volkov, S. (2018). Border aggregation model. *Ann. Appl. Probab.* **28** 1604–1633. MR3809473 https://doi.org/10.1214/17-AAP1339
[18] Witten, T.A. and Sander, L.M. (1983). Diffusion-limited aggregation. *Phys. Rev. B (3)* **27** 5686–5697. MR0704464 https://doi.org/10.1103/physrevb.27.5686

Bernoulli **28**(4), 2022, 2578–2605
https://doi.org/10.3150/21-BEJ1430

Matrix means and a novel high-dimensional shrinkage phenomenon

ASAD LODHIA[1,a], KEITH LEVIN[2,b] and ELIZAVETA LEVINA[3,c]

[1]*Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142-1027.* [a]*asad.i.lodhia@gmail.com*
[2]*Department of Statistics, University of Wisconsin-Madison, 1220 Medical Sciences Center, 1300 University Ave, Madison, WI 53706-1510.* [b]*kdlevin@wisc.edu*
[3]*Department of Statistics, University of Michigan, 323 West Hall, 1085 S. University Ave, Ann Arbor, MI 48109-1107.* [c]*elevina@umich.edu*

Many statistical settings call for estimating a population parameter, most typically the population mean, based on a sample of matrices. The most natural estimate of the population mean is the arithmetic mean, but there are many other matrix means that may behave differently, especially in high dimensions. Here we consider the matrix harmonic mean as an alternative to the arithmetic matrix mean. We show that in certain high-dimensional regimes, the harmonic mean yields an improvement over the arithmetic mean in estimation error as measured by the operator norm. Counter-intuitively, studying the asymptotic behavior of these two matrix means in a spiked covariance estimation problem, we find that this improvement in operator norm error does not imply better recovery of the leading eigenvector. We also show that a Rao-Blackwellized version of the harmonic mean is equivalent to a linear shrinkage estimator studied previously in the high-dimensional covariance estimation literature, while applying a similar Rao-Blackwellization to regularized sample covariance matrices yields a novel nonlinear shrinkage estimator. Simulations complement the theoretical results, illustrating the conditions under which the harmonic matrix mean yields an empirically better estimate.

Keywords: Matrix means; positive definite cone; covariance estimation; shrinkage Estimators

References

[1] Absil, P.-A., Mahony, R. and Sepulchre, R. (2004). Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.* **80** 199–220. MR2035507 https://doi.org/10.1023/B:ACAP.0000013855.14971.91
[2] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley Interscience. MR1990662
[3] 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
[4] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233
[5] Benaych-Georges, F. and Nadakuditi, R.R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007
[6] Berthet, Q. and Rigollet, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127
[7] Bhatia, R. (2007). *Positive Definite Matrices*. *Princeton Series in Applied Mathematics*. Princeton, NJ: Princeton Univ. Press. MR2284176
[8] Bickel, P.J. and Levina, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

1350-7265

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

[10] Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10** 186–198.

[11] Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949 https://doi.org/10.1198/jasa.2011.tm10560

[12] Cai, T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281 https://doi.org/10.1007/s00440-014-0562-z

[13] Cai, T.T., Ren, Z. and Zhou, H.H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electron. J. Stat.* **10** 1–59. MR3466172 https://doi.org/10.1214/15-EJS1081

[14] Cai, T.T., Zhang, C.-H. and Zhou, H.H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 https://doi.org/10.1214/09-AOS752

[15] Cai, T.T. and Zhou, H.H. (2012). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statist. Sinica* **22** 1319–1349. MR3027084

[16] Cai, T.T. and Zhou, H.H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607 https://doi.org/10.1214/12-AOS998

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

[18] Chen, Y., Wiesel, A., Eldar, Y.C. and Hero, A.O. (2010). Shrinkage algorithms for MMSE covariance estimation. *IEEE Trans. Signal Process.* **58** 5016–5029. MR2722661 https://doi.org/10.1109/TSP.2010.2053029

[19] 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

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

[21] 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

[22] 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

[23] Edelman, A., Arias, T.A. and Smith, S.T. (1999). The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* **20** 303–353. MR1646856 https://doi.org/10.1137/S0895479895290954

[24] El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012 https://doi.org/10.1214/07-AOS581

[25] Fan, J., Liao, Y. and Liu, H. (2016). An overview of the estimation of large covariance and precision matrices. *Econom. J.* **19** C1–C32. MR3501529 https://doi.org/10.1111/ectj.12061

[26] Fan, Z., Johnstone, I.M. and Sun, Y. (2018). Spiked covariances and principal components analysis in high-dimensional random effects models. Available at arXiv:1806.09529.

[27] Fan, Z., Sun, Y. and Wang, Z. (2021). Principal components in linear mixed models with general bulk. *Ann. Statist.* **49** 1489–1513. MR4302572 https://doi.org/10.1214/20-aos2010

[28] Fisher, T.J. and Sun, X. (2011). Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Comput. Statist. Data Anal.* **55** 1909–1918. MR2765053 https://doi.org/10.1016/j.csda.2010.12.006

[29] Hoyle, D.C. and Rattray, M. (2007). Statistical mechanics of learning multiple orthogonal signals: Asymptotic theory and fluctuation effects. *Phys. Rev. E (3)* **75** 016101, 13. MR2324655 https://doi.org/10.1103/PhysRevE.75.016101

[30] Johnstone, I.M. and Lu, A.Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

[31] Kolaczyk, E.D., Lin, L., Rosenberg, S., Walters, J. and Xu, J. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *Ann. Statist.* **48** 514–538. MR4065172 https://doi.org/10.1214/19-AOS1820

[32] Kong, W. and Valiant, G. (2017). Spectrum estimation from samples. *Ann. Statist.* **45** 2218–2247. MR3718167 https://doi.org/10.1214/16-AOS1525

[33] Konno, Y. (1988). Exact moments of the multivariate F and beta distributions. *J. Japan Statist. Soc.* **18** 123–130. MR0980810

[34] Kubokawa, T. and Inoue, A. (2014). Estimation of covariance and precision matrices under scale-invariant quadratic loss in high dimension. *Electron. J. Stat.* **8** 130–158. MR3165436 https://doi.org/10.1214/14-EJS878

[35] Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Ann. Statist.* **44** 928–953. MR3485949 https://doi.org/10.1214/15-AOS1393

[36] 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

[37] 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

[38] 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

[39] Levin, K., Lodhia, A. and Levina, E. (2019). Recovering low-rank structure from multiple networks with unknown edge distributions. *J. Mach. Learn. Res.* To appear. arXiv:1906.07265.

[40] Li, W., Chen, J., Qin, Y., Bai, Z. and Yao, J. (2013). Estimation of the population spectral distribution from a large dimensional sample covariance matrix. *J. Statist. Plann. Inference* **143** 1887–1897. MR3095079 https://doi.org/10.1016/j.jspi.2013.06.017

[41] Lodhia, A. (2021). Harmonic means of Wishart random matrices. *Random Matrices Theory Appl.* **10** Paper No. 2150016, 24. MR4260211 https://doi.org/10.1142/S2010326321500167

[42] Lunagómez, S., Olhede, S.C. and Wolfe, P.J. (2020). Modeling network populations via graph distances. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2020.1763803.

[43] Male, C. (2012). The norm of polynomials in large random and deterministic matrices. *Probab. Theory Related Fields* **154** 477–532. With an appendix by Dimitri Shlyakhtenko. MR3000553 https://doi.org/10.1007/s00440-011-0375-2

[44] Marčenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72** 507–536. MR0208649

[45] Marrinan, T., Beveridge, J.R., Draper, B., Kirby, M. and Peterson, C. (2014). Finding the subspace mean or median to fit your need. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1082–1089.

[46] Mingo, J.A. and Speicher, R. (2017). *Free Probability and Random Matrices*. *Fields Institute Monographs* **35**. New York: Springer; Toronto, ON: Fields Institute for Research in Mathematical Sciences. MR3585560 https://doi.org/10.1007/978-1-4939-6942-5

[47] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0652932

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

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

[50] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0346957

[51] Rao, N.R., Mingo, J.A., Speicher, R. and Edelman, A. (2008). Statistical eigen-inference from large Wishart matrices. *Ann. Statist.* **36** 2850–2885. MR2485015 https://doi.org/10.1214/07-AOS583

[52] Rothman, A.J., Levina, E. and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* **104** 177–186. MR2504372 https://doi.org/10.1198/jasa.2009.0101

[53] Schwartzman, A. (2016). Lognormal distributions and geometric averages of symmetric positive definite matrices. *Int. Stat. Rev.* **84** 456–486. MR3580425 https://doi.org/10.1111/insr.12113

[54] Smith, S.T. (2005). Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.* **53** 1610–1630. MR2148599 https://doi.org/10.1109/TSP.2005.845428

[55] Sporns, O. (2012). *Discovering the Human Connectome*. Cambridge: MIT Press.

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

[57] Touloumis, A. (2015). Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings. *Comput. Statist. Data Anal.* **83** 251–261. MR3281809 https://doi.org/10.1016/j.csda.2014.10.018

[58] 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

[59] Wu, W.B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844. MR2024760 https://doi.org/10.1093/biomet/90.4.831

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

[61] Yu, Y., Wang, T. and Samworth, R.J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika* **102** 315–323. MR3371006 https://doi.org/10.1093/biomet/asv008

Bernoulli **28**(4), 2022, 2606–2633
https://doi.org/10.3150/21-BEJ1431

A pseudo-marginal sequential Monte Carlo online smoothing algorithm

PIERRE GLOAGUEN[1,a], SYLVAIN LE CORFF[2,b] and JIMMY OLSSON[3,c]

[1]*AgroParisTech, Paris, France.* [a]*pierre.gloaguem@agroparistech.fr*
[2]*Télécom SudParis, Paris, France.* [b]*sylvain_lecorff@telecom-sudparis.eu*
[3]*KTH Royal Institute of Technology, Stockholm, Sweden.* [c]*jimmyol@kth.se*

We consider online computation of expectations of additive state functionals under general path probability measures proportional to products of unnormalised transition densities. These transition densities are assumed to be intractable but possible to estimate, with or without bias. Using pseudo-marginalisation techniques we are able to extend the particle-based, rapid incremental smoother (PaRIS) algorithm proposed in [*Bernoulli* **23**(3) (2017) 1951–1996] to this setting. The resulting algorithm, which has a linear complexity in the number of particles and constant memory requirements, applies to a wide range of challenging path-space Monte Carlo problems, including smoothing in partially observed diffusion processes and models with intractable likelihood. The algorithm is furnished with several theoretical results, including a central limit theorem, establishing its convergence and numerical stability. Moreover, under strong mixing assumptions we establish a novel $O(n\varepsilon)$ bound on the asymptotic bias of the algorithm, where n is the path length and ε controls the bias of the transition-density estimators.

Keywords: Central limit theorem; exponential concentration; partially observed diffusions; particle smoothing; pseudo-marginal methods; sequential Monte Carlo methods

References

[1] 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
[2] Bally, V. and Talay, D. (1996). The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods Appl.* **2** 93–128. MR1401964 https://doi.org/10.1515/mcma.1996.2.2.93
[3] 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
[4] Beskos, A., Papaspiliopoulos, O., Roberts, G.O. and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 333–382. MR2278331 https://doi.org/10.1111/j.1467-9868.2006.00552.x
[5] Candanedo, L.M., Feldheim, V. and Deramaix, D. (2017). A methodology based on hidden Markov models for occupancy detection and a case study in a low energy residential building. *Energy Build.* **148** 327–341.
[6] Cappé, O. (2011). Online EM algorithm for hidden Markov models. *J. Comput. Graph. Statist.* **20** 728–749. MR2878999 https://doi.org/10.1198/jcgs.2011.09109
[7] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. *Springer Series in Statistics*. New York: Springer. MR2159833
[8] Chopin, N. and Papaspiliopoulos, O. (2020). *An Introduction to Sequential Monte Carlo*. *Springer Series in Statistics*. Cham: Springer. MR4215639 https://doi.org/10.1007/978-3-030-47845-2
[9] Cornebise, J., Moulines, É. and Olsson, J. (2008). Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **18** 461–480. MR2461889 https://doi.org/10.1007/s11222-008-9089-4
[10] Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. *Probability and Its Applications (New York)*. New York: Springer. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

1350-7265

[11] Del Moral, P., Doucet, A. and Singh, S.S. (2010). A backward particle interpretation of Feynman-Kac formulae. *ESAIM Math. Model. Numer. Anal.* **44** 947–975. MR2731399 https://doi.org/10.1051/m2an/2010048
[12] Del Moral, P., Jacod, J. and Protter, P. (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields* **120** 346–368. MR1843179 https://doi.org/10.1007/PL00008786
[13] Douc, R., Garivier, A., Moulines, E. and Olsson, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* **21** 2109–2145. MR2895411 https://doi.org/10.1214/10-AAP735
[14] Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10** 197–208.
[15] Durham, G.B. and Gallant, A.R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20** 297–338. MR1939904 https://doi.org/10.1198/073500102288618397
[16] Fearnhead, P., Papaspiliopoulos, O. and Roberts, G.O. (2008). Particle filters for partially observed diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 755–777. MR2523903 https://doi.org/10.1111/j.1467-9868.2008.00661.x
[17] Fearnhead, P., Papaspiliopoulos, O., Roberts, G.O. and Stuart, A. (2010). Random-weight particle filtering of continuous time processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 497–512. MR2758525 https://doi.org/10.1111/j.1467-9868.2010.00744.x
[18] Gloaguen, P., Etienne, M.-P. and Le Corff, S. (2018). Online sequential Monte Carlo smoother for partially observed diffusion processes. *EURASIP J. Adv. Signal Process.* **2018** 9.
[19] Gloaguen, P., Le Corff, S. and Olsson, J. (2022). Supplement to "A pseudo-marginal sequential Monte Carlo online smoothing algorithm." https://doi.org/10.3150/21-BEJ1431SUPP
[20] Godsill, S.J., Doucet, A. and West, M. (2004). Monte Carlo smoothing for nonlinear times series. *J. Amer. Statist. Assoc.* **99** 156–168. MR2054295 https://doi.org/10.1198/016214504000000151
[21] Le Gland, F. and Mevel, L. (1997). Recursive estimation in HMMs. In *Proc. IEEE Conf. Decis. Control* 3468–3473.
[22] Le Gland, F. and Oudjane, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.* **14** 144–187. MR2023019 https://doi.org/10.1214/aoap/1075828050
[23] Marin, J.-M., Pudlo, P., Robert, C.P. and Ryder, R.J. (2012). Approximate Bayesian computational methods. *Stat. Comput.* **22** 1167–1180. MR2992292 https://doi.org/10.1007/s11222-011-9288-2
[24] Martin, J.S., Jasra, A., Singh, S.S., Whiteley, N., Del Moral, P. and McCoy, E. (2014). Approximate Bayesian computation for smoothing. *Stoch. Anal. Appl.* **32** 397–420. MR3197814 https://doi.org/10.1080/07362994.2013.879262
[25] McGree, J.M., Drovandi, C.C., White, G. and Pettitt, A.N. (2016). A pseudo-marginal sequential Monte Carlo algorithm for random effects models in Bayesian sequential design. *Stat. Comput.* **26** 1121–1136. MR3535048 https://doi.org/10.1007/s11222-015-9596-z
[26] Olsson, J. and Alenlöv, J.W. (2020). Particle-based online estimation of tangent filters with application to parameter estimation in nonlinear state-space models. *Ann. Inst. Statist. Math.* **72** 545–576. MR4067236 https://doi.org/10.1007/s10463-018-0698-1
[27] Olsson, J. and Ströjby, J. (2011). Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators. *Electron. J. Stat.* **5** 1090–1122. MR2836770 https://doi.org/10.1214/11-EJS632
[28] Olsson, J. and Westerborn, J. (2017). Efficient particle-based online smoothing in general hidden Markov models: The PaRIS algorithm. *Bernoulli* **23** 1951–1996. MR3624883 https://doi.org/10.3150/16-BEJ801
[29] Papavasiliou, A. (2006). Parameter estimation and asymptotic stability in stochastic filtering. *Stochastic Process. Appl.* **116** 1048–1065. MR2238613 https://doi.org/10.1016/j.spa.2006.01.002
[30] Michelot, T., Langrock, R. and Patterson, T.A. (2016). MoveHMM: An R package for the statistical modelling of animal movement data using hidden Markov models. *Methods Ecol. Evol.* **7** 1308–1315.
[31] Pitt, M.K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. MR1702328 https://doi.org/10.2307/2670179
[32] Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE* 257–286.

[33] Rauch, H.E., Tung, F. and Striebel, C.T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA J.* **3** 1445–1450. MR0181489 https://doi.org/10.2514/3.3166
[34] Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. *Institute of Mathematical Statistics Textbooks* **3**. Cambridge: Cambridge Univ. Press. MR3154309 https://doi.org/10.1017/CBO9781139344203
[35] Särkkä, S., Vehtari, A. and Lampinen, J. (2007). Rao-Blackwellized particle filter for multiple target tracking. *Inf. Fusion* **8** 2–15.
[36] Uhlenbeck, G.E. and Ornstein, L.S. (1930). On the theory of the Brownian motion. *Phys. Rev.* **36** 823.
[37] Yau, C., Papaspiliopoulos, O., Roberts, G.O. and Holmes, C. (2011). Bayesian non-parametric hidden Markov models with applications in genomics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 37–57. MR2797735 https://doi.org/10.1111/j.1467-9868.2010.00756.x
[38] Yonekura, S. and Beskos, A. (2020). Online smoothing for diffusion processes observed with noise. arXiv:2003.12247.

Bernoulli **28**(4), 2022, 2634–2662
https://doi.org/10.3150/21-BEJ1432

Deviation inequalities and Cramér-type moderate deviations for the explosive autoregressive process

HUI JIANG[1,a], YILONG WAN[1,b] and GUANGYU YANG[2,c]

[1]*Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, China.* [a]*huijiang@nuaa.edu.cn,* [b]*wanyil@nuaa.edu.cn*
[2]*School of Mathematics and Statistics, Zhengzhou University, Zhengzhou, China.* [c]*guangyu@zzu.edu.cn*

This paper concerns the asymptotic properties of the quadratic functionals and associated ordinary least squares estimator in the explosive first-order Gaussian autoregressive process. By the deviation inequalities for multiple Wiener-Itô integrals and asymptotic analysis techniques, Cramér-type moderate deviations are achieved under the explosive and mildly explosive frameworks. As applications, the global and local powers for the unit root test are shown to approach one at exponential rates. Simulation experiments are conducted to confirm the theoretical results.

Keywords: Cramér-type moderate deviations; deviation inequality; explosive autoregressive process; multiple Wiener-Itô integrals; unit root

References

[1] Anderson, T.W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Stat.* **30** 676–687. MR0107347 https://doi.org/10.1214/aoms/1177706198
[2] Bercu, B. (2001). On large deviations in the Gaussian autoregressive process: Stable, unstable and explosive cases. *Bernoulli* **7** 299–316. MR1828507 https://doi.org/10.2307/3318740
[3] Bercu, B., Gamboa, F. and Rouault, A. (1997). Large deviations for quadratic forms of stationary Gaussian processes. *Stochastic Process. Appl.* **71** 75–90. MR1480640 https://doi.org/10.1016/S0304-4149(97)00071-9
[4] Bryc, W. and Dembo, A. (1997). Large deviations for quadratic functionals of Gaussian processes **10** 307–332. MR1455147 https://doi.org/10.1023/A:1022656331883
[5] Buchmann, B. and Chan, N.H. (2013). Unified asymptotic theory for nearly unstable AR(p) processes. *Stochastic Process. Appl.* **123** 952–985. MR3005011 https://doi.org/10.1016/j.spa.2012.09.014
[6] Chan, N.H. and Wei, C.Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15** 1050–1063. MR0902245 https://doi.org/10.1214/aos/1176350492
[7] Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401. MR0924877 https://doi.org/10.1214/aos/1176350711
[8] Chang, J., Shao, Q.-M. and Zhou, W.-X. (2016). Cramér-type moderate deviations for Studentized two-sample U-statistics with applications. *Ann. Statist.* **44** 1931–1956. MR3546439 https://doi.org/10.1214/15-AOS1375
[9] Djellout, H., Guillin, A. and Wu, L. (2006). Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 393–416. MR2242954 https://doi.org/10.1016/j.anihpb.2005.04.006
[10] Douissi, S., Es-Sebaiy, K. and Viens, F.G. (2019). Berry-Esséen bounds for parameter estimation of general Gaussian processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **16** 633–664. MR3949273 https://doi.org/10.30757/alea.v16-23
[11] Es-Sebaiy, K. and Viens, F.G. (2019). Optimal rates for parameter estimation of stationary Gaussian processes. *Stochastic Process. Appl.* **129** 3018–3054. MR3985554 https://doi.org/10.1016/j.spa.2018.08.010

[12] Fan, X., Grama, I. and Liu, Q. (2020). Cramér moderate deviation expansion for martingales with one-sided Sakhanenko's condition and its applications. *J. Theoret. Probab.* **33** 749–787. MR4091583 https://doi.org/10.1007/s10959-019-00949-2

[13] Fan, X., Grama, I., Liu, Q. and Shao, Q.-M. (2019). Self-normalized Cramér type moderate deviations for martingales. *Bernoulli* **25** 2793–2823. MR4003565 https://doi.org/10.3150/18-BEJ1071

[14] Fan, X. and Shao, Q.-M. (2018). Berry-Esseen bounds for self-normalized martingales. *Commun. Math. Stat.* **6** 13–27. MR3767959 https://doi.org/10.1007/s40304-017-0122-9

[15] Fang, X., Luo, L. and Shao, Q.-M. (2020). A refined Cramér-type moderate deviation for sums of local statistics. *Bernoulli* **26** 2319–2352. MR4091111 https://doi.org/10.3150/20-BEJ1195

[16] Fuller, W.A. (1996). *Introduction to Statistical Time Series*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. MR1365746

[17] Gao, F., Xiong, J. and Zhao, X. (2018). Moderate deviations and nonparametric inference for monotone functions. *Ann. Statist.* **46** 1225–1254. MR3798002 https://doi.org/10.1214/17-AOS1583

[18] Giraitis, L. and Phillips, P.C.B. (2006). Uniform limit theory for stationary autoregression. *J. Time Series Anal.* **27** 51–60. MR2235146 https://doi.org/10.1111/j.1467-9892.2005.00452.x

[19] Hu, Y. and Nualart, D. (2010). Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Statist. Probab. Lett.* **80** 1030–1038. MR2638974 https://doi.org/10.1016/j.spl.2010.02.018

[20] Hu, Y., Nualart, D. and Zhou, H. (2019). Parameter estimation for fractional Ornstein-Uhlenbeck processes of general Hurst parameter. *Stat. Inference Stoch. Process.* **22** 111–142. MR3918739 https://doi.org/10.1007/s11203-017-9168-2

[21] Jiang, H., Liu, H. and Zhou, Y. (2020). Asymptotic properties for the parameter estimation in Ornstein-Uhlenbeck process with discrete observations. *Electron. J. Stat.* **14** 3192–3229. MR4146360 https://doi.org/10.1214/20-EJS1738

[22] Jiang, H. Wan, Y., Yang, G. (2022). Supplement to "Deviation inequalities and Cramér-type moderate deviations for the explosive autoregressive process." https://doi.org/10.3150/21-BEJ1432SUPP

[23] Jiang, H. and Zhang, N. (2020). Cramér-type moderate deviations for statistics in the non-stationary Ornstein-Uhlenbeck process. *Stochastics* **92** 478–496. MR4085761 https://doi.org/10.1080/17442508.2019.1637877

[24] Liu, W. and Shao, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. *Ann. Statist.* **38** 1913–1935. MR2662363 https://doi.org/10.1214/09-AOS774

[25] Miao, Y. and Shen, S. (2009). Moderate deviation principle for autoregressive processes. *J. Multivariate Anal.* **100** 1952–1961. MR2543078 https://doi.org/10.1016/j.jmva.2009.06.005

[26] Miao, Y., Wang, Y. and Yang, G. (2015). Moderate deviation principles for empirical covariance in the neighbourhood of the unit root. *Scand. J. Stat.* **42** 234–255. MR3318034 https://doi.org/10.1111/sjos.12104

[27] Nourdin, I. and Tran, T.T.D. (2019). Statistical inference for Vasicek-type model driven by Hermite processes. *Stochastic Process. Appl.* **129** 3774–3791. MR3997661 https://doi.org/10.1016/j.spa.2018.10.005

[28] Phillips, P.C.B. (1988). Regression theory for near-integrated time series. *Econometrica* **56** 1021–1043. MR0964147 https://doi.org/10.2307/1911357

[29] Phillips, P.C.B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *J. Econometrics* **136** 115–130. MR2328587 https://doi.org/10.1016/j.jeconom.2005.08.002

[30] Proïa, F. (2020). Moderate deviations in a class of stable but nearly unstable processes. *J. Statist. Plann. Inference* **208** 66–81. MR4076349 https://doi.org/10.1016/j.jspi.2020.01.009

[31] 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

[32] White, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Stat.* **29** 1188–1197. MR0100952 https://doi.org/10.1214/aoms/1177706450

[33] Worms, J. (1999). Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* **4** no. 8. MR1684149 https://doi.org/10.1214/EJP.v4-45

[34] Worms, J. (2001). Large and moderate deviations upper bounds for the Gaussian autoregressive process. *Statist. Probab. Lett.* **51** 235–243. MR1822730 https://doi.org/10.1016/S0167-7152(00)00134-6

Bernoulli **28**(4), 2022, 2663–2688
https://doi.org/10.3150/21-BEJ1433

Estimation of Wasserstein distances in the Spiked Transport Model

JONATHAN NILES-WEED[1,a] and PHILIPPE RIGOLLET[2,b]

[1] *Courant Institute of Mathematical Sciences & Center for Data Science, New York University, 251 Mercer Street, New York, NY 10012-1185, USA.* [a] *jnw@cims.nyu.edu*
[2] *Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA.* [b] *rigollet@math.mit.edu*

Abstract. We propose a new statistical model, the *spiked transport model*, which formalizes the assumption that two probability distributions differ only on a low-dimensional subspace. We study the minimax rate of estimation for the Wasserstein distance under this model and show that this low-dimensional structure can be exploited to avoid the curse of dimensionality. As a byproduct of our minimax analysis, we establish a lower bound showing that, in the absence of such structure, the plug-in estimator is nearly rate-optimal for estimating the Wasserstein distance in high dimension. We also give evidence for a statistical-computational gap and conjecture that any computationally efficient estimator is bound to suffer from the curse of dimensionality.

Keywords: Wasserstein distance; optimal transport; high-dimensional statistics

References

[1] Ahidar-Coutrix, A., Le Gouic, T. and Paris, Q. (2020). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177** 323–368. MR4095017 https://doi.org/10.1007/s00440-019-00950-0
[2] Alaux, J., Grave, E., Cuturi, M. and Joulin, A. (2019). Unsupervised hyper-alignment for multilingual word embeddings. In *ICLR*.
[3] Alvarez-Melis, D., Jaakkola, T.S. and Jegelka, S. (2018). Structured optimal transport. In *AISTATS*.
[4] Arjovsky, , Chintala, , and Bottou, Wasserstein generative adversarial networks. *ICML*, 2017.
[5] Ba, K.D., Nguyen, H.L., Nguyen, H.N. and Rubinfeld, R. (2011). Sublinear time algorithms for Earth mover's distance. *Theory Comput. Syst.* **48** 428–442. MR2763110 https://doi.org/10.1007/s00224-010-9265-8
[6] Bandeira, A.S., Perry, A. and Wein, A.S. (2018). Notes on computational-to-statistical gaps: Predictions using statistical physics. *Port. Math.* **75** 159–186. MR3892753 https://doi.org/10.4171/PM/2014
[7] Berthet, Q. and Rigollet, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127
[8] Bigot, J., Cazelles, E. and Papadakis, N. (2019). Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electron. J. Stat.* **13** 5120–5150. MR4041704 https://doi.org/10.1214/19-EJS1637
[9] Blum, A., Dwork, C., McSherry, F. and Nissim, K. (2005). Practical privacy: The sulq framework. In *PODS*.
[10] Bobkov, S. and Ledoux, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** v+126. MR4028181 https://doi.org/10.1090/memo/1259
[11] Bobkov, S.G. and Götze, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772 https://doi.org/10.1006/jfan.1998.3326
[12] Boissard, E. and Le Gouic, T. (2014). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 539–563. MR3189084 https://doi.org/10.1214/12-AIHP517

1350-7265

[13] Bolley, F., Guillin, A. and Villani, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* **137** 541–593. MR2280433 https://doi.org/10.1007/s00440-006-0004-7
[14] Brennan, M., Bresler, G. and Huleihel, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *COLT*.
[15] Bubeck, S., Lee, Y.T., Price, E. and Razenshteyn, I.P. (2019). Adversarial Examples from Computational Constraints. In *ICML*.
[16] Cai, T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281 https://doi.org/10.1007/s00440-014-0562-z
[17] Cai, T.T., Liang, T. and Rakhlin, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann. Statist.* **45** 1403–1430. MR3670183 https://doi.org/10.1214/16-AOS1488
[18] Canas, G. and Rosasco, L. (2012). Learning probability measures with respect to optimal transport metrics. In *NIPS*.
[19] Cazelles, E., Seguy, V., Bigot, J., Cuturi, M. and Papadakis, N. (2018). Geodesic PCA versus log-PCA of histograms in the Wasserstein space. *SIAM J. Sci. Comput.* **40** B429–B456. MR3780753 https://doi.org/10.1137/17M1143459
[20] Claici, S. and Solomon, J. (2018). Wasserstein coresets for Lipschitz costs. arXiv:1805.07412.
[21] Courty, N., Flamary, R., Tuia, D. and Rakotomamonjy, A. (2016). Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1853–1865.
[22] Del Barrio, E., Gamboa, F., Gordaliza, P. and Loubes, J.-M. (2019). Obtaining fairness using optimal transport theory. In *ICML*.
[23] del Barrio, E., Gordaliza, P., Lescornel, H. and Loubes, J.-M. (2019). Central limit theorem and bootstrap procedure for Wasserstein's variations with an application to structural relationships between distributions. *J. Multivariate Anal.* **169** 341–362. MR3875604 https://doi.org/10.1016/j.jmva.2018.09.014
[24] del Barrio, E., Inouzhe, H., Loubes, J.-M., Matrán, C., Mayo-Íscar and optimalFlow, A. (2019). Optimal-transport approach to flow cytometry gating and population matching. arXiv:1907.08006.
[25] Dereich, S., Scheutzow, M. and Schottstedt, R. (2013). Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183–1203. MR3127919 https://doi.org/10.1214/12-AIHP489
[26] Deshpande, I., Hu, Y.-T., Sun, R., Pyrros, A., Siddiqui, N., Koyejo, S., Zhao, Z., Forsyth, D., Schwing and Max-sliced, A. (2019). Wasserstein distance and its use for GANs. arXiv:1904.05877.
[27] Deutsch, F. (1995). The angle between subspaces of a Hilbert space. In *Approximation Theory, Wavelets and Applications (Maratea, 1994). NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci.* **454** 107–130. Dordrecht: Kluwer Academic. MR1340886
[28] Diakonikolas, I., Kane, D.M. and Stewart, A. (2017). Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures (extended abstract). In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017* 73–84. Los Alamitos, CA: IEEE Computer Soc. MR3734219 https://doi.org/10.1109/FOCS.2017.16
[29] Dixmier, J. (1949). Étude sur les variétés et les opérateurs de Julia, avec quelques applications. *Bull. Soc. Math. France* **77** 11–101. MR0032937
[30] Dudley, R.M. (1968). The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Stat.* **40** 40–50. MR0236977 https://doi.org/10.1214/aoms/1177697802
[31] Dumitrascu, B., Villar, S., Mixon, D.G. and Engelhardt, B.E. (2021). Optimal marker gene selection for cell type discrimination in single cell analyses. *Nat. Commun.* **12** 1–8.
[32] Feldman, V., Grigorescu, E., Reyzin, L., Vempala, S.S. and Xiao, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** Art. 8, 37. MR3664576 https://doi.org/10.1145/3046674
[33] Fernando, B., Habrard, A., Sebban, M. and Tuytelaars, T. (2013). Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE International Conference on Computer Vision* 2960–2967.
[34] Feydy, J., Charlier, B., Vialard, F. and Peyré, G. (2017). Optimal transport for diffeomorphic registration. In *MICCAI*.

[35] Flamary, R., Cuturi, M., Courty, N. and Rakotomamonjy, A. (2018). Wasserstein discriminant analysis. *Mach. Learn.* **107** 1923–1945. MR3854402 https://doi.org/10.1007/s10994-018-5717-1

[36] Forrow, A., Hütter, J.-C., Nitzan, M., Rigollet, P., Schiebinger, G. and Weed, J. (2019). Statistical optimal transport via factored couplings. In *AISTATS*.

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

[38] Friedman, J.H. and Tukey, J.W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.* **23** 881–890.

[39] Genevay, A., Peyré, G. and Cuturi, M. (2018). Learning generative models with sinkhorn divergences. In *AISTATS*.

[40] Gong, B., Shi, Y., Sha, F. and Grauman, K. (2012). Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* 2066–2073. IEEE.

[41] Gozlan, N. (2007). Characterization of Talagrand's like transportation-cost inequalities on the real line. *J. Funct. Anal.* **250** 400–425. MR2352486 https://doi.org/10.1016/j.jfa.2007.05.025

[42] Gozlan, N. (2009). A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.* **37** 2480–2498. MR2573565 https://doi.org/10.1214/09-AOP470

[43] Gozlan, N. and Léonard, C. (2007). A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** 235–283. MR2322697 https://doi.org/10.1007/s00440-006-0045-y

[44] Gozlan, N. and Léonard, C. (2010). Transport inequalities. A survey. *Markov Process. Related Fields* **16** 635–736. MR2895086

[45] Grave, E., Joulin, A. and Berthet, Q. (2019). Unsupervised alignment of embeddings with wasserstein procrustes. In *AISTATS*.

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

[47] Janati, H., Cuturi, M. and Gramfort, A. (2019). Wasserstein regularization for sparse multi-task regression. In *AISTATS*.

[48] Jiao, J., Han, Y. and Weissman, T. (2018). Minimax estimation of the L_1 distance. *IEEE Trans. Inf. Theory* **64** 6672–6706. MR3860754 https://doi.org/10.1109/TIT.2018.2846245

[49] Johnson, W.B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. In *Conference in Modern Analysis and Probability (New Haven, Conn., 1982)*. *Contemp. Math.* **26** 189–206. Providence, RI: Amer. Math. Soc. MR0737400 https://doi.org/10.1090/conm/026/737400

[50] 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

[51] Kearns, M. (1998). Efficient noise-tolerant learning from statistical queries. *J. ACM* **45** 983–1006. MR1678849 https://doi.org/10.1145/293347.293351

[52] Klatt, M., Tameling, C. and Munk, A. (2020). Empirical regularized optimal transport: Statistical theory and applications. *SIAM J. Math. Data Sci.* **2** 419–443. MR4105566 https://doi.org/10.1137/19M1278788

[53] Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R. and Rohde, G.K. (2019). Generalized sliced Wasserstein distances. Preprint, arXiv:1902.00434.

[54] Kouw, W.M. and Loog, M. (2019). A review of domain adaptation without target labels. *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 766–785.

[55] Kroshnin, A., Spokoiny, V. and Suvorikova, A. (2021). Statistical inference for Bures-Wasserstein barycenters. *Ann. Appl. Probab.* **31** 1264–1298. MR4278784 https://doi.org/10.1214/20-aap1618

[56] Kruskal, J.B. (1969). Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new "index of condensation". In *Statistical Computation* 427–440. Elsevier.

[57] Kruskal, J.B. (1972). Linear transformation of multivariate data to reveal clustering. *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences* **1** 181–191.

[58] Lafzi, A., Moutinho, C., Picelli, S. and Heyn, H. (2018). Tutorial: Guidelines for the experimental design of single-cell RNA sequencing studies. *Nat. Protoc.* **13** 2742–2757.

[59] Lavenant, H., Claici, S., Chien, E. and Solomon, J. (2018). Dynamical optimal transport on discrete surfaces. *ACM Trans. Graph.* **37** 250:1–250:16.

[60] Le Gouic, T., Paris, Q., Rigollet, P. and Stromme, A.J. (2019). Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space. arXiv:1908.00828.
[61] 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
[62] Liang, T. (2019). On the Minimax Optimality of Estimating the Wasserstein Metric. arXiv:1908.10324.
[63] Lim, S., Lee, S.-E., Chang, S. and Ye, J.C. (2019). CycleGAN with a Blur Kernel for Deconvolution Microscopy: Optimal Transport Geometry. arXiv:1908.09414.
[64] Lin, T., Zheng, Z., Chen, E.Y., Cuturi, M. and Jordan, M.I. (2021). On projection robust optimal transport: Sample complexity and model misspecification. In *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event* 262–270.
[65] Ma, Z. and Wu, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300
[66] Marton, K. (1996). Bounding $\overline{d}$-distance by informational divergence: A method to prove measure concentration. *Ann. Probab.* **24** 857–866. MR1404531 https://doi.org/10.1214/aop/1039639365
[67] Marton, K. (1996). A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** 556–571. MR1392329 https://doi.org/10.1007/BF02249263
[68] Maurey, B. (1991). Some deviation inequalities. *Geom. Funct. Anal.* **1** 188–197. MR1097258 https://doi.org/10.1007/BF01896377
[69] McWhirter, C., Mixon, D.G. and Villar, S. (2020). SqueezeFit: Label-aware dimensionality reduction by semidefinite programming. *IEEE Trans. Inf. Theory* **66** 3878–3892. MR4115139 https://doi.org/10.1109/tit.2019.2962681
[70] Montavon, G., Müller, K. and Cuturi, M. (2016). Wasserstein training of restricted Boltzmann machines. In *NIPS* 3711–3719.
[71] Niles-Weed, J., Rigollet, P. (2022). Supplement to "Estimation of Wasserstein distances in the Spiked Transport Model." https://doi.org/10.3150/21-BEJ1433SUPP
[72] Panaretos, V.M. and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938
[73] Paty, F.-P. and Cuturi, M. (2019). Subspace Robust Wasserstein Distances. In *ICML*.
[74] Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Found. Trends Mach. Learn.* **11** 355–607.
[75] Pitié, F., Kokaram, A.C. and Dahyot, R. (2007). Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.* **107** 123–137.
[76] Rabin, J., Peyré, G., Delon, J. and Bernot, M. (2011). Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision* 435–446. Springer.
[77] Ramdas, A., García Trillos, N. and Cuturi, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy* **19** Paper No. 47, 15. MR3608466 https://doi.org/10.3390/e19020047
[78] Rigollet, P. and Weed, J. (2018). Entropic optimal transport is maximum-likelihood deconvolution. *C. R. Math. Acad. Sci. Paris* **356** 1228–1235. MR3907589 https://doi.org/10.1016/j.crma.2018.10.010
[79] Rigollet, P. and Weed, J. (2019). Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Inf. Inference* **8** 691–717. MR4045481 https://doi.org/10.1093/imaiai/iaz006
[80] Rolet, A., Cuturi, M. and Peyré, G. (2016). Fast dictionary learning with a smoothed wasserstein loss. MIFODS Semester on Learning under complex structure.
[81] Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A. and Lander, E.S. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176** 928–943.
[82] Schmitz, M.A., Heitz, M., Bonneel, N., Ngolè, F., Coeurjolly, D., Cuturi, M., Peyré, G. and Starck, J.-L. (2018). Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning. *SIAM J. Imaging Sci.* **11** 643–678. MR3769700 https://doi.org/10.1137/17M1140431
[83] Seguy, V. and Cuturi, M. (2015). Principal geodesic analysis for probability measures under the optimal transport metric. In *NIPS*.
[84] Singh, S. and Póczos, B. (2018). Minimax distribution estimation in Wasserstein distance. arXiv:1802.08855.

[85] Solomon, J., de Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T. and Guibas, L. (2015). Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.* **34** 66:1–66:11.

[86] Solomon, J., Peyré, G., Kim, V.G. and Sra, S. (2016). Entropic metric alignment for correspondence problems. *ACM Trans. Graph.* **35** 72:1–72:13.

[87] Staib, M., Claici, S., Solomon, J.M. and Jegelka, S. (2017). Parallel streaming wasserstein barycenters. In *NIPS*.

[88] Talagrand, M. (1996). Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6** 587–600. MR1392331 https://doi.org/10.1007/BF02249265

[89] Tameling, C. and Munk, A. (2018). Computational strategies for statistical inference based on empirical optimal transport. In *2018 IEEE Data* 175–179.

[90] Timan, A.F. (1994). *Theory of Approximation of Functions of a Real Variable*. New York: Dover. Translated from the Russian by J. Berry, Translation edited and with a preface by J. Cossar, Reprint of the 1963 English translation. MR1262128

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

[92] Valiant, G. and Valiant, P. (2010). A CLT and tight lower bounds for estimating entropy. *Electron. Colloq. Comput. Complex.* **17** 183.

[93] Valiant, G. and Valiant, P. (2011). The power of linear estimators. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011* 403–412. Los Alamitos, CA: IEEE Computer Soc. MR2932716 https://doi.org/10.1109/FOCS.2011.81

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

[95] Wang, F., Liang, S., Kumar, T., Navin, N. and Chen, K. (2019). Scmarker: ab initio marker selection for single cell transcriptome profiling. *PLoS Comput. Biol.* **15** e1007445.

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

[97] Weed, J. and Berthet, Q. (2019). Estimation of smooth densities in wasserstein distance. In *COLT*.

[98] Wu, Y. and Yang, P. (2019). Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *Ann. Statist.* **47** 857–883. MR3909953 https://doi.org/10.1214/17-AOS1665

[99] Yang, K.D., Damodaran, K., Venkatchalapathy, S., Soylemezoglu, A.C., Shivashankar, G. and Uhler, C. (2018). Autoencoder and optimal transport to infer single-cell trajectories of biological processes. bioRxiv 455469.

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

Bernoulli **28**(4), 2022, 2689–2715
https://doi.org/10.3150/21-BEJ1434

A presmoothing approach for estimation in the semiparametric Cox mixture cure model

ENI MUSTA[1,2,b], VALENTIN PATILEA[3,c] and INGRID VAN KEILEGOM[1,a]

[1] *ORSTAT, KU Leuven, Belgium.* [a] *ingrid.vankeilegom@kuleuven.be*
[2] *Korteweg de Vries Institute for Mathematics, University of Amsterdam, Netherlands.* [b] *e.musta@uva.nl*
[3] *CREST, Ensai, France.* [c] *valentin.patilea@ensai.fr*

A challenge when dealing with survival analysis data is accounting for a cure fraction, meaning that some subjects will never experience the event of interest. Mixture cure models have been frequently used to estimate both the probability of being cured and the time to event for the susceptible subjects, by usually assuming a parametric (logistic) form of the incidence. We propose a new estimation procedure for a parametric cure rate that relies on a preliminary smooth estimator and is independent of the model assumed for the latency. On a second stage one can assume a semiparametric model for the latency and estimate also the survival distribution of the uncured subject. For the particular case of the logistic/Cox model, we investigate the theoretical properties of the estimators and show through simulations that presmoothing leads to more accurate results compared to the maximum likelihood estimator. To illustrate the practical use, we apply the new estimation procedure to two studies of melanoma survival data.

Keywords: Cure models; kernel smoothing; logistic model; survival analysis

References

[1] Aerts, M., Hens, N. and Simonoff, J.S. (2010). Model selection in regression based on pre-smoothing. *J. Appl. Stat.* **37** 1455–1472. MR2758660 https://doi.org/10.1080/02664760903046086
[2] Amico, M., Van Keilegom, I. and Legrand, C. (2019). The single-index/Cox mixture cure model. *Biometrics* **75** 452–462. MR3999169 https://doi.org/10.1111/biom.12999
[3] Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646
[4] Berkson, J. and Gage, R.P. (1952). Survival curve for cancer patients following treatment. *J. Amer. Statist. Assoc.* **47** 501–515.
[5] Boag, J.W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Roy. Statist. Soc. B* **11** 15–53.
[6] Burke, K. and Patilea, V. (2021). A likelihood-based approach for cure regression models. *TEST* **30** 693–712. MR4297274 https://doi.org/10.1007/s11749-020-00738-8
[7] Cai, C., Zou, Y., Peng, Y. and Zhang, J. (2012). smcure: An R-Package for estimating semiparametric mixture cure models. *Comput. Methods Programs Biomed.* **108** 1255–1260.
[8] Chen, X., Linton, O. and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71** 1591–1608. MR2000259 https://doi.org/10.1111/1468-0262.00461
[9] Cristóbal Cristóbal, J.A., Faraldo Roca, P. and González Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Ann. Statist.* **15** 603–609. MR0888428 https://doi.org/10.1214/aos/1176350363
[10] Du, Y. and Akritas, M.G. (2002). Uniform strong representation of the conditional Kaplan-Meier process. *Math. Methods Statist.* **11** 152–182. MR1941314
[11] Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1041–1046.

1350-7265

[12] Ferraty, F., González-Manteiga, W., Martínez-Calvo, A. and Vieu, P. (2012). Presmoothing in functional linear regression. *Statist. Sinica* **22** 69–94. MR2933168 https://doi.org/10.5705/ss.2010.085
[13] Fleming, T.R. and Harrington, D.P. (2011). *Counting Processes and Survival Analysis* **169**. New York: Wiley.
[14] Han, X. (2017). *Statistical Methods for Analysis of Genetic and Survival Data with Latent Heterogeneity*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.)—New York University. MR3755013
[15] Kuk, A.Y. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79** 531–541.
[16] Lee, T.E., Fisher, D.O., Blomberg, S.P. and Wintle, B.A. (2017). Extinct or still out there? Disentangling influences on extinction and rediscovery helps to clarify the fate of species on the edge. *Glob. Change Biol.* **23** 621–634.
[17] Li, C.-S. and Taylor, J.M. (2002). A semi-parametric accelerated failure time cure model. *Stat. Med.* **21** 3235–3247.
[18] Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Ann. Inst. Statist. Math.* **60** 545–574. MR2434411 https://doi.org/10.1007/s10463-007-0120-x
[19] Müller, U.U. and Van Keilegom, I. (2019). Goodness-of-fit tests for the cure rate in a mixture cure model. *Biometrika* **106** 211–227. MR3912392 https://doi.org/10.1093/biomet/asy058
[20] Musta, E., Patilea, V. and Van Keilegom, I. (2022). Supplement to “A presmoothing approach for estimation in the semiparametric Cox mixture cure model.” https://doi.org/10.3150/21-BEJ1434SUPP
[21] Patilea, V. and Van Keilegom, I. (2020). A general approach for cure models in survival analysis. *Ann. Statist.* **48** 2323–2346. MR4134797 https://doi.org/10.1214/19-AOS1889
[22] Peng, Y. and Dear, K.B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56** 237–243.
[23] Sposto, R. (2002). Cure model analysis in cancer: An application to data from the Children’s Cancer Group. *Stat. Med.* **21** 293–312. https://doi.org/10.1002/sim.987
[24] Stringer, S., Denys, D., Kahn, R.S. and Derks, E.M. (2016). What cure models can teach us about genome-wide survival analysis. *Behav. Genet.* **46** 269–280.
[25] Sy, J.P. and Taylor, J.M.G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56** 227–236. MR1767631 https://doi.org/10.1111/j.0006-341X.2000.00227.x
[26] Taylor, J.M. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics* **51** 899–907.
[27] Van Keilegom, I. and Akritas, M.G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.* **27** 1745–1784. MR1742508 https://doi.org/10.1214/aos/1017939150
[28] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes, with Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2
[29] Wycinka, E. and Jurkiewicz, T. (2017). Mixture cure models in prediction of time to default: Comparison with logit and Cox models. In *Contemporary Trends and Challenges in Finance* 221–231. Springer.
[30] Xu, J. and Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canad. J. Statist.* **42** 1–17. MR3181580 https://doi.org/10.1002/cjs.11197
[31] Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of “permanent employment” in Japan. *J. Amer. Statist. Assoc.* **87** 284–292.
[32] Zhang, J. and Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Stat. Med.* **26** 3157–3171. MR2380509 https://doi.org/10.1002/sim.2748

Bernoulli **28**(4), 2022, 2716–2743
https://doi.org/10.3150/21-BEJ1435

Asymptotic behavior of a low-temperature non-cascading 2-GREM dynamics at extreme time scales

LUIZ RENATO FONTES[1,a], SUSANA FRÓMETA[2,c] and LEONEL ZUAZNÁBAR[1,b]

[1]*USP - Instituto de Matemática e Estatística, R. do Matão, 1010 - Butantã, São Paulo - SP, CEP: 05508-090, Brasil.* [a]*lrfontes@usp.br,* [b]*lzuaznabar@ime.usp.br*
[2]*UFRGS - Instituto de Matemática e Estatística, Campus do Vale, Av. Bento Gonçalves, Porto Alegre - RS, 9500, CEP: 91509-900, Brasil.* [c]*susana.frometa@ufrgs.br*

We derive the scaling limit for the Hierarchical Random Hopping dynamics for the non cascading 2-GREM at low temperatures and time scales where the dynamics is close to equilibrium. The *fine tuning* phenomenon plays a role (under certain choices of parameters of the model), yielding three dynamical regimes. In contrast to the *cascading* case, the pairs of first and second level energies have fluctuations which scale with the volume, and this leads to a family of time scales where we see the dynamics moving through only a part of the full low-temperature energy landscape.

Keywords: GREM; random hopping dynamics; low temperature; fine tuning temperature; scaling limit; extreme time-scale; ergodic time-scale; K process; spin glasses

References

[1] Ben Arous, G., Bovier, A. and Černý, J. (2008). Universality of the REM for dynamics of mean-field spin glasses. *Comm. Math. Phys.* **282** 663–695. MR2426140 https://doi.org/10.1007/s00220-008-0565-7

[2] Ben Arous, G., Bovier, A. and Gayrard, V. (2003). Glauber dynamics of the random energy model. I. Metastable motion on the extreme states. *Comm. Math. Phys.* **235** 379–425. MR1974509 https://doi.org/10.1007/s00220-003-0798-4

[3] Ben Arous, G., Bovier, A. and Gayrard, V. (2003). Glauber dynamics of the random energy model. II. Aging below the critical temperature. *Comm. Math. Phys.* **236** 1–54. MR1977880 https://doi.org/10.1007/s00220-003-0799-3

[4] Ben Arous, G., Dembo, A. and Guionnet, A. (2001). Aging of spherical spin glasses. *Probab. Theory Related Fields* **120** 1–67. MR1856194 https://doi.org/10.1007/PL00008774

[5] Ben Arous, G. and Gayrard, V. (2008). Elementary potential theory on the hypercube. *Electron. J. Probab.* **13** 1726–1807. MR2448129 https://doi.org/10.1214/EJP.v13-527

[6] Ben Arous, G. and Gün, O. (2012). Universality and extremal aging for dynamics of spin glasses on subexponential time scales. *Comm. Pure Appl. Math.* **65** 77–127. MR2846638 https://doi.org/10.1002/cpa.20372

[7] Bezerra, S.C., Fontes, L.R.G., Gava, R.J., Gayrard, V. and Mathieu, P. (2012). Scaling limits and aging for asymmetric trap models on the complete graph and K processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **9** 303–321. MR3069367

[8] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[9] Bouchaud, J.P. and Dean, D.S. (1995). Aging on Parisi's tree. *J. Phys. I France* **5** 265–286.

[10] Bovier, A. and Gayrard, V. (2013). Convergence of clock processes in random environments and ageing in the p-spin SK model. *Ann. Probab.* **41** 817–847. MR3077527 https://doi.org/10.1214/11-AOP705

[11] Bovier, A. and Hartung, L. (2014). The extremal process of two-speed branching Brownian motion. *Electron. J. Probab.* **19** no. 18. MR3164771 https://doi.org/10.1214/EJP.v19-2982

1350-7265

[12] Bovier, A. and Kurkova, I. (2004). Derrida's generalised random energy models. I. Models with finitely many hierarchies. *Ann. Inst. Henri Poincaré Probab. Stat.* **40** 439–480. MR2070334 https://doi.org/10.1016/j.anihpb.2003.09.002

[13] Cortines, A., Gold, J. and Louidor, O. (2018). Dynamical freezing in a spin glass system with logarithmic correlations. *Electron. J. Probab.* **23** Paper No. 59, 31. MR3814253 https://doi.org/10.1214/18-ejp181

[14] Derrida, B. (1985). A generalization of the random energy model which includes correlations between energies. *J. Phys. Lett.* **46** 401–407.

[15] Fontes, L.R., Frómeta, S. Zuaznábar, L. (2022). Supplement to "Asymptotic behavior of a low-temperature non-cascading 2-GREM dynamics at extreme time scales." https://doi.org/10.3150/21-BEJ1435SUPP

[16] Fontes, L.R. and Gayrard, V. (2019). Asymptotic behavior and aging of a low temperature cascading 2-GREM dynamics at extreme time scales. *Electron. J. Probab.* **24** Paper No. 142. MR4049078 https://doi.org/10.1214/19-ejp395

[17] Fontes, L.R. and Peixoto, G.R.C. (2013). Elementary results on K processes with weights. *Markov Process. Related Fields* **19** 343–370. MR3113947

[18] Fontes, L.R. and Peixoto, G.R.C. (2021). Infinite level GREM-like K-processes existence and convergence. *J. Stat. Phys.* **182** Paper No. 50, 31. MR4224161 https://doi.org/10.1007/s10955-021-02713-5

[19] Fontes, L.R.G., Gava, R.J. and Gayrard, V. (2014). The K-process on a tree as a scaling limit of the GREM-like trap model. *Ann. Appl. Probab.* **24** 857–897. MR3178499 https://doi.org/10.1214/13-AAP937

[20] Fontes, L.R.G. and Lima, P.H.S. (2009). Convergence of symmetric trap models in the hypercube. In *New Trends in Mathematical Physics* (V. Sidoravičius, ed.) 285–297. Springer, Dordrecht.

[21] Fontes, L.R.G. and Mathieu, P. (2008). K-processes, scaling limit and aging for the trap model in the complete graph. *Ann. Probab.* **36** 1322–1358. MR2435851 https://doi.org/10.1214/07-AOP360

[22] Gayrard, V. and Gün, O. (2016). Aging in the GREM-like trap model. *Markov Process. Related Fields* **22** 165–202. MR3523983

[23] Jara, M., Landim, C. and Teixeira, A. (2011). Quenched scaling limits of trap models. *Ann. Probab.* **39** 176–223. MR2778800 https://doi.org/10.1214/10-AOP554

[24] Jara, M., Landim, C. and Teixeira, A. (2014). Universality of trap models in the ergodic time scale. *Ann. Probab.* **42** 2497–2557. MR3265173 https://doi.org/10.1214/13-AOP886

[25] Sasaki, M. and Nemoto, K. (2000). Analysis on aging in the generalized random energy model. *J. Phys. Soc. Jpn.* **69** 3045–3050.

Bernoulli **28**(4), 2022, 2744–2787
https://doi.org/10.3150/21-BEJ1436

Estimation of the ℓ_2-norm and testing in sparse linear regression with unknown variance

ALEXANDRA CARPENTIER[1,a], OLIVIER COLLIER[2,b], LAETITIA COMMINGES[3,c], ALEXANDRE B. TSYBAKOV[4,d] and YUHAO WANG[5,e]

[1]*Institut für Mathematik, Universität Potsdam, Potsdam, Germany.* [a]*carpentier@uni-potsdam.de*
[2]*Modal'X, Université Paris-Nanterre, Nanterre and CREST, Paris, France.* [b]*olivier.collier@parisnanterre.fr*
[3]*CEREMADE, Université Paris-Dauphine and CREST, Paris, France.* [c]*laetitia.comminges@dauphine.psl.eu*
[4]*CREST, ENSAE and Institut Polytechnique de Paris, Paris, France.* [d]*alexandre.tsybakov@ensae.fr*
[5]*Tsinghua University, Beijing, China and Shanghai Qi Zhi Institute, Shanghai, China.* [e]*yuhaow@tsinghua.edu.cn*

We consider the related problems of estimating the ℓ_2-norm and the squared ℓ_2-norm in sparse linear regression with unknown variance, as well as the problem of testing the hypothesis that the regression parameter is null under sparse alternatives with ℓ_2 separation. We establish the minimax optimal rates of estimation (respectively, testing) in these three problems.

Keywords: Sparse linear regression; signal detection; non-linear functional estimation

References

[1] Adamczak, R. and Wolff, P. (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** 531–586. MR3383337 https://doi.org/10.1007/s00440-014-0579-3
[2] Arias-Castro, E., Candès, E.J. and Plan, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 https://doi.org/10.1214/11-AOS910
[3] Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648
[4] Bellec, P.C., Lecué, G. and Tsybakov, A.B. (2018). Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. MR3852663 https://doi.org/10.1214/17-AOS1670
[5] 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
[6] Cai, T.T. and Low, M.G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. MR2253108 https://doi.org/10.1214/009053605000000147
[7] Carpentier, A., Collier, O., Comminges, L., Tsybakov, A.B. and Wang, Y. (2019). Minimax rate of testing in sparse linear regression. *Autom. Remote Control* **80** 1817–1834.
[8] Carpentier, A. and Verzelen, N. (2021). Optimal sparsity testing in linear regression model. *Bernoulli* **27** 727–750. MR4255213 https://doi.org/10.3150/20-bej1224
[9] Collier, O., Comminges, L. and Tsybakov, A.B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.* **45** 923–958. MR3662444 https://doi.org/10.1214/15-AOS1432
[10] Collier, O., Comminges, L., Tsybakov, A.B. and Verzelen, N. (2018). Optimal adaptive estimation of linear functionals under sparsity. *Ann. Statist.* **46** 3130–3150. MR3851767 https://doi.org/10.1214/17-AOS1653
[11] Comminges, L., Collier, O., Ndaoud, M. and Tsybakov, A.B. (2021). Adaptive robust estimation in sparse vector model. *Ann. Statist.* **49** 1347–1377. MR4298867 https://doi.org/10.1214/20-aos2002
[12] Comminges, L. and Dalalyan, A.S. (2013). Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *Electron. J. Stat.* **7** 146–190. MR3020417 https://doi.org/10.1214/13-EJS766

1350-7265

[13] Derumigny, A. (2018). Improved bounds for square-root lasso and square-root slope. *Electron. J. Stat.* **12** 741–766. MR3769194 https://doi.org/10.1214/18-EJS1410

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

[15] Donoho, D.L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323. MR1081043 https://doi.org/10.1016/0885-064X(90)90025-9

[16] Guo, Z., Wang, W., Cai, T.T. and Li, H. (2019). Optimal estimation of genetic relatedness in high-dimensional linear models. *J. Amer. Statist. Assoc.* **114** 358–369. MR3941260 https://doi.org/10.1080/01621459.2017.1407774

[17] Ingster, Y.I., Tsybakov, A.B. and Verzelen, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131 https://doi.org/10.1214/10-EJS589

[18] Ingster, Yu.I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

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

[20] Lecué, G. and Mendelson, S. (2017). Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc. (JEMS)* **19** 881–904. MR3612870 https://doi.org/10.4171/JEMS/682

[21] Mukherjee, R. and Sen, S. (2020). On minimax exponents of sparse testing. Preprint, arXiv:2003.00570.

[22] Rudelson, M. and Vershynin, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Not. IMRN* **19** 9594–9617. MR3431603 https://doi.org/10.1093/imrn/rnu243

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

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

[25] 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

[26] Verzelen, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electron. J. Stat.* **6** 38–90. MR2879672 https://doi.org/10.1214/12-EJS666

[27] Verzelen, N. and Gassiat, E. (2018). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli* **24** 3683–3710. MR3788186 https://doi.org/10.3150/17-BEJ975

Bernoulli **28**(4), 2022, 2788–2815
https://doi.org/10.3150/21-BEJ1437

A unified performance analysis of likelihood-informed subspace methods

TIANGANG CUI[1,a] and XIN T. TONG[2,b]

[1]*Monash University, School of Mathematics.* [a]*tiangang.cui@monash.edu*
[2]*National University of Singapore, Department of Mathematics.* [b]*mattxin@nus.edu.sg*

The likelihood-informed subspace (LIS) method offers a viable route to reducing the dimensionality of high-dimensional probability distributions arising in Bayesian inference. LIS identifies an intrinsic low-dimensional linear subspace where the target distribution differs the most from some tractable reference distribution. Such a subspace can be identified using the leading eigenvectors of a Gram matrix of the gradient of the log-likelihood function. Then, the original high-dimensional target distribution is approximated through various forms of marginalization of the likelihood function, in which the approximated likelihood only has support on the intrinsic low-dimensional subspace. This approximation enables the design of inference algorithms that can scale sub-linearly with the apparent dimensionality of the problem. Intuitively, the accuracy of the approximation, and hence the performance of the inference algorithms, are influenced by three factors—the dimension truncation error in identifying the subspace, Monte Carlo error in estimating the Gram matrices, and Monte Carlo error in constructing marginalizations. This work establishes a unified framework to analyze each of these three factors and their interplay. Under mild technical assumptions, we establish error bounds for a range of existing dimension reduction techniques based on the principle of LIS. Our error bounds also provide useful insights into the accuracy of these methods. In addition, we analyze the integration of LIS with sampling methods such as Markov Chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC). We also demonstrate the applicability of our analysis on a linear inverse problem with Gaussian prior, which shows that all the estimates can be dimension-independent if the prior covariance is a trace-class operator. Finally, we demonstrate various aspects of our theoretical claims on two nonlinear inverse problems.

Keywords: Dimension reduction; approximation error; likelihood informed subspace; Monte Carlo estimation

References

[1] Agapiou, S., Dashti, M. and Helin, T. (2021). Rates of contraction of posterior distributions based on p-exponential priors. *Bernoulli* **27** 1616–1642. MR4278794 https://doi.org/10.3150/20-bej1285
[2] Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D. and Stuart, A.M. (2017). Importance sampling: Intrinsic dimension and computational cost. *Statist. Sci.* **32** 405–431. MR3696003 https://doi.org/10.1214/17-STS611
[3] Agapiou, S., Roberts, G.O. and Vollmer, S.J. (2018). Unbiased Monte Carlo: Posterior estimation for intractable/infinite-dimensional models. *Bernoulli* **24** 1726–1786. MR3757514 https://doi.org/10.3150/16-BEJ911
[4] Andrieu, C. and Roberts, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648 https://doi.org/10.1214/07-AOS574
[5] Andrieu, C. and 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
[6] Beskos, A., Crisan, D. and Jasra, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.* **24** 1396–1445. MR3211000 https://doi.org/10.1214/13-AAP951
[7] Beskos, A., Girolami, M., Lan, S., Farrell, P.E. and Stuart, A.M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.* **335** 327–351. MR3612501 https://doi.org/10.1016/j.jcp.2016.12.041

[8] Beskos, A., Jasra, A., Law, K., Marzouk, Y. and Zhou, Y. (2018). Multilevel sequential Monte Carlo with dimension-independent likelihood-informed proposals. *SIAM/ASA J. Uncertain. Quantificat.* **6** 762–786. MR3810512 https://doi.org/10.1137/17M1120993

[9] Bigoni, D., Zahm, O., Spantini, A. and Marzouk, Y. (2019). Greedy inference with layers of lazy maps. ArXiv Preprint. Available at arXiv:1906.00031.

[10] Bobkov, S. and Ledoux, M. (1997). Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** 383–400. MR1440138 https://doi.org/10.1007/s004400050090

[11] Bobkov, S.G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27** 1903–1921. MR1742893 https://doi.org/10.1214/aop/1022874820

[12] Bobkov, S.G. and Ledoux, M. (2000). From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.* **10** 1028–1052. MR1800062 https://doi.org/10.1007/PL00001645

[13] Bobkov, S.G. and Ledoux, M. (2009). Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Probab.* **37** 403–427. MR2510011 https://doi.org/10.1214/08-AOP407

[14] Brascamp, H.J. and Lieb, E.H. (1976). On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480 https://doi.org/10.1016/0022-1236(76)90004-5

[15] Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G. and Wilcox, L.C. (2012). Extreme-scale UQ for Bayesian inverse problems governed by PDEs. In *SC'12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis* 1–11. IEEE.

[16] Bui-Thanh, T., Ghattas, O., Martin, J. and Stadler, G. (2013). A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion. *SIAM J. Sci. Comput.* **35** A2494–A2523. MR3126997 https://doi.org/10.1137/12089586X

[17] Cai, T.T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496 https://doi.org/10.1214/009053606000000830

[18] Constantine, P.G., Kent, C. and Bui-Thanh, T. (2016). Accelerating Markov chain Monte Carlo with active subspaces. *SIAM J. Sci. Comput.* **38** A2779–A2805. MR3543164 https://doi.org/10.1137/15M1042127

[19] Cotter, S.L., Roberts, G.O., Stuart, A.M. and White, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540 https://doi.org/10.1214/13-STS421

[20] Cui, T. and Dolgov, S. (2022). Deep composition of tensor trains using squared inverse Rosenblatt transports. *Found. Comput. Math.* To appear. https://doi.org/10.1007/s10208-021-09537-5

[21] Cui, T., Fox, C. and O'Sullivan, M.J. (2011). Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. *Water Resour. Res.* **47**.

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

[23] Cui, T., Martin, J., Marzouk, Y.M., Solonen, A. and Spantini, A. (2014). Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Probl.* **30** 114015. MR3274599 https://doi.org/10.1088/0266-5611/30/11/114015

[24] Cui, T. and Tong, X.T. (2022). Supplement to "A unified performance analysis of likelihood-informed subspace methods." https://doi.org/10.3150/21-BEJ1437SUPP

[25] Cui, T. and Zahm, O. (2021). Data-free likelihood-informed dimension reduction of Bayesian inverse problems. *Inverse Probl.* **37** Paper No. 045009. MR4234453 https://doi.org/10.1088/1361-6420/abeafb

[26] Dashti, M. and Stuart, A.M. (2011). Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal.* **49** 2524–2542. MR2873245 https://doi.org/10.1137/100814664

[27] Detommaso, G., Cui, T., Marzouk, Y., Spantini, A. and Scheichl, R. (2018). A Stein variational Newton method. *Adv. Neural Inf. Process. Syst.* **31** 9169–9179.

[28] Dodwell, T.J., Ketelsen, C., Scheichl, R. and Teckentrup, A.L. (2019). Multilevel Markov chain Monte Carlo. *SIAM Rev.* **61** 509–545. MR3989241 https://doi.org/10.1137/19M126966X

[29] Drineas, P. and Ipsen, I.C.F. (2019). Low-rank matrix approximations do not need a singular value gap. *SIAM J. Matrix Anal. Appl.* **40** 299–319. MR3916954 https://doi.org/10.1137/18M1163658

[30] Flath, H.P., Wilcox, L.C., Akçelik, V., Hill, J., van Bloemen Waanders, B. and Ghattas, O. (2011). Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank

partial Hessian approximations. *SIAM J. Sci. Comput.* **33** 407–432. MR2783201 https://doi.org/10.1137/090780717

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

[32] Haario, H., Laine, M., Lehtinen, M., Saksman, E. and Tamminen, J. (2004). Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 591–607. MR2088292 https://doi.org/10.1111/j.1467-9868.2004.02053.x

[33] Hall, P. and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269 https://doi.org/10.1214/009053606000000957

[34] Iglesias, M.A., Lin, K. and Stuart, A.M. (2014). Well-posed Bayesian geometric inverse problems arising in subsurface flow. *Inverse Probl.* **30** 114001. MR3274585 https://doi.org/10.1088/0266-5611/30/11/114001

[35] Kaipio, J.P., Kolehmainen, V., Somersalo, E. and Vauhkonen, M. (2000). Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Probl.* **16** 1487–1522. MR1800606 https://doi.org/10.1088/0266-5611/16/5/321

[36] Karhunen, K. (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae Ser. A. I. Math.-Phys.* **1947** 79. MR0023013

[37] Ledoux, M. (1994). A simple analytic proof of an inequality by P. Buser. *Proc. Amer. Math. Soc.* **121** 951–959. MR1186991 https://doi.org/10.2307/2160298

[38] Lie, H.C., Sullivan, T.J. and Teckentrup, A. (2021). Error bounds for some approximate posterior measures in Bayesian inference. In *Numerical Mathematics and Advanced Applications—ENUMATH 2019*. *Lecture Notes in Computational Science and Engineering* **139** 275–283. Cham: Springer. MR4266506 https://doi.org/10.1007/978-3-030-55874-1_26

[39] Liu, Q. and Wang, D. (2016). Stein variational gradient descent: A general purpose Bayesian inference algorithm. *Adv. Neural Inf. Process. Syst.* **29** 2378–2386.

[40] Loève, M. (1978). *Probability Theory. II*, 4th ed. *Graduate Texts in Mathematics, Vol. 46*. New York: Springer. MR0651018

[41] Martin, J., Wilcox, L.C., Burstedde, C. and Ghattas, O. (2012). A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J. Sci. Comput.* **34** A1460–A1487. MR2970260 https://doi.org/10.1137/110845598

[42] Marzouk, Y., Moselhy, T., Parno, M. and Spantini, A. (2017). Sampling via measure transport: An introduction. In *Handbook of Uncertainty Quantification. Vol. 1, 2, 3* 785–825. Cham: Springer. MR3821485

[43] Morzfeld, M., Tong, X.T. and Marzouk, Y.M. (2019). Localization for MCMC: Sampling high-dimensional posterior distributions with local structure. *J. Comput. Phys.* **380** 1–28. MR3901708 https://doi.org/10.1016/j.jcp.2018.12.008

[44] Murray, I., MacKay, D. and Adams, R.P. (2008). The Gaussian process density sampler. *Adv. Neural Inf. Process. Syst.* **21** 9–16.

[45] Otto, F. and Villani, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620 https://doi.org/10.1006/jfan.1999.3557

[46] Petra, N., Martin, J., Stadler, G. and Ghattas, O. (2014). A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems. *SIAM J. Sci. Comput.* **36** A1525–A1555. MR3233941 https://doi.org/10.1137/130934805

[47] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993

[48] Rudolf, D. and Sprungk, B. (2018). On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm. *Found. Comput. Math.* **18** 309–343. MR3777781 https://doi.org/10.1007/s10208-016-9340-x

[49] Sanz-Alonso, D. (2018). Importance sampling and necessary sample size: An information theory approach. *SIAM/ASA J. Uncertain. Quantificat.* **6** 867–879. MR3813241 https://doi.org/10.1137/16M1093549

[50] Spantini, A., Bigoni, D. and Marzouk, Y. (2018). Inference via low-dimensional couplings. *J. Mach. Learn. Res.* **19** Paper No. 71. MR3899768

[51] Spantini, A., Solonen, A., Cui, T., Martin, J., Tenorio, L. and Marzouk, Y. (2015). Optimal low-rank approximations of Bayesian linear inverse problems. *SIAM J. Sci. Comput.* **37** A2451–A2487. MR3418226 https://doi.org/10.1137/140977308

[52] Stewart, G.W. (1980). The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM J. Numer. Anal.* **17** 403–409. MR0581487 https://doi.org/10.1137/0717034

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

[54] Tabak, E.G., Trigila, G. and Zhao, W. (2020). Conditional density estimation and simulation through optimal transport. *Mach. Learn.* **109** 665–688. MR4094530 https://doi.org/10.1007/s10994-019-05866-3

[55] Tabak, E.G. and Turner, C.V. (2013). A family of nonparametric density estimation algorithms. *Comm. Pure Appl. Math.* **66** 145–164. MR2999294 https://doi.org/10.1002/cpa.21423

[56] Teixeira Parente, M., Wallin, J. and Wohlmuth, B. (2020). Generalized bounds for active subspaces. *Electron. J. Stat.* **14** 917–943. MR4065177 https://doi.org/10.1214/20-EJS1684

[57] Tong, X.T., Morzfeld, M. and Marzouk, Y.M. (2020). MALA-within-Gibbs samplers for high-dimensional distributions with sparse conditional structure. *SIAM J. Sci. Comput.* **42** A1765–A1788. MR4111677 https://doi.org/10.1137/19M1284014

[58] Trigila, G. and Tabak, E.G. (2016). Data-driven optimal transport. *Comm. Pure Appl. Math.* **69** 613–648. MR3465084 https://doi.org/10.1002/cpa.21588

[59] Yu, Y., Wang, T. and Samworth, R.J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika* **102** 315–323. MR3371006 https://doi.org/10.1093/biomet/asv008

[60] Zahm, O., Cui, T., Law, K., Spantini, A. and Marzouk, Y. (2018). Certified dimension reduction in nonlinear Bayesian inverse problems. ArXiv Preprint. Available at arXiv:1807.03712.

Bernoulli **28**(4), 2022, 2816–2832
https://doi.org/10.3150/21-BEJ1438

Measuring association with Wasserstein distances

JOHANNES C.W. WIESEL[a]

Columbia University, Department of Statistics, 1255 Amsterdam Avenue, New York, NY 10027, USA.
[a]*johannes.wiesel@columbia.edu*

Let $\pi \in \Pi(\mu, \nu)$ be a coupling between two probability measures μ and ν on a Polish space. In this article we propose and study a class of nonparametric measures of association between μ and ν, which we call Wasserstein correlation coefficients. These coefficients are based on the Wasserstein distance between ν and the disintegration π_{x_1} of π with respect to the first coordinate. We also establish basic statistical properties of this new class of measures: we develop a statistical theory for strongly consistent estimators and determine their convergence rate in the case of compactly supported measures μ and ν. Throughout our analysis we make use of the so-called adapted/bicausal Wasserstein distance, in particular we rely on results established in [Backhoff, Bartl, Beiglböck, Wiesel. Estimating processes in adapted Wasserstein distance. 2020]. Our approach applies to probability laws on general Polish spaces.

Keywords: Independence; measure of association; correlation; optimal transport; empirical measure; (bicausal/adapted) Wasserstein distance

References

[1] Backhoff, J., Bartl, D., Beiglböck, M. and Wiesel, J. (2021+). Estimating processes in adapted Wasserstein distance. *Ann. Appl. Probab.* Avaible at arXiv:2002.07261.
[2] Backhoff-Veraguas, J., Bartl, D., Beiglböck, M. and Eder, M. (2020). Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.* **24** 601–632. MR4118990 https://doi.org/10.1007/s00780-020-00426-3
[3] Blum, J.R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055
[4] Cao, S. and Bickel, P.J. (2020). Correlations with tailored extremal properties. arXiv preprint. Available at arXiv:2008.10177.
[5] Chatterjee, S. (2020). A new coefficient of correlation. *J. Amer. Statist. Assoc.* 1–21.
[6] Cifarelli, D.M. and Regazzini, E. (2017). On the centennial anniversary of Gini's theory of statistical relations. *Metron* **75** 227–242. MR3695007 https://doi.org/10.1007/s40300-017-0108-0
[7] Gini, C. (1915). Nuovi contributi alla teoria delle relazioni statistiche. *Premiate Officine Grafiche di Carlo Ferrari*.
[8] Deb, N., Ghosal, P. and Sen, B. (2020). Measuring association on topological spaces using kernels and geometric graphs. arXiv preprint. Available at arXiv:2010.01768.
[9] Dette, H., Siburg, K.F. and Stoimenov, P.A. (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.* **40** 21–41. MR3024030 https://doi.org/10.1111/j.1467-9469.2011.00767.x
[10] Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7
[11] Friedman, J.H. and Rafsky, L.C. (1983). Graph-theoretic measures of multivariate association and prediction. *Ann. Statist.* **11** 377–391. MR0696054
[12] Gamboa, F., Klein, T. and Lagnoux, A. (2018). Sensitivity analysis based on Cramér–von Mises distance. *SIAM/ASA J. Uncertain. Quantificat.* **6** 522–548. MR3784121 https://doi.org/10.1137/15M1025621
[13] Geenens, G. and Lafaye de Micheaux, P. (2020). The Hellinger correlation. *J. Amer. Statist. Assoc.* 1–15.

[14] Giacomo Nies, T., Staudt, T. and Munk, A. (2021). Transport dependency: Optimal transport based dependency measures. arXiv preprint. Available at arXiv:2105.02073.

[15] Griessenberger, F., Junker, R.R. and Trutschnig, W. (2021). On a multivariate copula-based dependence measure and its estimation. arXiv preprint. Available at arXiv:2109.12883.

[16] Junker, R.R., Griessenberger, F. and Trutschnig, W. (2021). Estimating scale-invariant directed dependence of bivariate distributions. *Comput. Statist. Data Anal.* **153** 107058. MR4141460 https://doi.org/10.1016/j.csda.2020.107058

[17] Ke, C. and Yin, X. (2020). Expected conditional characteristic function-based measures for testing independence. *J. Amer. Statist. Assoc.* **115** 985–996. MR4107694 https://doi.org/10.1080/01621459.2019.1604364

[18] Lassalle, R. (2018). Causal transport plans and their Monge-Kantorovich problems. *Stoch. Anal. Appl.* **36** 452–484. MR3784142 https://doi.org/10.1080/07362994.2017.1422747

[19] Linfoot, E.H. (1957). An informational measure of correlation. *Inf. Control* **1** 85–89. MR0092706

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

[21] Mordant, G. and Segers, J. (2022). Measuring dependence between random vectors via optimal transport. *J. Multivariate Anal.* **189** 104912. MR4349898 https://doi.org/10.1016/j.jmva.2021.104912

[22] Móri, T.F. and Székely, G.J. (2019). Four simple axioms of dependence measures. *Metrika* **82** 1–16. MR3897521 https://doi.org/10.1007/s00184-018-0670-3

[23] Móri, T.F. and Székely, G.J. (2020). The Earth mover's correlation. *Ann. Univ. Sci. Budapest. Sect. Comput.* **50** 249–268. MR4163686 https://doi.org/10.1007/s10701-018-0208-6

[24] Ozair, S., Lynch, C., Bengio, Y., van den Oord, A., Levine, S. and Sermanet, P. (2019). Wasserstein dependency measure for representation learning. arXiv preprint. Available at arXiv:1903.11780.

[25] Pflug, G.C. (2009). Version-independence and nested distributions in multistage stochastic optimization. *SIAM J. Optim.* **20** 1406–1420. MR2587733 https://doi.org/10.1137/080718401

[26] Pflug, G.C. and Pichler, A. (2012). A distance for multistage stochastic optimization models. *SIAM J. Optim.* **22** 1–23. MR2902682 https://doi.org/10.1137/110825054

[27] Puccetti, G. (2019). Measuring linear correlation between random vectors. Available at SSRN 3116066.

[28] Rényi, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hung.* **10** 441–451. MR0115203 https://doi.org/10.1007/BF02024507

[29] Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1–14. MR0428579

[30] Scarsini, M. (1984). On measures of concordance. *Stochastica* **8** 201–218. MR0796650

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

[32] Shi, H., Drton, M. and Han, F. (2020). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* 1–16.

[33] Shi, H., Drton, M. and Han, F. (2020). On the power of Chatterjee rank correlation. arXiv preprint. Available at arXiv:2008.11619.

[34] Sommerfeld, M. and Munk, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 219–238. MR3744719 https://doi.org/10.1111/rssb.12236

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

[36] Trutschnig, W. (2011). On a strong metric on the space of copulas and its induced dependence measure. *J. Math. Anal. Appl.* **384** 690–705. MR2825218 https://doi.org/10.1016/j.jmaa.2011.06.013

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

[38] Wiesel, J.C.W. (2022). Supplement to "Measuring association with Wasserstein distances." https://doi.org/10.3150/21-BEJ1438SUPP

[39] Xiao, Y. and Wang, W.Y. (2019). Disentangled representation learning with Wasserstein total correlation. arXiv preprint. Available at arXiv:1912.12818.

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

Bernoulli **28**(4), 2022, 2833–2861
https://doi.org/10.3150/21-BEJ1439

Local scaling limits of Lévy driven fractional random fields

VYTAUTĖ PILIPAUSKAITĖ[1,a] and DONATAS SURGAILIS[2,b]

[1] *University of Luxembourg, Department of Mathematics, 6 Avenue de la Fonte, 4364 Esch-sur-Alzette, Luxembourg.* [a] *vytaute.pilipauskaite@gmail.com*
[2] *Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, 03225 Vilnius, Lithuania.* [b] *donatas.surgailis@mif.vu.lt*

We obtain a complete description of local anisotropic scaling limits for a class of fractional random fields X on $\mathbb{R}^2$ written as stochastic integral with respect to infinitely divisible random measure. The scaling procedure involves increments of X over points the distance between which in the horizontal and vertical directions shrinks as $O(\lambda)$ and $O(\lambda^\gamma)$ respectively as $\lambda \downarrow 0$, for some $\gamma > 0$. We consider two types of increments of X: usual increment and rectangular increment, leading to the respective concepts of γ-tangent and γ-rectangent random fields. We prove that for above X both types of local scaling limits exist for any $\gamma > 0$ and undergo a transition, being independent of $\gamma > \gamma_0$ and $\gamma < \gamma_0$, for some $\gamma_0 > 0$; moreover, the 'unbalanced' scaling limits ($\gamma \neq \gamma_0$) are (H_1, H_2)-multi self-similar with one of H_i, $i = 1,2$, equal to 0 or 1. The paper extends Pilipauskaitė and Surgailis (*Stochastic Process. Appl.* **127** (2017) 2751–2779) and Surgailis (*Stochastic Process. Appl.* **130** (2020) 7518–7546) on large-scale anisotropic scaling of random fields on $\mathbb{Z}^2$ and Benassi et al. (*Bernoulli* **10** (2004) 357–373) on 1-tangent limits of isotropic fractional Lévy random fields.

Keywords: Fractional random field; local anisotropic scaling limit; rectangular increment; Lévy random measure; scaling transition; multi self-similar random field

References

[1] Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* U.S. Government printing office, Washington, D.C.

[2] Ayache, A., Leger, S. and Pontier, M. (2002). Drap brownien fractionnaire. *Potential Anal.* **17** 31–43. MR1906407 https://doi.org/10.1023/A:1015260803576

[3] Bardet, J.-M. and Surgailis, D. (2011). Measuring the roughness of random paths by increment ratios. *Bernoulli* **17** 749–780. MR2787614 https://doi.org/10.3150/10-BEJ291

[4] Barndorff-Nielsen, O.E., Benth, F.E. and Veraart, A.E.D. (2018). *Ambit Stochastics*. *Probability Theory and Stochastic Modelling* **88**. Cham: Springer. MR3839270 https://doi.org/10.1007/978-3-319-94129-5

[5] Basse-O'Connor, A., Graversen, S.-E. and Pedersen, J. (2012). Multiparameter processes with stationary increments: Spectral representation and integration. *Electron. J. Probab.* **17** no. 74. MR2968681 https://doi.org/10.1214/EJP.v17-2287

[6] Basse-O'Connor, A., Lachièze-Rey, R. and Podolskij, M. (2017). Power variation for a class of stationary increments Lévy driven moving averages. *Ann. Probab.* **45** 4477–4528. MR3737916 https://doi.org/10.1214/16-AOP1170

[7] Basse-O'Connor, A., Pilipauskaitė, V. and Podolskij, M. (2021). Power variations for fractional type infinitely divisible random fields. *Electron. J. Probab.* **26** Paper No. 55. MR4254797 https://doi.org/10.1214/21-ejp617

[8] Benassi, A., Cohen, S. and Istas, J. (2004). On roughness indices for fractional fields. *Bernoulli* **10** 357–373. MR2046778 https://doi.org/10.3150/bj/1082380223

[9] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Stat.* **42** 1656–1670. MR0383482 https://doi.org/10.1214/aoms/1177693164

[10] Biermé, H., Durieu, O. and Wang, Y. (2017). Invariance principles for operator-scaling Gaussian random fields. *Ann. Appl. Probab.* **27** 1190–1234. MR3655864 https://doi.org/10.1214/16-AAP1229

[11] Biermé, H., Meerschaert, M.M. and Scheffler, H.-P. (2007). Operator scaling stable random fields. *Stochastic Process. Appl.* **117** 312–332. MR2290879 https://doi.org/10.1016/j.spa.2006.07.004

[12] Bolin, D. (2014). Spatial Matérn fields driven by non-Gaussian noise. *Scand. J. Stat.* **41** 557–579. MR3249417 https://doi.org/10.1111/sjos.12046

[13] Cohen, S. (2012). Fractional Lévy fields. In *Lévy Matters II*. *Lecture Notes in Math.* **2061** 1–95. Heidelberg: Springer. MR3014146 https://doi.org/10.1007/978-3-642-31407-0_1

[14] Cohen, S. and Istas, J. (2013). *Fractional Fields and Applications*. *Mathématiques & Applications (Berlin) [Mathematics & Applications]* **73**. Heidelberg: Springer. MR3088856 https://doi.org/10.1007/978-3-642-36739-7

[15] Falconer, K.J. (2002). Tangent fields and the local structure of random fields. *J. Theoret. Probab.* **15** 731–750. MR1922445 https://doi.org/10.1023/A:1016276016983

[16] Fyodorov, Y.V., Khoruzhenko, B.A. and Simm, N.J. (2016). Fractional Brownian motion with Hurst index $H = 0$ and the Gaussian unitary ensemble. *Ann. Probab.* **44** 2980–3031. MR3531684 https://doi.org/10.1214/15-AOP1039

[17] Genton, M.G., Perrin, O. and Taqqu, M.S. (2007). Self-similarity and Lamperti transformation for random fields. *Stoch. Models* **23** 397–411. MR2341075 https://doi.org/10.1080/15326340701471018

[18] Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Table of Integrals, Series, and Products*, 6th ed. San Diego, CA: Academic Press. MR1773820

[19] Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics. XLIX. On the Matérn correlation family. *Biometrika* **93** 989–995. MR2285084 https://doi.org/10.1093/biomet/93.4.989

[20] Hansen, L.V. and Thorarinsdottir, T.L. (2013). A note on moving average models for Gaussian random fields. *Statist. Probab. Lett.* **83** 850–855. MR3040314 https://doi.org/10.1016/j.spl.2012.12.009

[21] Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff Publishing. MR0322926

[22] Jónsdóttir, K.Ý., Rønn-Nielsen, A., Mouridsen, K. and Jensen, E.B.V. (2013). Lévy-based modelling in brain imaging. *Scand. J. Stat.* **40** 511–529. MR3091695 https://doi.org/10.1002/sjos.12000

[23] Kelbert, M.Y., Leonenko, N.N. and Ruiz-Medina, M.D. (2005). Fractional random fields associated with stochastic fractional heat equations. *Adv. in Appl. Probab.* **37** 108–133. MR2135156 https://doi.org/10.1239/aap/1113402402

[24] Lamperti, J. (1962). Semi-stable stochastic processes. *Trans. Amer. Math. Soc.* **104** 62–78. MR0138128 https://doi.org/10.2307/1993933

[25] Lee, J. (2021). Hurst estimation for operator scaling random fields. *Statist. Probab. Lett.* **178** Paper No. 109188. MR4291555 https://doi.org/10.1016/j.spl.2021.109188

[26] Leonenko, N.N., Ruiz-Medina, M.D. and Taqqu, M.S. (2011). Fractional elliptic, hyperbolic and parabolic random fields. *Electron. J. Probab.* **16** 1134–1172. MR2820073 https://doi.org/10.1214/EJP.v16-891

[27] Lodhia, A., Sheffield, S., Sun, X. and Watson, S.S. (2016). Fractional Gaussian fields: A survey. *Probab. Surv.* **13** 1–56. MR3466837 https://doi.org/10.1214/14-PS243

[28] Loh, W.-L. (2015). Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. *Ann. Statist.* **43** 2766–2794. MR3405611 https://doi.org/10.1214/15-AOS1365

[29] Pakkanen, M.S. (2014). Limit theorems for power variations of ambit fields driven by white noise. *Stochastic Process. Appl.* **124** 1942–1973. MR3170230 https://doi.org/10.1016/j.spa.2014.01.005

[30] Pakkanen, M.S. and Réveillac, A. (2016). Functional limit theorems for generalized variations of the fractional Brownian sheet. *Bernoulli* **22** 1671–1708. MR3474829 https://doi.org/10.3150/15-BEJ707

[31] Pilipauskaitė, V. and Surgailis, D. (2014). Joint temporal and contemporaneous aggregation of random-coefficient AR(1) processes. *Stochastic Process. Appl.* **124** 1011–1035. MR3138604 https://doi.org/10.1016/j.spa.2013.10.004

[32] Pilipauskaitė, V. and Surgailis, D. (2016). Anisotropic scaling of the random grain model with application to network traffic. *J. Appl. Probab.* **53** 857–879. MR3570099 https://doi.org/10.1017/jpr.2016.45

[33] Pilipauskaitė, V. and Surgailis, D. (2017). Scaling transition for nonlinear random fields with long-range dependence. *Stochastic Process. Appl.* **127** 2751–2779. MR3660890 https://doi.org/10.1016/j.spa.2016.12.011

[34] Pilipauskaitė, V. and Surgailis, D. (2021). Scaling limits of linear random fields on $\mathbb{Z}^2$ with general dependence axis. In *In and Out of Equilibrium. 3. Celebrating Vladas Sidoravicius*. *Progress in Probability* **77** 683–710. Cham: Birkhäuser/Springer. MR4237288

[35] Pratt, J.W. (1960). On interchanging limits and integrals. *Ann. Math. Stat.* **31** 74–77. MR0123673 https://doi.org/10.1214/aoms/1177705988

[36] Puplinskaitė, D. and Surgailis, D. (2015). Scaling transition for long-range dependent Gaussian random fields. *Stochastic Process. Appl.* **125** 2256–2271. MR3322863 https://doi.org/10.1016/j.spa.2014.12.011

[37] Puplinskaitė, D. and Surgailis, D. (2016). Aggregation of autoregressive random fields and anisotropic long-range dependence. *Bernoulli* **22** 2401–2441. MR3498033 https://doi.org/10.3150/15-BEJ733

[38] Rajput, B.S. and Rosiński, J. (1989). Spectral representations of infinitely divisible processes. *Probab. Theory Related Fields* **82** 451–487. MR1001524 https://doi.org/10.1007/BF00339998

[39] Samorodnitsky, G. (2016). *Stochastic Processes and Long Range Dependence*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3561100 https://doi.org/10.1007/978-3-319-45575-4

[40] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*. *Stochastic Modeling*. New York: CRC Press. MR1280932

[41] Surgailis, D. (2019). Anisotropic scaling limits of long-range dependent linear random fields on $\mathbb{Z}^3$. *J. Math. Anal. Appl.* **472** 328–351. MR3906377 https://doi.org/10.1016/j.jmaa.2018.11.027

[42] Surgailis, D. (2020). Scaling transition and edge effects for negatively dependent linear random fields on $\mathbb{Z}^2$. *Stochastic Process. Appl.* **130** 7518–7546. MR4167214 https://doi.org/10.1016/j.spa.2020.08.005

[43] Takenaka, S. (1991). Integral-geometric construction of self-similar stable processes. *Nagoya Math. J.* **123** 1–12. MR1126180 https://doi.org/10.1017/S0027763000003627

[44] Wallin, J. and Bolin, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scand. J. Stat.* **42** 872–890. MR3391697 https://doi.org/10.1111/sjos.12141

[45] Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Int. Stat. Inst.* **40** 974–994. MR0173287

Bernoulli **28**(4), 2022, 2862–2873
https://doi.org/10.3150/21-BEJ1440

Robust importance sampling with adaptive winsorization

PAULO ORENSTEIN[a]

Instituto de Matemática Pura e Aplicada, Estr. Dona Castorina, 110, Rio de Janeiro 22460-320, Brazil.
[a]*pauloo@impa.br*

Importance sampling is a widely used technique to estimate properties of a distribution. The resulting estimator is unbiased but may have huge, potentially infinite, variance. This paper proposes trading-off some bias for variance by adaptively winsorizing the importance sampling estimator. The novel procedure is based on the Balancing Principle (or Lepskii's Method). As a consequence, it offers a principled way to perform winsorization with finite-sample guarantees in the form of an oracle inequality. In various examples, the proposed estimator is shown to have smaller mean squared error and mean absolute deviation than leading alternatives such as the traditional importance sampling estimator or winsorizing it via cross-validation.

Keywords: Importance sampling; Winsorization; Balancing Principle; Lepskii's method; robustness

References

[1] Bertin, K., Lacour, C. and Rivoirard, V. (2016). Adaptive pointwise estimation of conditional density function. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 939–980. MR3498017 https://doi.org/10.1214/14-AIHP665
[2] Bousquet-Mélou, M. (2014). On the importance sampling of self-avoiding walks. *Combin. Probab. Comput.* **23** 725–748. MR3249222 https://doi.org/10.1017/S0963548313000643
[3] Chan, J.C., Hou, C. and Yang, T.T. (2020). Robust estimation and inference for importance sampling estimators with infinite variance. In *Essays in Honor of Cheng Hsiao* Emerald Publishing Limited.
[4] Delyon, B. and Portier, F. (2016). Integral approximation by kernel smoothing. *Bernoulli* **22** 2177–2208. MR3498027 https://doi.org/10.3150/15-BEJ725
[5] Fithian, W. and Wager, S. (2015). Semiparametric exponential families for heavy-tailed data. *Biometrika* **102** 486–493. MR3371018 https://doi.org/10.1093/biomet/asu065
[6] Hämarik, U. and Raus, T. (2009). About the balancing principle for choice of the regularization parameter. *Numer. Funct. Anal. Optim.* **30** 951–970. MR2589759 https://doi.org/10.1080/01630560903393139
[7] Ionides, E.L. (2008). Truncated importance sampling. *J. Comput. Graph. Statist.* **17** 295–311. MR2439961 https://doi.org/10.1198/106186008X320456
[8] Knuth, D.E. (1976). Mathematics and computer science: Coping with finiteness. *Science* **194** 1235–1242. MR0534161 https://doi.org/10.1126/science.194.4271.1235
[9] Lacour, C., Massart, P. and Rivoirard, V. (2017). Estimator selection: A new method with applications to kernel density estimation. *Sankhya A* **79** 298–335. MR3707423 https://doi.org/10.1007/s13171-017-0107-5
[10] Lepskiĭ, O.V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. MR1091202 https://doi.org/10.1137/1135065
[11] Madras, N. and Slade, G. (2013). *The Self-Avoiding Walk*. *Modern Birkhäuser Classics*. New York: Birkhäuser/Springer. MR2986656 https://doi.org/10.1007/978-1-4614-6025-1
[12] Mathé, P. (2006). The Lepskiĭ principle revisited. *Inverse Probl.* **22** L11–L15. MR2235633 https://doi.org/10.1088/0266-5611/22/3/L02
[13] Northrop, P.J., Attalides, N. and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 93–120. MR3611679 https://doi.org/10.1111/rssc.12159

[14] Orenstein, P. (2022). Supplement to "Robust importance sampling with adaptive winsorization." https://doi.org/10.3150/21-BEJ1440SUPPA, https://doi.org/10.3150/21-BEJ1440SUPPB
[15] Riihimäki, J. and Vehtari, A. (2014). Laplace approximation for logistic Gaussian process density estimation and regression. *Bayesian Anal.* **9** 425–447. MR3217002 https://doi.org/10.1214/14-BA872
[16] Sen, R., Shanmugam, K., Dimakis, A.G. and Shakkottai, S. (2017). Identifying best interventions through online importance sampling. In *International Conference on Machine Learning* 3057–3066.
[17] Vehtari, A., Gelman, A. and Gabry, J. (2015). Pareto smoothed importance sampling. Preprint. Available at: arXiv:1507.02646.

Bernoulli **28**(4), 2022, 2874–2908
https://doi.org/10.3150/21-BEJ1441

Upper functions for sample paths of Lévy(-type) processes

FRANZISKA KÜHN[a]

TU Dresden, Fakultät Mathematik, Institut für Mathematische Stochastik, 01062 Dresden, Germany.
[a]*franziska.kuehn1@tu-dresden.de*

We study the small-time asymptotics of sample paths of Lévy processes and Lévy-type processes. Namely, we investigate under which conditions the limit

$$\limsup_{t\to 0} \frac{1}{f(t)} |X_t - X_0|$$

is finite resp. infinite with probability 1. We establish integral criteria in terms of the infinitesimal characteristics and the symbol of the process. Our results apply to a wide class of processes, including solutions to Lévy-driven SDEs and stable-like processes. For the particular case of Lévy processes, we recover and extend earlier results from the literature. Moreover, we present a new maximal inequality for Lévy-type processes, which is of independent interest.

Keywords: Lévy process; Feller process; martingale problem; sample path behaviour; small-time asymptotics; upper function; maximal inequality

References

[1] Bass, R.F. (1988). Uniqueness in law for pure jump Markov processes. *Probab. Theory Related Fields* **79** 271–287. MR0958291 https://doi.org/10.1007/BF00320922

[2] Bass, R.F. and Tang, H. (2009). The martingale problem for a class of stable-like processes. *Stochastic Process. Appl.* **119** 1144–1167. MR2508568 https://doi.org/10.1016/j.spa.2008.06.003

[3] Bertoin, J. (1996). *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge: Cambridge Univ. Press. MR1406564

[4] Bertoin, J., Doney, R.A. and Maller, R.A. (2008). Passage of Lévy processes across power law boundaries at small times. *Ann. Probab.* **36** 160–197. MR2370602 https://doi.org/10.1214/009117907000000097

[5] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR1324786

[6] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. MR0898871 https://doi.org/10.1017/CBO9780511721434

[7] Blumenthal, R.M. and Getoor, R.K. (1961). Sample functions of stochastic processes with stationary independent increments. *J. Math. Mech.* **10** 493–516. MR0123362

[8] Böttcher, B., Schilling, R. and Wang, J. (2013). *Lévy Matters. III: Lévy-Type Processes: Construction, Approximation and Sample Path Properties*. *Lecture Notes in Math.* **2099**. Cham: Springer. MR3156646 https://doi.org/10.1007/978-3-319-02684-8

[9] Chen, L.H.Y. (1978). A short note on the conditional Borel-Cantelli lemma. *Ann. Probab.* **6** 699–700. MR0496420

[10] Cho, S., Kim, P. and Lee, J. General Law of iterated logarithm for Markov processes. Preprint, arXiv:2102.01917.

[11] 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

[12] Fristedt, B. (1974). Sample functions of stochastic processes with stationary, independent increments. In *Advances in Probability and Related Topics, Vol. 3* 241–396. MR0400406

[13] Gīhman, Ĭ.Ī. and Skorohod, A.V. (1974). *The Theory of Stochastic Processes. I. Die Grundlehren der Mathematischen Wissenschaften, Band 210*. New York: Springer. MR0346882

[14] Grzywny, T. (2014). On Harnack inequality and Hölder regularity for isotropic unimodal Lévy processes. *Potential Anal.* **41** 1–29. MR3225805 https://doi.org/10.1007/s11118-013-9360-y

[15] Gut, A. (2005). *Probability: A Graduate Course*. *Springer Texts in Statistics*. New York: Springer. MR2125120

[16] Hoh, W. (1998). Pseudo-Differential Operators Generating Markov Processes. Universität Bielefeld. Habilitationsschrift.

[17] Itô, K. (2004). *Stochastic Processes*. Berlin: Springer.

[18] Jacob, N. (2001–2005). *Pseudo Differential Operators and Markov Processes. Vols I–III*. London: Imperial College Press.

[19] Kallenberg, O. (1981). Splitting at backward times in regenerative sets. *Ann. Probab.* **9** 781–799. MR0628873

[20] Khintchine, A. (1939). Sur la croissance locale des processus stochastiques homogènes à accroissements indépendants. *Bull. Acad. Sci. URSS Sér. Math.* **1939** 487–508. MR0002054

[21] Khoshnevisan, D. and Schilling, R. (2016). *From Lévy-Type Processes to Parabolic SPDEs* (L. Quer-Sardanyons and F. Utzet, eds.). *Advanced Courses in Mathematics. CRM Barcelona*. Cham: Birkhäuser/Springer. MR3587832 https://doi.org/10.1007/978-3-319-34120-0

[22] Kim, P., Kumagai, T. and Wang, J. (2017). Laws of the iterated logarithm for symmetric jump processes. *Bernoulli* **23** 2330–2379. MR3648033 https://doi.org/10.3150/16-BEJ812

[23] Knopova, V. and Schilling, R.L. (2013). A note on the existence of transition probability densities of Lévy processes. *Forum Math.* **25** 125–149. MR3010850 https://doi.org/10.1515/form.2011.108

[24] Knopova, V. and Schilling, R.L. (2014). On the small-time behaviour of Lévy-type processes. *Stochastic Process. Appl.* **124** 2249–2265. MR3188355 https://doi.org/10.1016/j.spa.2014.02.008

[25] Kolokoltsov, V.N. (2011). *Markov Processes, Semigroups and Generators*. *De Gruyter Studies in Mathematics* **38**. Berlin: de Gruyter. MR2780345

[26] Kühn, F. (2017). *Lévy Matters. VI: Lévy-Type Processes: Moments, Construction and Heat Kernel Estimates*. *Lecture Notes in Math.* **2187**. Cham: Springer. MR3701414

[27] Kühn, F. (2018). Solutions of Lévy-driven SDEs with unbounded coefficients as Feller processes. *Proc. Amer. Math. Soc.* **146** 3591–3604. MR3803683 https://doi.org/10.1090/proc/14022

[28] Kühn, F. (2018). On martingale problems and Feller processes. *Electron. J. Probab.* **23** Paper No. 13, 18. MR3771750 https://doi.org/10.1214/18-EJP142

[29] Kühn, F. (2020). Existence of (Markovian) solutions to martingale problems associated with Lévy-type operators. *Electron. J. Probab.* **25** Paper No. 16, 26. MR4073677 https://doi.org/10.1214/20-ejp424

[30] Kühn, F. and Schilling, R.L. (2019). On the domain of fractional Laplacians and related generators of Feller processes. *J. Funct. Anal.* **276** 2397–2439. MR3926121 https://doi.org/10.1016/j.jfa.2018.12.011

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

[32] Negoro, A. (1994). Stable-like processes: Construction of the transition density and the behavior of sample paths near $t = 0$. *Osaka J. Math.* **31** 189–214. MR1262797

[33] Pruitt, W.E. (1981). The growth of random walks and Lévy processes. *Ann. Probab.* **9** 948–956. MR0632968

[34] Reker, J. Short-time behavior of solutions to Lévy-driven SDEs. Preprint, arXiv:2008.00526.

[35] Sato, K. (2013). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. MR3185174

[36] Savov, M. (2009). Small time two-sided LIL behavior for Lévy processes at zero. *Probab. Theory Related Fields* **144** 79–98. MR2480786 https://doi.org/10.1007/s00440-008-0142-1

[37] Schilling, R.L. (1998). Growth and Hölder conditions for the sample paths of Feller processes. *Probab. Theory Related Fields* **112** 565–611. MR1664705 https://doi.org/10.1007/s004400050201

[38] Schilling, R.L. and Schnurr, A. (2010). The symbol associated with the solution of a stochastic differential equation. *Electron. J. Probab.* **15** 1369–1393. MR2721050 https://doi.org/10.1214/EJP.v15-807

[39] Skorohod, A.V. (1991). *Random Processes with Independent Increments*. *Mathematics and Its Applications (Soviet Series)* **47**. Dordrecht: Kluwer Academic. MR1155400 https://doi.org/10.1007/978-94-011-3710-2

[40] Taylor, S.J. (1967). Sample path properties of a transient stable process. *J. Math. Mech.* **16** 1229–1246. MR0208684

[41] Wee, I.S. and Kim, Y.K. (1988). General laws of the iterated logarithm for Lévy processes. *J. Korean Statist. Soc.* **17** 30–45. MR0968135

Bernoulli **28**(4), 2022, 2909–2940
https://doi.org/10.3150/21-BEJ1442

Quantifying deviations from separability in space-time functional processes

HOLGER DETTE[a], GAUTHIER DIERICKX[b] and TIM KUTTA[c]

Ruhr-Universität Bochum, D-44780 Bochum. [a]*holger.dette@rub.de,* [b]*gauthier.dierickx@rub.de,* [c]*tim.kutta@rub.de*

The estimation of covariance operators of spatio-temporal data is in many applications only computationally feasible under simplifying assumptions, such as separability of the covariance into strictly temporal and spatial factors. Powerful tests for this assumption have been proposed in the literature. However, as real world systems, such as climate data are notoriously inseparable, validating this assumption by statistical tests, seems inherently questionable. In this paper we present an alternative approach: By virtue of separability measures, we quantify how strongly the data's covariance operator diverges from a separable approximation. Confidence intervals localize these measures with statistical guarantees. This method provides users with a flexible tool, to weigh the computational gains of a separable model against the associated increase in bias. As separable approximations we consider the established methods of partial traces and partial products, and develop weak convergence principles for the corresponding estimators. Moreover, we also prove such results for estimators of optimal, separable approximations, which are arguably of most interest in applications. In particular we present for the first time statistical inference for this object, which has been confined to estimation previously.

Besides confidence intervals, our results encompass tests for approximate separability. All methods proposed in this paper are free of nuisance parameters and do not require computationally expensive resampling procedures. A simulation study underlines the advantages of our approach and its applicability is demonstrated by the investigation of German annual temperature data.

Keywords: Space-time processes; approximate separability; partial traces; optimal approximation

References

Aitchison, J. (1964). Confidence-region tests. *J. Roy. Statist. Soc. Ser. B* **26** 462–476. MR0174119

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

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

Bel, L., Bar-Hen, A., Petit, R. and Cheddadi, R. (2011). Spatio-temporal functional regression on paleoecological data. *J. Appl. Stat.* **38** 695–704. MR2773575 https://doi.org/10.1080/02664760903563650

Berkes, I., Horváth, L. and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Process. Appl.* **123** 385–403. MR3003356 https://doi.org/10.1016/j.spa.2012.10.003

Bhatia, R. (2003). Partial traces and entropy inequalities. *Linear Algebra Appl.* **370** 125–132. MR1994322 https://doi.org/10.1016/S0024-3795(03)00386-0

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

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

Bücher, A. and Kojadinovic, I. (2016). A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli* **22** 927–968. MR3449804 https://doi.org/10.3150/14-BEJ682

Cardot, H. and Degras, D. (2018). Online principal component analysis in high dimension: Which algorithm to choose? *Int. Stat. Rev.* **86** 29–50. MR3796510 https://doi.org/10.1111/insr.12220

1350-7265

Chattopadhyay, A., Hassanzadeh, P. and Pasha, S. (2020). Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Sci. Rep.* **10** 1317.

Constantinou, P., Kokoszka, P. and Reimherr, M. (2017). Testing separability of space-time functional processes. *Biometrika* **104** 425–437. MR3698263 https://doi.org/10.1093/biomet/asx013

Constantinou, P., Kokoszka, P. and Reimherr, M. (2018). Testing separability of functional time series. *J. Time Series Anal.* **39** 731–747. MR3849524 https://doi.org/10.1111/jtsa.12302

Cowan, C., Sabharwal, J. and Wu, S. (2016). 09. Space–time codependence of retinal ganglion cells can be explained by novel and separable components of their receptive fields. *Physiol. Rep.* **4**.

Cressie, N.A.C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. New York: Wiley. MR3559472

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494 https://doi.org/10.2307/2669946

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

Crujeiras, R.M., Fernández-Casal, R. and González-Manteiga, W. (2010). Nonparametric test for separability of spatio-temporal processes. *Environmetrics* **21** 382–399. MR2842250 https://doi.org/10.1002/env.1006

Davies, E.B. (2007). *Linear Operators and Their Spectra*. *Cambridge Studies in Advanced Mathematics* **106**. Cambridge: Cambridge Univ. Press. MR2359869 https://doi.org/10.1017/CBO9780511618864

Dehling, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete* **63** 393–432. MR0705631 https://doi.org/10.1007/BF00542537

Dehling, H., Mikosch, T. and Sörensen, M. (2002). *Empirical Process Techniques for Dependent Data*. Basel: Birkhäuser.

Dehling, H. and Philipp, W. (1982). Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.* **10** 689–701. MR0659538

Depireux, D., Klein, D. and Shamma, S. (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol.* **85** 1220–34.

Dette, H., Dierickx, G. and Kutta, T. (2022). Supplement to "Quantifying deviations from separability in space-time functional processes." https://doi.org/10.3150/21-BEJ1442SUPP

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

Dette, H. and Kutta, T. (2021). Detecting structural breaks in eigensystems of functional time series. *Electron. J. Stat.* **15** 944–983. MR4255289 https://doi.org/10.1214/20-ejs1796

Dunford, N. and Schwartz, J.T. (1958). *Linear Operators. I. General Theory*. *Pure and Applied Mathematics, Vol. 7*. New York, Interscience Publishers. MR0117523

Fuentes, M. (2006). Testing for separability of spatial-temporal covariance functions. *J. Statist. Plann. Inference* **136** 447–466. MR2211349 https://doi.org/10.1016/j.jspi.2004.07.004

Garvey, C., Meng, C. and Nagy, J.G. (2018). Singular value decomposition approximation via Kronecker summations for imaging applications. *SIAM J. Matrix Anal. Appl.* **39** 1836–1857. MR3892426 https://doi.org/10.1137/18M1164147

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

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475 https://doi.org/10.1198/016214502760047113

Gneiting, T., Genton, M.G. and Guttorp, P. (2006). *Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry*, 1st ed. **107**. London: Chapman & Hall/CRC Press. Chapter 4, pp. 151–175.

Gohberg, I., Goldberg, S. and Kaashoek, M.A. (1993). *Classes of Linear Operators. Vol. II*. *Operator Theory: Advances and Applications* **63**. Basel: Birkhäuser. MR1246332 https://doi.org/10.1007/978-3-0348-8558-4_1

Gohberg, I., Goldberg, S. and Kaashoek, M.A. (2003). *Basic Classes of Linear Operators*. Basel: Birkhäuser. MR2015498 https://doi.org/10.1007/978-3-0348-7980-4

Golub, G.H. and Van Loan, C.F. (2013). *Matrix Computations*, 4th ed. *Johns Hopkins Studies in the Mathematical Sciences*. Baltimore, MD: Johns Hopkins Univ. Press. MR3024913

Gross, L. (1970). *Lectures in Modern Analysis and Applications II*, 1st ed. *Lecture Notes in Math.* **140**. Berlin Heidelberg: Springer.

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

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

Huang, H.-C. and Hsu, N.-J. (2004). Modeling transport effects on ground-level ozone using a non-stationary space-time model. *Environmetrics* **15** 251–268.

Jun, M. and Stein, M.L. (2004). Statistical comparison of observed and cmaq modeled daily sulfate levels. *Atmos. Environ.* **38** 4427–4436.

Kalogianni, K., De Munck, J., Nolte, G., Vardy, A., van der Helm, F. and Daffertshofer, A. (2018). 02. Spatial resolution for eeg source reconstruction-a simulation study on seps. *J. Neurosci. Methods* **301**.

Krall, J.R., Chang, H.H., Ebelt-Sarnat, S., Peng, R.D. and Waller, L. (2015). Current methods and challenges for epidemiological studies of the associations between chemical constituents of particulate matter and health. *Curr. Envir. Health Rep.* **2** 388–398.

Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Ann. Probab.* **8** 1003–1036. MR0602377

Kyriakidis, P.C. and Journel, A.G. (1999). Geostatistical space-time models: A review. *Math. Geol.* **31** 651–684. MR1694654 https://doi.org/10.1023/A:1007528426688

Laurini, M.P. (2019). A spatio-temporal approach to estimate patterns of climate change. *Environmetrics* **30** e2542. MR3908109 https://doi.org/10.1002/env.2542

Li, B., Genton, M.G. and Sherman, M. (2007). A nonparametric assessment of properties of space-time covariance functions. *J. Amer. Statist. Assoc.* **102** 736–744. MR2370863 https://doi.org/10.1198/016214507000000202

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

Lu, N. and Zimmerman, D.L. (2005). Testing for directional symmetry in spatial dependence using the periodogram. *J. Statist. Plann. Inference* **129** 369–385. MR2126855 https://doi.org/10.1016/j.jspi.2004.06.058

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

Masak, T., Sarkar, S. and Panaretos, V.M. (2020). Separable expansions for covariance estimation. https://arxiv.org/pdf/2007.12175.pdf.

Matsuda, Y. and Yajima, Y. (2004). On testing for separable correlations of multivariate time series. *J. Time Series Anal.* **25** 501–528. MR2073235 https://doi.org/10.1111/j.1467-9892.2004.01795.x

Mazer, J., Vinje, W., Mcdermott, J., Schiller, P. and Gallant, J. (2002). 03. Spatial frequency and orientation tuning dynamics in area v1. *Proc. Natl. Acad. Sci. USA* **99** 1645–50.

Merlevède, F. (2003). On the central limit theorem and its weak invariance principle for strongly mixing sequences with values in a Hilbert space via martingale approximation. *J. Theoret. Probab.* **16** 625–653. MR2009196 https://doi.org/10.1023/A:1025668415566

Mitchell, M.W., Genton, M.G. and Gumpertz, M.L. (2005). Testing for separability of space-time covariances. *Environmetrics* **16** 819–831. MR2216653 https://doi.org/10.1002/env.737

Mitchell, M.W., Genton, M.G. and Gumpertz, M.L. (2006). A likelihood ratio test for separability of covariances. *J. Multivariate Anal.* **97** 1025–1043. MR2276147 https://doi.org/10.1016/j.jmva.2005.07.005

Porcu, E., Alegria, A. and Furrer, R. (2018). Modelling temporally evolving and spatially globally dependent data. *Int. Stat. Rev.* **86** 344–377. MR3852415 https://doi.org/10.1111/insr.12266

Porcu, E., Furrer, R. and Nychka, D. (2021). 30 years of space-time covariance functions. *Wiley Interdiscip. Rev.: Comput. Stat.* **13** e1512. MR4218945 https://doi.org/10.1002/wics.1512

Rabiner, R.L. and Schafer, R.W. (1978). *Digital Processing of Speech Signals*. *Signal Processing Series* **100**. Prentice-Hall.

Rosenblatt, M. (1971). *Markov Processes. Structure and Asymptotic Behavior*. *Die Grundlehren der Mathematischen Wissenschaften, Band 184*. New York: Springer. MR0329037

Samur, J.D. (1984). Convergence of sums of mixing triangular arrays of random vectors with stationary rows. *Ann. Probab.* **12** 390–426. MR0735845

Samur, J.D. (1987). On the invariance principle for stationary ϕ-mixing triangular arrays with infinitely divisible limits. *Probab. Theory Related Fields* **75** 245–259. MR0885465 https://doi.org/10.1007/BF00354036

Scaccia, L. and Martin, R.J. (2005). Testing axial symmetry and separability of lattice processes. *J. Statist. Plann. Inference* **131** 19–39. MR2136004 https://doi.org/10.1016/j.jspi.2004.01.003

Sherman, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2815783

Skup, M. (2010). Longitudinal fMRI analysis: A review of methods. *Stat. Interface* **3** 235–252. MR2659514 https://doi.org/10.4310/SII.2010.v3.n2.a10

Stein, M.L. (2005). Space-time covariance functions. *J. Amer. Statist. Assoc.* **100** 310–321. MR2156840 https://doi.org/10.1198/016214504000000854

Tingley, M.P. and Huybers, P. (2013). Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. *Nature* **496** 201–205.

Tzeng, J. (2013). Split-and-combine singular value decomposition for large-scale matrix. *J. Appl. Math.* 683053. MR3033589 https://doi.org/10.1155/2013/683053

Van Loan, C.F. and Pitsianis, N. (1993). Approximation with Kronecker products. In *Linear Algebra for Large Scale and Real-Time Applications (Leuven, 1992)*. *NATO Adv. Sci. Inst. Ser. E: Appl. Sci.* **232** 293–314. Dordrecht: Kluwer Academic. MR1250183

Venezia, J., Martin, A.-G., Hickok, G. and Richards, V. (2019). 04. Identification of the spectrotemporal modulations that support speech intelligibility in hearing-impaired and normal-hearing listeners. *J. Speech Lang. Hear. Res.* **62** 1–17.

Weidmann, J. (1980). *Linear Operators in Hilbert Spaces*. *Graduate Texts in Mathematics* **68**. New York-Berlin: Springer. MR0566954

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J. and Evans, A.C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* **4** 58–73.

Zunino, A. and Mosegaard, K. (2018). 10. An efficient method to solve large linearizable inverse problems under Gaussian and separability assumptions. *Comput. Geosci.* **122** 77–86.

Bernoulli **28**(4), 2022, 2941–2967
https://doi.org/10.3150/21-BEJ1443

Tracy-Widom limit for the largest eigenvalue of high-dimensional covariance matrices in elliptical distributions

WEN JUN[a], XIE JIAHUI[b], YU LONG[c] and ZHOU WANG[d]

Department of Statistics and Data Science, National University of Singapore, Singapore. [a]*jun.wen@u.nus.edu,* [b]*jiahui.xie@u.nus.edu,* [c]*stayl@nus.edu.sg,* [d]*wangzhou@nus.edu.sg*

Let X be an $M \times N$ random matrix consisting of independent M-variate elliptically distributed column vectors $\mathbf{x}_1,\dots,\mathbf{x}_N$ with general population covariance matrix Σ. In the literature, the quantity XX^* is referred to as the sample covariance matrix after scaling, where X^* is the transpose of X. In this article, we prove that the limiting behavior of the scaled largest eigenvalue of XX^* is universal for a wide class of elliptical distributions, namely, the scaled largest eigenvalue converges weakly to the same limit regardless of the distributions that $\mathbf{x}_1,\dots,\mathbf{x}_N$ follow as $M,N \to \infty$ with $M/N \to \phi_0 > 0$ if the weak fourth moment of the radius of $\mathbf{x}_1$ exists. In particular, via comparing the Green function with that of the sample covariance matrix of multivariate normally distributed data, we conclude that the limiting distribution of the scaled largest eigenvalue is the celebrated Tracy-Widom law.

Keywords: Sample covariance matrices; elliptical distributions; edge universality; Tracy-Widom distribution; tail probability

References

[1] Anderson, T. and Fang, K. (1990). Statistical inference in elliptically contoured and related distributions Technical report, Standford Univ., Stanford, California. https://apps.dtic.mil/dtic/tr/fulltext/u2/a230672.pdf.

[2] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273

[3] 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

[4] Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

[5] Bai, Z.D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. MR1711663

[6] Bao, Z., Pan, G. and Zhou, W. (2013). Local density of the spectrum on the edge for sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864 https://doi.org/10.1214/14-AOS1281

[7] Bao, Z., Pan, G. and Zhou, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864 https://doi.org/10.1214/14-AOS1281

[8] Benaych-Georges, F. and Knowles, A. (2016). Lectures on the local semicircle law for Wigner matrices. ArXiv e-prints. Available at arXiv:1601.04055.

[9] Cai, T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281 https://doi.org/10.1007/s00440-014-0562-z

[10] Ding, X. (2017). Asymptotics of empirical eigen-structure for high dimensional sample covariance matrices of general form. ArXiv e-prints. Available at arXiv:1708.06296.

[11] Ding, X. and Yang, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.* **28** 1679–1738. MR3809475 https://doi.org/10.1214/17-AAP1341

1350-7265

[12] Dobriban, E. and Liu, S. (2018). A new theory for sketching in linear regression. ArXiv e-prints. Available at arXiv:1810.06089.
[13] Dobriban, E. and Sheng, Y. (2021). Distributed linear regression by averaging. *Ann. Statist.* **49** 918–943. MR4255113 https://doi.org/10.1214/20-aos1984
[14] El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592 https://doi.org/10.1214/009117906000000917
[15] El Karoui, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248 https://doi.org/10.1214/08-AAP548
[16] Erdësh, L. (2011). Universality of Wigner random matrices: A survey of recent results. *Russian Math. Surveys* **66** 507.
[17] Erdős, L., Knowles, A. and Yau, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922 https://doi.org/10.1007/s00023-013-0235-y
[18] Erdős, L., Yau, H.-T. and Yin, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. MR2981427 https://doi.org/10.1007/s00440-011-0390-3
[19] Erdős, L., Yau, H.-T. and Yin, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147 https://doi.org/10.1016/j.aim.2011.12.010
[20] 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
[21] Hu, J., Li, W. and Zhou, W. (2019). Central limit theorem for mutual information of large MIMO systems with elliptically correlated channels. *IEEE Trans. Inf. Theory* **65** 7168–7180. MR4030880 https://doi.org/10.1109/TIT.2019.2913760
[22] 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
[23] Johnstone, I.M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010 https://doi.org/10.1214/08-AOS605
[24] Johnstone, I.M. and Nadler, B. (2017). Roy's largest root test under rank-one alternatives. *Biometrika* **104** 181–193. MR3626473 https://doi.org/10.1093/biomet/asw060
[25] Jun, W. Jiahui, X. Long, Y. Wang, Z. (2022). Supplement to "Tracy-Widom limit for the largest eigenvalue of high-dimensional covariance matrices in elliptical distributions." https://doi.org/10.3150/21-BEJ1443SUPP
[26] Knowles, A. and Yin, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 https://doi.org/10.1007/s00440-016-0730-4
[27] Lee, J.O. and Yin, J. (2014). A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.* **163** 117–173. MR3161313 https://doi.org/10.1215/00127094-2414767
[28] Li, W. and Yao, J. (2018). On structure testing for component covariance matrices of a high dimensional mixture. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 293–318. MR3763693 https://doi.org/10.1111/rssb.12248
[29] Onatski, A. (2008). The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.* **18** 470–490. MR2398763 https://doi.org/10.1214/07-AAP454
[30] Onatski, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070 https://doi.org/10.3982/ECTA6964
[31] Pillai, N.S. and Yin, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978 https://doi.org/10.1214/13-AAP939
[32] Silverstein, J.W. (2009). The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices. In *Random Matrix Theory and Its Applications* (Z.D. Bai, Y. Chen and Y.-C. Yang, eds.). *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **18** 1–25. Hackensack, NJ: World Sci. Publ. MR2603192 https://doi.org/10.1142/9789814273121_0001
[33] Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices **108** 1033–1056. J. Stat. Phys., 5-6, Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. MR1933444 https://doi.org/10.1023/A:1019739414239
[34] Tracy, C.A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246
[35] Yang, X., Zheng, X. and Chen, J. (2021). Testing high-dimensional covariance matrices under the elliptical distribution and beyond. *J. Econometrics* **221** 409–423. MR4215033 https://doi.org/10.1016/j.jeconom.2020.05.017

Bernoulli **28**(4), 2022, 2968–2997
https://doi.org/10.3150/21-BEJ1444

Nonparametric inference for reversed mean models with panel count data

LI LIU[1,a], WEN SU[2,b], GUOSHENG YIN[2,c], XINGQIU ZHAO[3,d] and YING ZHANG[4,e]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, 430072, China.* [a]*lliu.math@whu.edu.cn*
[2]*Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong.*
[b]*jenna.wen.su@connect.hku.hk,* [c]*gyin@hku.hk*
[3]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong.*
[d]*xingqiu.zhao@polyu.edu.hk*
[4]*Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, USA.* [e]*ying.zhang@unmc.edu*

Panel count data typically refer to data arising from studies with recurrent events, in which subjects are observed only at discrete time points rather than under continuous observations. We investigate a general situation where a recurrent event process is eventually truncated by an informative terminal event and we are particularly interested in behaviors of the recurrent event process near the terminal event. We propose a reversed mean model for estimating the mean function of the recurrent event process. We develop a two-stage sieve likelihood-based method to estimate the mean function, which overcomes the computational difficulties arising from a nuisance functional parameter involved in the likelihood. The consistency and the convergence rate of the two-stage estimator are established. Allowing for the convergence rate slower than the standard rate, we develop the general weak convergence theory of M-estimators with a nuisance functional parameter, and then apply it to the proposed estimator for deriving the asymptotic normality. Furthermore, a class of two-sample tests is developed. The proposed methods are evaluated with extensive simulation studies and illustrated with panel count data from the Chinese Longitudinal Healthy Longevity Study.

Keywords: Nonparametric tests; recurrent events; reversed mean model; terminal event

References

Akritas, M.G. (2000). The central limit theorem under censoring. *Bernoulli* **6** 1109–1120. MR1809737 https://doi.org/10.2307/3318473

Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. *Springer Series in Statistics*. New York: Springer. MR1198884 https://doi.org/10.1007/978-1-4612-4348-9

Andrews, D.F. and Herzberg, A.M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer.

Balakrishnan, N. and Zhao, X. (2009). New multi-sample nonparametric tests for panel count data. *Ann. Statist.* **37** 1112–1149. MR2509069 https://doi.org/10.1214/08-AOS599

Breslow, N.E. (1984). Extra-Poisson variation in log-linear models. *Appl. Statist.* **33** 38–44.

Chan, K.C.G. and Wang, M.-C. (2010). Backward estimation of stochastic processes with failure events as time origins. *Ann. Appl. Stat.* **4** 1602–1620. MR2758343 https://doi.org/10.1214/09-AOAS319

Chan, K.C.G. and Wang, M.-C. (2017). Semiparametric modeling and estimation of the terminal behavior of recurrent marker processes before failure events. *J. Amer. Statist. Assoc.* **112** 351–362. MR3646577 https://doi.org/10.1080/01621459.2016.1140051

Cheng, S.C. and Wei, L.J. (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87** 89–97. MR1766830 https://doi.org/10.1093/biomet/87.1.89

Diao, G., Zeng, D., Hu, K. and Ibrahim, J.G. (2017). Modeling event count data in the presence of informative dropout with application to bleeding and transfusion events in myelodysplastic syndrome. *Stat. Med.* **36** 3475–3494. MR3696504 https://doi.org/10.1002/sim.7351

Hinde, R., ed. (1982). Compound Poisson regression models. In *GLIM 82: Proceedings of the International Conference in Generalized Linear Models*. (R. Gilchrist, ed.) 109–121. Berlin: Springer.

Hu, X.J., Lagakos, S.W. and Lockhart, R.A. (2009). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statist. Sinica* **19** 561–580. MR2514175

Hu, X.J., Sun, J. and Wei, L.-J. (2003). Regression parameter estimation from panel counts. *Scand. J. Stat.* **30** 25–43. MR1963891 https://doi.org/10.1111/1467-9469.00316

Huang, C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc.* **99** 1153–1165. MR2109503 https://doi.org/10.1198/016214504000001033

Huang, C.-Y., Wang, M.-C. and Zhang, Y. (2006). Analysing panel count data with informative observation times. *Biometrika* **93** 763–775. MR2285070 https://doi.org/10.1093/biomet/93.4.763

Kalbfleisch, J.D. and Lawless, J.F. (1985). The analysis of panel data under a Markov assumption. *J. Amer. Statist. Assoc.* **80** 863–871. MR0819585

Kong, S. and Nan, B. (2016). Semiparametric approach to regression with a covariate subject to a detection limit. *Biometrika* **103** 161–174. MR3465828 https://doi.org/10.1093/biomet/asv055

Kong, S., Nan, B., Kalbfleisch, J.D., Saran, R. and Hirth, R. (2018). Conditional modeling of longitudinal data with terminal event. *J. Amer. Statist. Assoc.* **113** 357–368. MR3803470 https://doi.org/10.1080/01621459.2016.1255637

Li, Z., Tosteson, T.D. and Bakitas, M.A. (2013). Joint modeling quality of life and survival using a terminal decline model in palliative care studies. *Stat. Med.* **32** 1394–1406. MR3045908 https://doi.org/10.1002/sim.5635

Liu, L., Su, W. Yin, G., Zhao, X. Zhang, Y. (2022). Supplement to "Nonparametric inference for reversed mean models with panel count data." https://doi.org/10.3150/21-BEJ1444SUPP

Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94** 705–718. MR2410018 https://doi.org/10.1093/biomet/asm057

Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count data using monotone *B*-splines. *J. Amer. Statist. Assoc.* **104** 1060–1070. MR2750237 https://doi.org/10.1198/jasa.2009.tm08086

Ma, L. and Sundaram, R. (2018). Analysis of gap times based on panel count data with informative observation times and unknown start time. *J. Amer. Statist. Assoc.* **113** 294–305. MR3803465 https://doi.org/10.1080/01621459.2016.1246369

Schumaker, L.L. (2007). *Spline Functions: Basic Theory*, 3rd ed. *Cambridge Mathematical Library*. Cambridge: Cambridge Univ. Press. MR2348176 https://doi.org/10.1017/CBO9780511618994

Sun, J. and Fang, H.-B. (2003). A nonparametric test for panel count data. *Biometrika* **90** 199–208. MR1966560 https://doi.org/10.1093/biomet/90.1.199

Sun, J. and Kalbfleisch, J.D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5** 279–289. MR1329298

Sun, J. and Wei, L.J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 293–302. MR1749540 https://doi.org/10.1111/1467-9868.00232

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *J. Amer. Statist. Assoc.* **107** 688–700. MR2980077 https://doi.org/10.1080/01621459.2012.682528

Thall, P.F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44** 197–209. MR0931635 https://doi.org/10.2307/2531907

Thall, P.F. and Lachin, J.M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *J. Amer. Statist. Assoc.* **83** 339–347.

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

Wellner, J.A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28** 779–814. MR1792787 https://doi.org/10.1214/aos/1015951998

Wellner, J.A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35** 2106–2142. MR2363965 https://doi.org/10.1214/009053607000000181

Zeng, D. and Cai, J. (2010). A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika* **97** 699–712. MR2672492 https://doi.org/10.1093/biomet/asq039

Zeng, Y., Vaupel, J., Xiao, Z., Liu, Y. and Zhang, C. (2017). Chinese Longitudinal Healthy Longevity Survey (CLHLS). In *Inter-University Consortium for Political and Social Research [Distributor]* 1998–2014. Ann Arbor, MI. https://doi.org/10.3886/ICPSR36692.v1

Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika* **93** 777–790. MR2285071 https://doi.org/10.1093/biomet/93.4.777

Zhang, Y. and Jamshidian, M. (2003). The gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics* **59** 1099–1106. MR2025135 https://doi.org/10.1111/j.0006-341X.2003.00126.x

Zhao, H., Li, Y. and Sun, J. (2013). Analyzing panel count data with a dependent observation process and a terminal event. *Canad. J. Statist.* **41** 174–191. MR3030791 https://doi.org/10.1002/cjs.11143

Zhao, X. and Zhang, Y. (2017). Asymptotic normality of nonparametric M-estimators with applications to hypothesis testing for panel count data. *Statist. Sinica* **27** 931–950. MR3675037

Zhao, X., Zhou, J. and Sun, L. (2011). Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. *Biometrics* **67** 404–414. MR2829009 https://doi.org/10.1111/j.1541-0420.2010.01458.x

Zhao, H., Zhang, Y., Zhao, X. and Yu, Z. (2019). A nonparametric regression model for panel count data analysis. *Statist. Sinica* **29** 809–826. MR3931389

Zhou, J., Zhang, H., Sun, L. and Sun, J. (2017). Joint analysis of panel count data with an informative observation process and a dependent terminal event. *Lifetime Data Anal.* **23** 560–584. MR3705825 https://doi.org/10.1007/s10985-016-9375-y

Bernoulli **28**(4), 2022, 2998–3022
https://doi.org/10.3150/21-BEJ1445

Coverage error optimal confidence intervals for local polynomial regression

SEBASTIAN CALONICO[1,a], MATIAS D. CATTANEO[2,b] and MAX H. FARRELL[3,c]

[1]*Department of Health Policy and Management, Columbia University, New York, New York, U.S.A.* [a]*sebastian.calonico@columbia.edu*
[2]*Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey, U.S.A.* [b]*cattaneo@princeton.edu*
[3]*Booth School of Business, University of Chicago, Chicago, Illinois, U.S.A.* [c]*max.farrell@chicagobooth.edu*

This paper studies higher-order inference properties of nonparametric local polynomial regression methods under random sampling. We prove Edgeworth expansions for t statistics and coverage error expansions for interval estimators that (i) hold uniformly in the data generating process, (ii) allow for the uniform kernel, and (iii) cover estimation of derivatives of the regression function. The terms of the higher-order expansions, and their associated rates as a function of the sample size and bandwidth sequence, depend on the smoothness of the population regression function, the smoothness exploited by the inference procedure, and on whether the evaluation point is in the interior or on the boundary of the support. We prove that robust bias corrected confidence intervals have the fastest coverage error decay rates in all cases, and we use our results to deliver novel, inference-optimal bandwidth selectors. The main methodological results are implemented in companion R and Stata software packages.

Keywords: Edgeworth expansion; Cramér condition; nonparametric regression; robust bias correction; bandwidth selection; optimal inference; minimax bound

References

[1] Bahadur, R.R. and Savage, L.J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Stat.* **27** 1115–1122. MR0084241 https://doi.org/10.1214/aoms/1177728077
[2] Beran, R. (1982). Estimated sampling distributions: The bootstrap and competitors. *Ann. Statist.* **10** 212–225. MR0642733
[3] Bhattacharya, R.N. and Ranga Rao, R. (1976). *Normal Approximation and Asymptotic Expansions*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0436272
[4] Calonico, S., Cattaneo, M.D. Farrell, M.H. (2022). Supplement to "Coverage error optimal confidence intervals for local polynomial regression." https://doi.org/10.3150/21-BEJ1445SUPP
[5] Calonico, S., Cattaneo, M.D. and Farrell, M.H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *J. Amer. Statist. Assoc.* **113** 767–779. MR3832225 https://doi.org/10.1080/01621459.2017.1285776
[6] Calonico, S., Cattaneo, M.D. and Farrell, M.H. (2019). `nprobust`: Nonparametric kernel-based estimation and robust bias-corrected inference. *J. Stat. Softw.* **91** 1–33.
[7] Calonico, S., Cattaneo, M.D. and Farrell, M.H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *Econom. J.* **23** 192–210. MR4108924 https://doi.org/10.1093/ectj/utz022
[8] Calonico, S., Cattaneo, M.D. and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* **82** 2295–2326. MR3301169 https://doi.org/10.3982/ECTA11757
[9] Cattaneo, M.D., Farrell, M.H. and Feng, Y. (2020). Large sample properties of partitioning-based series estimators. *Ann. Statist.* **48** 1718–1741. MR4124341 https://doi.org/10.1214/19-AOS1865

[10] Chen, S.X. and Qin, Y.S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* **87** 946–953. MR1813987 https://doi.org/10.1093/biomet/87.4.946
[11] Chen, S.X. and Qin, Y.S. (2002). Confidence intervals based on local linear smoother. *Scand. J. Stat.* **29** 89–99. MR1894383 https://doi.org/10.1111/1467-9469.00273
[12] Cheng, G. and Chen, Y.-C. (2019). Nonparametric inference via bootstrapping the debiased estimator. *Electron. J. Stat.* **13** 2194–2256. MR3980957 https://doi.org/10.1214/19-EJS1575
[13] Cheng, M.-Y., Fan, J. and Marron, J.S. (1997). On automatic boundary corrections. *Ann. Statist.* **25** 1691–1708. MR1463570 https://doi.org/10.1214/aos/1031594737
[14] Edgeworth, F.Y. (1883). The law of error. *The London, Edinburgh and Dublin Philosophical Magazine* **5** 300–309.
[15] Edgeworth, F.Y. (1906). The generalised law of error, or law of great numbers. *J. Roy. Stat. Soc.* **69** 497–539.
[16] Fan, J., Gasser, T., Gijbels, I., Brockmann, M. and Engel, J. (1997). Local polynomial regression: Optimal kernels and asymptotic minimax efficiency. *Ann. Inst. Statist. Math.* **49** 79–99. MR1450693 https://doi.org/10.1023/A:1003162622169
[17] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. London: CRC Press. MR1383587
[18] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. *Springer Series in Statistics*. New York: Springer. MR1964455 https://doi.org/10.1007/b97702
[19] Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics* **22** 215–232. MR1097375 https://doi.org/10.1080/02331889108802305
[20] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. New York: Springer. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7
[21] Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. MR1165587 https://doi.org/10.1214/aos/1176348651
[22] Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** 695–711. MR1165588 https://doi.org/10.1214/aos/1176348652
[23] Hall, P. and Jing, B.-Y. (1995). Uniform coverage bounds for confidence intervals and Berry-Esseen theorems for Edgeworth expansion. *Ann. Statist.* **23** 363–375. MR1332571 https://doi.org/10.1214/aos/1176324525
[24] Hall, P. and Kang, K.-H. (2001). Bootstrapping nonparametric density estimators with empirically chosen bandwidths. *Ann. Statist.* **29** 1443–1468. MR1873338 https://doi.org/10.1214/aos/1013203461
[25] Neumann, M.H. (1997). Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure. *Statistics* **29** 1–36. MR1438531 https://doi.org/10.1080/02331889708802572
[26] Tuvaandorj, P. (2020). Regression discontinuity designs, white noise models, and minimax. *J. Econometrics* **218** 587–608. MR4149240 https://doi.org/10.1016/j.jeconom.2020.04.030

Bernoulli **28**(4), 2022, 3023–3050
https://doi.org/10.3150/21-BEJ1446

Graphical modeling of stochastic processes driven by correlated noise

SØREN WENGEL MOGENSEN[1,2,3,a] and NIELS RICHARD HANSEN[1,b]

[1] *Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark.*
[a] *soren.wengel_mogensen@control.lth.se,* [b] *niels.r.hansen@math.ku.dk*
[2] *Section for Cognitive Systems, Technical University of Denmark, Kgs. Lyngby, Denmark*
[3] *Current affiliation: Department of Automatic Control, Lund University, Lund, Sweden*

We study a class of graphs that represent local independence structures in stochastic processes allowing for correlated noise processes. Several graphs may encode the same local independencies and we characterize such equivalence classes of graphs. In the worst case, the number of conditions in our characterizations grows superpolynomially as a function of the size of the node set in the graph. We show that deciding Markov equivalence of graphs from this class is coNP-complete which suggests that our characterizations cannot be improved upon substantially. We prove a global Markov property in the case of a multivariate Ornstein-Uhlenbeck process which is driven by correlated Brownian motions.

Keywords: Graphical models; stochastic processes; local independence; Markov equivalence; Ornstein–Uhlenbeck processes

References

[1] Aalen, O.O. (1987). Dynamic modelling and causality. *Scand. Actuar. J.* **3-4** 177–190. MR0943579 https://doi.org/10.1016/j.rser.2011.04.029

[2] Aalen, O.O. and Gjessing, H.K. (2004). Survival models based on the Ornstein–Uhlenbeck process. *Lifetime Data Anal.* **10** 407–423. MR2125423 https://doi.org/10.1007/s10985-004-4775-9

[3] Aalen, O.O., Røysland, K., Gran, J.M. and Ledergerber, B. (2012). Causality, mediation and time: A dynamic viewpoint. *J. Roy. Statist. Soc. Ser. A* **175** 831–861. MR2993496 https://doi.org/10.1111/j.1467-985X.2011.01030.x

[4] Ali, R.A., Richardson, T.S. and Spirtes, P. (2009). Markov equivalence for ancestral graphs. *Ann. Statist.* **37** 2808–2837. MR2541448 https://doi.org/10.1214/08-AOS626

[5] Bartoszek, K., Glémin, S., Kaj, I. and Lascoux, M. (2017). Using the Ornstein–Uhlenbeck process to model the evolution of interacting populations. *J. Theoret. Biol.* **429** 35–45. MR3672925 https://doi.org/10.1016/j.jtbi.2017.06.011

[6] Bormetti, G., Cazzola, V. and Delpini, D. (2010). Option pricing under Ornstein-Uhlenbeck stochastic volatility: A linear model. *Int. J. Theor. Appl. Finance* **13** 1047–1063. MR2738761 https://doi.org/10.1142/S0219024910006108

[7] Chandrasekaran, V., Srebro, N. and Harsha, P. (2008). Complexity of inference in graphical models. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI).*

[8] Chickering, D.M., Heckerman, D. and Meek, C. (2004). Large-sample learning of Bayesian networks is NP-hard. *J. Mach. Learn. Res.* **5** 1287–1330. MR2248018

[9] Commenges, D. and Gégout-Petit, A. (2009). A general dynamical statistical model with causal interpretation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 719–736. MR2749916 https://doi.org/10.1111/j.1467-9868.2009.00703.x

[10] Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2009). *Introduction to Algorithms*, 3rd ed. Cambridge, MA: MIT Press. MR2572804

[11] Cox, D.R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.* **8** 204–218. MR1243593

[12] Danks, D. and Plis, S. (2013). Learning causal structure from undersampled time series. In *JMLR: Workshop and Conference Proceedings* **10** 1–10.

[13] Didelez, V. (2000). Graphical Models for Event History Analysis based on Local Independence Ph.D. thesis, Universität Dortmund.

[14] Didelez, V. (2006). Graphical models for composable finite Markov processes. *Scand. J. Stat.* **34** 169–185. MR2325249 https://doi.org/10.1111/j.1467-9469.2006.00528.x

[15] Didelez, V. (2008). Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 245–264. MR2412641 https://doi.org/10.1111/j.1467-9868.2007.00634.x

[16] Ditlevsen, S. and Lansky, P. (2005). Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model. *Phys. Rev. E (3)* **71** 011907. MR2139313 https://doi.org/10.1103/PhysRevE.71.011907

[17] Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353. MR2354948 https://doi.org/10.1016/j.jeconom.2005.06.032

[18] Eichler, M. (2012). Graphical modelling of multivariate time series. *Probab. Theory Related Fields* **153** 233–268. MR2925574 https://doi.org/10.1007/s00440-011-0345-8

[19] Eichler, M. (2012). Causal inference in time series analysis. In *Causality: Statistical Perspectives and Applications* (C. Berzuini, P. Dawid and L. Bernardinelli, eds.) 327–354. New York: Wiley.

[20] Eichler, M. (2013). Causal inference with multiple time series: Principles and problems. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **371** 20110613. MR3081173 https://doi.org/10.1098/rsta.2011.0613

[21] Eichler, M. and Didelez, V. (2007). Causal reasoning in graphical time series models. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*.

[22] Eichler, M. and Didelez, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16** 3–32. MR2575937 https://doi.org/10.1007/s10985-009-9143-3

[23] Garey, M.R. and Johnson, D.S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. *A Series of Books in the Mathematical Sciences*. San Francisco, CA: W. H. Freeman and Co. MR0519066

[24] Granger, C.W.J. and Newbold, P. (1986). *Forecasting Economic Time Series*, 2nd ed. San Diego: Academic Press.

[25] Heath, R.A. (2000). The Ornstein-Uhlenbeck model for decision time in cognitive tasks: An example of control of nonlinear network dynamics. *Psychol. Res.* **63** 183–191.

[26] Hopcroft, J. and Tarjan, R. (1973). Algorithm 447: Efficient algorithms for graph manipulation. *Commun. ACM* **16** 372–378.

[27] Hyttinen, A., Plis, S., Järvisalo, M., Eberhardt, F. and Danks, D. (2016). Causal discovery from subsampled time series data by constraint optimization. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models (PGM)*.

[28] Jacobsen, M. (1993). A brief account of the theory of homogeneous Gaussian diffusions in finite dimensions. In *Frontiers in Pure and Applied Probability* (H. Niemi et al., eds.) **1** 86–94.

[29] Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR2778120

[30] Koster, J.T.A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Stat.* **26** 413–431. MR1712043 https://doi.org/10.1111/1467-9469.00157

[31] Lauritzen, S. (1996). *Graphical Models*. Oxford: Clarendon Press.

[32] Lee, M.-L.T. and Whitmore, G.A. (2006). Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. *Statist. Sci.* **21** 501–513. MR2380714 https://doi.org/10.1214/088342306000000330

[33] Liptser, R.S. and Shiryayev, A.N. (1977). *Statistics of Random Processes. I. General Theory*. *Applications of Mathematics, Vol. 5*. New York: Springer. MR0474486

[34] Maathuis, M., Drton, M., Lauritzen, S. and Wainwright, M., eds. (2019). *Handbook of Graphical Models*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. Boca Raton, FL: CRC Press. MR3889064

[35] Meek, C. (2001). Finding a path is harder than finding a tree. *J. Artificial Intelligence Res.* **15** 383–389. MR1884083 https://doi.org/10.1613/jair.914

[36] Mogensen, S.W. (2020). Graphical modeling in dynamical systems Ph.D. thesis, Univ. Copenhagen.
[37] Mogensen, S.W. and Hansen, N.R. (2020). Markov equivalence of marginalized local independence graphs. *Ann. Statist.* **48** 539–559. MR4065173 https://doi.org/10.1214/19-AOS1821
[38] Mogensen, S.W. and Hansen, N.R. Supplement to "Graphical modeling of stochastic processes driven by correlated noise." https://doi.org/10.3150/21-BEJ1446SUPP
[39] Mogensen, S.W., Malinsky, D. and Hansen, N.R. (2018). Causal learning for partially observed stochastic dynamical systems. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[40] Nodelman, U., Shelton, C.R. and Koller, D. (2002). Continuous time Bayesian networks. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[41] Nodelman, U., Shelton, C.R. and Koller, D. (2003). Learning continuous time Bayesian networks. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[42] Pavliotis, G.A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. *Texts in Applied Mathematics* **60**. New York: Springer. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7
[43] Ricciardi, L.M. and Sacerdote, L. (1979). The Ornstein-Uhlenbeck process as a model for neuronal activity. *Biol. Cybernet.* **35** 1–9.
[44] Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30** 145–157. MR1963898 https://doi.org/10.1111/1467-9469.00323
[45] Richardson, T. and Spirtes, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015
[46] Richardson, T.S. (1996). A discovery algorithm for directed cyclic graphs. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[47] Richardson, T.S. (1996). A polynomial-time algorithm for deciding Markov equivalence of directed cyclic graphical models. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[48] Richardson, T.S. (1997). A characterization of Markov equivalence for directed cyclic graphs. *Internat. J. Approx. Reason.* **17** 107–162.
[49] Rogers, L.C.G. and Williams, D. (2000). *Diffusions, Markov Processes, and Martingales. Vol. 2. Itô Calculus*. *Cambridge Mathematical Library*. Cambridge: Cambridge Univ. Press. MR1780932 https://doi.org/10.1017/CBO9781107590120
[50] Røysland, K. (2012). Counterfactual analyses with graphical models based on local independence. *Ann. Statist.* **40** 2162–2194. MR3059080 https://doi.org/10.1214/12-AOS1031
[51] Schöbel, R. and Zhu, J. (1999). Stochastic volatility with an Ornstein-Uhlenbeck process: An extension. *Eur. Finance Rev.* **3** 23–46.
[52] Schweder, T. (1970). Composable Markov processes. *J. Appl. Probab.* **7** 400–410. MR0264755 https://doi.org/10.2307/3211973
[53] Shimokawa, T., Pakdaman, K., Takahata, T., Tanabe, S. and Sato, S. (2000). A first-passage-time analysis of the periodically forced noisy leaky integrate-and-fire model. *Biol. Cybernet.* **83** 327–340.
[54] Sipser, M. (2013). *Introduction to the Theory of Computation*, 3rd ed. Boston: Thomson Course Technology.
[55] Sokol, A. and Hansen, N.R. (2014). Causal interpretation of stochastic differential equations. *Electron. J. Probab.* **19** 1–24. MR3275852 https://doi.org/10.1214/ejp.v19-2891
[56] Sonntag, D. and Peña, J.M. (2015). Chain graph interpretations and their relations revisited. *Internat. J. Approx. Reason.* **58** 39–56. MR3311407 https://doi.org/10.1016/j.ijar.2014.12.001
[57] Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI)*.
[58] Stein, E.M. and Stein, J.C. (1991). Stock price distributions with stochastic volatility: An analytic approach. *Rev. Financ. Stud.* **4** 727–752.
[59] Varando, G. and Hansen, N.R. (2020). Graphical continuous Lyapunov models. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[60] Verma, T. and Pearl, J. (1991). Equivalence and synthesis of causal models Technical Report R-150, Univ. California, Los Angeles.
[61] Wright, S. (1921). Correlation and causation. *J. Agric. Res.* **20** 557–585.
[62] Wright, S. (1934). The method of path coefficients. *Ann. Math. Stat.* **5** 161–215.

Bernoulli **28**(4), 2022, 3051–3078
https://doi.org/10.3150/22-BEJ1482

Conditional regression for single-index models

ALESSANDRO LANTERI[1,a], MAURO MAGGIONI[2,b] and STEFANO VIGOGNA[3,c]

[1]*DEMM, Università degli Studi di Milano, Milano, Italy.* [a]*alessandro.lanteri@unimi.it*
[2]*Department of Mathematics and Department of Applied Mathematics & Statistics, Johns Hopkins University, Baltimore, USA.* [b]*mauromaggionijhu@icloud.com*
[3]*MaLGa, DIBRIS, Università degli Studi di Genova, Genova, Italy.* [c]*vigogna@dibris.unige.it*

The single-index model is a statistical model for intrinsic regression where responses are assumed to depend on a single yet unknown linear combination of the predictors, allowing to express the regression function as $\mathbb{E}[Y|X] = f(\langle v, X\rangle)$ for some unknown *index* vector v and *link* function f. Conditional methods provide a simple and effective approach to estimate v by averaging moments of X conditioned on Y, but depend on parameters whose optimal choice is unknown and do not provide generalization bounds on f. In this paper we propose a new conditional method converging at $\sqrt{n}$ rate under an explicit parameter characterization. Moreover, we prove that polynomial partitioning estimates achieve the 1-dimensional min-max rate for regression of Hölder functions when combined to any $\sqrt{n}$-convergent index estimator. Overall this yields an estimator for dimension reduction and regression of single-index models that attains statistical optimality in quasilinear time.

Keywords: Single-index model; dimension reduction; nonparametric regression; finite-sample bounds

References

[1] Bach, F. (2017). Breaking the curse of dimensionality with convex neutral networks. *J. Mach. Learn. Res.* **18** Paper No. 19, 53. MR3634886
[2] Bhatia, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. New York: Springer. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8
[3] Bickel, P.J. and Li, B. (2007). Local polynomial regression on unknown manifolds. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 177–186. Beachwood, OH: IMS. MR2459188 https://doi.org/10.1214/074921707000000148
[4] Binev, P., Cohen, A., Dahmen, W. and DeVore, R. (2007). Universal algorithms for learning theory. II. Piecewise polynomial functions. *Constr. Approx.* **26** 127–152. MR2327596 https://doi.org/10.1007/s00365-006-0658-z
[5] Binev, P., Cohen, A., Dahmen, W., DeVore, R. and Temlyakov, V. (2005). Universal algorithms for learning theory. I. Piecewise constant functions. *J. Mach. Learn. Res.* **6** 1297–1321. MR2249856
[6] Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11** 368–385. MR0629795 https://doi.org/10.1016/0047-259X(81)90082-8
[7] Carroll, R.J., Fan, J., Gijbels, I. and Wand, M.P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842 https://doi.org/10.2307/2965697
[8] Carroll, R.J., Ruppert, D. and Welsh, A.H. (1998). Local estimating equations. *J. Amer. Statist. Assoc.* **93** 214–227. MR1614624 https://doi.org/10.2307/2669618
[9] Cook, R.D. (1998). Save: a method for dimension reduction and graphics in regression. *Communications in Statistics - Theory and Methods* **29** 2109–2121.
[10] Cook, R.D. and Lee, H. (1999). Dimension reduction in binary response regression. *J. Amer. Statist. Assoc.* **94** 1187–1200. MR1731482 https://doi.org/10.2307/2669934
[11] Coudret, R., Liquet, B. and Saracco, J. (2014). Comparison of sliced inverse regression approaches for underdetermined cases. *J. SFdS* **155** 72–96. MR3211755
[12] Cui, X., Härdle, W.K. and Zhu, L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39** 1658–1688. MR2850216 https://doi.org/10.1214/10-AOS871

1350-7265

[13] Dalalyan, A.S., Juditsky, A. and Spokoiny, V. (2008). A new algorithm for estimating the effective dimension-reduction subspace. *J. Mach. Learn. Res.* **9** 1648–1678. MR2438819
[14] Delecroix, M., Härdle, W. and Hristache, M. (1997). Efficient estimation in single-index regression. Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes. Humboldt University of Berlin. SFB 373 Discussion Paper 37.
[15] Delecroix, M. and Hristache, M. (1999). M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique. *Bull. Belg. Math. Soc. Simon Stevin* **6** 161–185. MR1705152
[16] Delecroix, M., Hristache, M. and Patilea, V. (2006). On semiparametric M-estimation in single-index regression. *J. Statist. Plann. Inference* **136** 730–769. MR2181975 https://doi.org/10.1016/j.jspi.2004.09.006
[17] Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12** 793–815. MR0751274 https://doi.org/10.1214/aos/1176346703
[18] Duan, N. and Li, K.-C. (1991). Slicing regression: A link-free regression method. *Ann. Statist.* **19** 505–530. MR1105834 https://doi.org/10.1214/aos/1176348109
[19] Eaton, M.L. (1986). A characterization of spherical distributions. *J. Multivariate Anal.* **20** 272–276. MR0866075 https://doi.org/10.1016/0047-259X(86)90083-7
[20] Gaïffas, S. and Lecué, G. (2007). Optimal rates and adaptation in the single-index model using aggregation. *Electron. J. Stat.* **1** 538–573. MR2369025 https://doi.org/10.1214/07-EJS077
[21] Ganti, R., Rao, N., Willett, R.M. and Nowak, R. (2015). Learning single index models in high dimensions. arXiv:1506.08910.
[22] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. *Springer Series in Statistics*. New York: Springer. MR1920390 https://doi.org/10.1007/b97848
[23] Hall, P. and Li, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.* **21** 867–889. MR1232523 https://doi.org/10.1214/aos/1176349155
[24] Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171 https://doi.org/10.1214/aos/1176349020
[25] Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995. MR1134488
[26] Horowitz, J.L. (1998). *Semiparametric Methods in Econometrics*. *Lecture Notes in Statistics* **131**. New York: Springer. MR1624936 https://doi.org/10.1007/978-1-4612-0621-7
[27] Hristache, M., Juditsky, A., Polzehl, J. and Spokoiny, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.* **29** 1537–1566. MR1891738 https://doi.org/10.1214/aos/1015345954
[28] Hristache, M., Juditsky, A. and Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29** 595–623. MR1865333 https://doi.org/10.1214/aos/1009210681
[29] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. MR1230981 https://doi.org/10.1016/0304-4076(93)90114-K
[30] Kakade, S.M., Kanade, V., Shamir, O. and Kalai, A.T. (2011). Efficient learning of generalized linear and single index models with isotonic regression. *Adv. Neural Inf. Process. Syst.* **24** 927–935.
[31] Kalai, A.T. and Sastry, R. (2009). The isotron algorithm: High-dimensional isotonic regression. In *Proceedings of the 22nd Annual Conference on Learning Theory (COLT)*.
[32] Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–438. MR0287628
[33] Klock, T., Lanteri, A. and Vigogna, S. (2021). Estimating multi-index models with response-conditional least squares. *Electronic Journal of Statistics* **15** Paper No. 1, 589–629.
[34] Kpotufe, S. (2011). k-nn regression adapts to local intrinsic dimension. *Adv. Neural Inf. Process. Syst.* **24** 729–737.
[35] Kpotufe, S. and Garg, V. (2013). Adaptivity to local smoothness and dimension in kernel regression. *Adv. Neural Inf. Process. Syst.* **26** 3075–3083.
[36] Lanteri, A., Maggioni, M. Vigogna, S. (2022). Supplement to "Conditional regression for single-index models." https://doi.org/10.3150/22-BEJ1482SUPP
[37] Li, B. (2018). *Sufficient Dimension Reduction: Methods and Applications with R. Chapman & Hall/CRC Monographs on Statistics and Applied Probability.* Boca Raton: CRC Press.
[38] Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102** 997–1008. MR2354409 https://doi.org/10.1198/016214507000000536

[39] Li, B., Zha, H. and Chiaromonte, F. (2005). Contour regression: A general approach to dimension reduction. *Ann. Statist.* **33** 1580–1616. MR2166556 https://doi.org/10.1214/009053605000000192

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

[41] Liao, W., Maggioni, M. and Vigogna, S. (2016). Learning adaptive multiscale approximations to data and functions near low-dimensional sets. In *2016 IEEE Information Theory Workshop (ITW)* 226–230. IEEE.

[42] Liao, W., Maggioni, M. and Vigogna, S. (2022). Multiscale regression on unknown manifolds. *Math. Eng.* **4** Paper No. 4, 1–25. MR4311326 https://doi.org/10.3934/mine.2022028

[43] Stewart, G.W. and Sun, J.G. (1990). *Matrix Perturbation Theory*. *Computer Science and Scientific Computing*. Boston, MA: Academic Press. MR1061154

[44] Stoker, T.M. (1986). Consistent estimation of scaled coefficients. *Econometrica* **54** 1461–1481. MR0868152 https://doi.org/10.2307/1914309

[45] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642

[46] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With applications to statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

[47] 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

[48] Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory* **22** 1112–1137. MR2328530 https://doi.org/10.1017/S0266466606060531

[49] 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

[50] Yu, Y., Wang, T. and Samworth, R.J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika* **102** 315–323. MR3371006 https://doi.org/10.1093/biomet/asv008