BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Thirty One Number Three August 2025 ISSN: 1350-7265

CONTENTS

(continued)

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

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Thirty One Number Three August 2025 ISSN: 1350-7265

CONTENTS

(continued)

BERNOULLI Volume 31 Number 3 August 2025 Pages 1693–2546 ISI/BS

Volume 31 Number 3   August 2025   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: https://www.bernoullisociety.org/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 https://www.bernoullisociety.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 2025 Bernoulli consists of 4 issues published in February, May, August and November.

Bernoulli Society
for Mathematical Statistics
and Probability

Bernoulli **31**(3), 2025, 1693–1708
https://doi.org/10.3150/22-BEJ1473

On the robustness of the minimim ℓ_2 interpolator

GEOFFREY CHINOT[1,a] and MATTHIEU LERASLE[2,b]

[1]*ETHZ, Rämistrasse 101, 8092 Zürich, Switzerland,* [a]*geoffrey.chinot@stat.math.ethz.ch*
[2]*CNRS, ENSAE, CREST, 5 avenue Henri Chatelier 91120 Palaiseau, France,* [b]*matthieu.lerasle@ensae.fr*

We analyse the interpolator with minimal ℓ_2-norm $\hat{\beta}$ in a general high dimensional linear regression framework where $\mathbb{Y} = \mathbb{X}\beta^* + \xi$ with $\mathbb{X}$ a random $n \times p$ matrix with independent $\mathcal{N}(0, \Sigma)$ rows. We prove that, with high probability, without assumption on the noise vector $\xi \in \mathbb{R}^n$, the ellipsoid risk $\|\hat{\beta} - \beta^*\|_\Sigma^2 = (\hat{\beta} - \beta^*)^T \Sigma(\hat{\beta} - \beta^*)$ is bounded from above by $(\|\beta^*\|_2^2 r_{cn}(\Sigma) \vee \|\xi\|^2)/n$, where c is an absolute constant and, for any $k \geq 1$, $r_k(\Sigma) = \sum_{i \geq k} \lambda_i(\Sigma)$ is the tail sum of the eigenvalues of Σ. These bounds show a transition in the rates. For high signal to noise ratios, the rates $\|\beta^*\|_2^2 r_{cn}(\Sigma)/n$ broadly improve the existing ones. For low signal to noise ratio, we also provide lower bound holding with large probability. General lower bounds are proved under minor restrictions on the noise ξ (see Theorem 1). Under assumptions on the sprectrum of Σ, this lower bound is of order $\|\xi\|_2^2/n$, matching the upper bound. Consequently, in the large noise regime, we are able to precisely track the ellipsoid risk with large probability. These results give new insight when the interpolation can be harmless in high dimensions.

Keywords: Interpolation problems; statistical learning; robustness

References

[1] Advani, M.S. and Saxe, A.M. (2017). High-dimensional dynamics of generalization error in neural networks. arXiv preprint. Available at arXiv:1710.03667.

[2] Amelunxen, D., Lotz, M., McCoy, M.B. and Tropp, J.A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453 https://doi.org/10.1093/imaiai/iau005

[3] Arora, S., Du, S.S., Hu, W., Li, Z., Salakhutdinov, R.R. and Wang, R. (2019). On exact computation with an infinitely wide neural net. *Adv. Neural Inf. Process. Syst.* 8139–8148.

[4] Bartlett, P.L., Long, P.M., Lugosi, G. and Tsigler, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 https://doi.org/10.1073/pnas.1907378117

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

[6] Belkin, M., Hsu, D. and Xu, J. (2020). Two models of double descent for weak features. *SIAM J. Math. Data Sci.* **2** 1167–1180. MR4186534 https://doi.org/10.1137/20M1336072

[7] Belkin, M., Hsu, D.J. and Mitra, P. (2018). Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate. *Adv. Neural Inf. Process. Syst.* 2300–2311.

[8] Belkin, M., Rakhlin, A. and Tsybakov, A.B. (2018). Does data interpolation contradict statistical optimality? arXiv preprint. Available at arXiv:1806.09471.

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

[10] Bietti, A. and Mairal, J. (2019). On the inductive bias of neural tangent kernels. *Adv. Neural Inf. Process. Syst.* 12873–12884.

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

[12] Bunea, F., Strimas-Mackey, S. and Wegkamp, M. (2020). Interpolation under latent factor regression models. arXiv preprint. Available at arXiv:2002.02525.

[13] Candès, E.J., Romberg, J.K. and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. MR2230846 https://doi.org/10.1002/cpa.20124

[14] Casella, G. (1980). Minimax ridge regression estimation. *Ann. Statist.* **8** 1036–1056. MR0585702

[15] Chatterji, N.S., Long, P.M. and Bartlett, P.L. (2021). When does gradient descent with logistic loss find interpolating two-layer networks? *J. Mach. Learn. Res.* **22** Paper No. 159, 48. MR4318515

[16] Chen, S.S., Donoho, D.L. and Saunders, M.A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094 https://doi.org/10.1137/S1064827596304010

[17] De Mol, C., De Vito, E. and Rosasco, L. (2009). Elastic-net regularization in learning theory. *J. Complexity* **25** 201–230. MR2513616 https://doi.org/10.1016/j.jco.2009.01.002

[18] Dirksen, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.* **20** no. 53, 29. MR3354613 https://doi.org/10.1214/EJP.v20-3760

[19] Fan, J., Li, R., Zhang, C.-H. and Zou, H. (2020). *Statistical Foundations of Data Science*. CRC press.

[20] Feldman, V. (2020). Does learning require memorization? A short tale about a long tail. In *STOC '20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* 954–959. New York: ACM. MR4141813 https://doi.org/10.1145/3357713.3384290

[21] Giraud, C. (2015). *Introduction to High-Dimensional Statistics*. *Monographs on Statistics and Applied Probability* **139**. Boca Raton, FL: CRC Press. MR3307991

[22] Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

[23] Jacot, A., Gabriel, F. and Hongler, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. *Adv. Neural Inf. Process. Syst.* 8571–8580.

[24] Koltchinskii, V. and Mendelson, S. (2015). Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN* 23 12991–13008. MR3431642 https://doi.org/10.1093/imrn/rnv096

[25] Lecué, G. and Mendelson, S. (2013). Learning subgaussian classes: Upper and minimax bounds. arXiv preprint. Available at arXiv:1305.4825.

[26] Lee, J., Xiao, L., Schoenholz, S.S., Bahri, Y., Novak, R., Sohl-Dickstein, J. and Pennington, J. (2020). Wide neural networks of any depth evolve as linear models under gradient descent. *J. Stat. Mech. Theory Exp.* 12 124002, 15. MR4241355 https://doi.org/10.1088/1742-5468/abc62b

[27] Liang, T. and Rakhlin, A. (2020). Just interpolate: Kernel "ridgeless" regression can generalize. *Ann. Statist.* **48** 1329–1347. MR4124325 https://doi.org/10.1214/19-AOS1849

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

[29] Mendelson, S. (2014). Learning without concentration. In *Conference on Learning Theory*. 25–39.

[30] Mendelson, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680. MR3565471 https://doi.org/10.1016/j.spa.2016.04.019

[31] Mendelson, S. (2017). On multiplier processes under weak moment assumptions. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **2169** 301–318. Cham: Springer. MR3645129

[32] Negrea, J., Dziugaite, G.K. and Roy, D. (2020). In defense of uniform convergence: Generalization via derandomization with an application to interpolating predictors. In *International Conference on Machine Learning*. 7263–7272. PMLR.

[33] Talagrand, M. (2014). *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **60**. Heidelberg: Springer. MR3184689 https://doi.org/10.1007/978-3-642-54075-2

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

[35] van de Geer, S.A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809 https://doi.org/10.1214/009053607000000929

[36] Vershynin, R. (2015). Estimation in high dimensions: A geometric perspective. In *Sampling Theory, a Renaissance*. *Appl. Numer. Harmon. Anal.* 3–66. Cham: Birkhäuser/Springer. MR3467418

[37] 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. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596
[38] Wainwright, M.J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge: Cambridge Univ. Press. MR3967104 https://doi.org/10.1017/9781108627771
[39] Wojtaszczyk, P. (2010). Stability and instance optimality for Gaussian measurements in compressed sensing. *Found. Comput. Math.* **10** 1–13. MR2591836 https://doi.org/10.1007/s10208-009-9046-4
[40] Yang, Z., Bai, Y. and Mei, S. (2021). Exact gap between generalization error and uniform convergence in random feature models. arXiv preprint. Available at arXiv:2103.04554.
[41] Zhang, C., Bengio, S., Hardt, M., Recht, B. and Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. arXiv preprint. Available at arXiv:1611.03530.
[42] Zhou, L., Sutherland, D.J. and Srebro, N. (2020). On uniform convergence and low-norm interpolation learning. arXiv preprint. Available at arXiv:2006.05942.
[43] 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 **31**(3), 2025, 1709–1734
https://doi.org/10.3150/24-BEJ1745

Limit theorems for SDEs with irregular drifts

JIANHAI BAO[a] and JIAQING HAO[b]

Center for Applied Mathematics, Tianjin University, 300072 Tianjin, P.R. China, [a]*jianhaibao@tju.edu.cn,* [b]*hjq_0227@tju.edu.cn*

In this paper, concerning SDEs with Hölder continuous drifts, which are merely dissipative at infinity, and SDEs with piecewise continuous drifts, we investigate the strong law of large numbers and the central limit theorem for underlying additive functionals and reveal the corresponding rates of convergence. To establish the limit theorems under consideration, the exponentially contractive property of solution processes under the (quasi-)Wasserstein distance plays an indispensable role. In order to achieve such contractive property, which is new and interesting in its own right for SDEs with Hölder continuous drifts or piecewise continuous drifts, the reflection coupling method is employed and meanwhile a sophisticated test function is built.

Keywords: Central limit theorem; dissipativity at infinity; Hölder continuous drift; piecewise continuous drift; strong law of large numbers

References

[1] Bao, J., Wang, F.-Y. and Yuan, C. (2015). Hypercontractivity for functional stochastic differential equations. *Stochastic Process. Appl.* **125** 3636–3656. MR3357623 https://doi.org/10.1016/j.spa.2015.04.001

[2] Bao, J., Wang, F.-Y. and Yuan, C. (2020). Limit theorems for additive functionals of path-dependent SDEs. *Discrete Contin. Dyn. Syst.* **40** 5173–5188. MR4128304 https://doi.org/10.3934/dcds.2020224

[3] Bao, J., Wang, F.-Y. and Yuan, C. (2020). Ergodicity for neutral type SDEs with infinite length of memory. *Math. Nachr.* **293** 1675–1690. MR4148827 https://doi.org/10.1002/mana.201800539

[4] Chen, C., Dang, T., Hong, J. and Zhou, T. (2023). CLT for approximating ergodic limit of SPDEs via a full discretization. *Stochastic Process. Appl.* **157** 1–41. MR4521762 https://doi.org/10.1016/j.spa.2022.11.015

[5] Hairer, M., Mattingly, J.C. and Scheutzow, M. (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149** 223–259. MR2773030 https://doi.org/10.1007/s00440-009-0250-6

[6] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. *Probability and Mathematical Statistics*. New York-London: Academic Press [Harcourt Brace Jovanovich, Publishers]. MR0624435

[7] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Berlin: Springer. MR0959133 https://doi.org/10.1007/978-3-662-02514-7

[8] Jin, D. (2025). Central limit theorem for temporal average of backward Euler-Maruyama method. *J. Comput. Math.* **43** 588–614. MR4829500 https://doi.org/10.4208/jcm.2311-m2023-0147

[9] Komorowski, T., Landim, C. and Olla, S. (2012). *Fluctuations in Markov Processes: Time Symmetry and Martingale Approximation*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **345**. Heidelberg: Springer. MR2952852 https://doi.org/10.1007/978-3-642-29880-6

[10] Komorowski, T. and Walczuk, A. (2012). Central limit theorem for Markov processes with spectral gap in the Wasserstein metric. *Stochastic Process. Appl.* **122** 2155–2184. MR2921976 https://doi.org/10.1016/j.spa.2012.03.006

[11] Kulik, A. (2018). *Ergodic Behavior of Markov Processes with Applications to Limit Theorems*. *De Gruyter Studies in Mathematics* **67**. Berlin: de Gruyter. MR3791835

[12] Leobacher, G. and Szölgyenyi, M. (2017). A strong order 1/2 method for multidimensional SDEs with discontinuous drift. *Ann. Appl. Probab.* **27** 2383–2418. MR3693529 https://doi.org/10.1214/16-AAP1262

1350-7265

[13] Lu, J., Tan, Y. and Xu, L. (2022). Central limit theorem and self-normalized Cramér-type moderate deviation for Euler-Maruyama scheme. *Bernoulli* **28** 937–964. MR4388925 https://doi.org/10.3150/21-bej1372

[14] Majka, M.B., Mijatović, A. and Szpruch, Ł. (2020). Nonasymptotic bounds for sampling algorithms without log-concavity. *Ann. Appl. Probab.* **30** 1534–1581. MR4132634 https://doi.org/10.1214/19-AAP1535

[15] Müller-Gronbach, T. and Yaroslavtseva, L. (2022). A strong order 3/4 method for SDEs with discontinuous drift coefficient. *IMA J. Numer. Anal.* **42** 229–259. MR4367654 https://doi.org/10.1093/imanum/draa078

[16] Müller-Gronbach, T., Sabanis, S. and Yaroslavtseva, L. (2024). Existence, uniqueness and approximation of solutions of SDEs with superlinear coefficients in the presence of discontinuities of the drift coefficient. *Commun. Math. Sci.* **23** 389–429. MR4846712 https://doi.org/10.4310/cms.241217213438

[17] Pagès, G. and Panloup, F. (2012). Ergodic approximation of the distribution of a stationary diffusion: Rate of convergence. *Ann. Appl. Probab.* **22** 1059–1100. MR2977986 https://doi.org/10.1214/11-AAP779

[18] Priola, E. and Wang, F.-Y. (2006). Gradient estimates for diffusion semigroups with singular coefficients. *J. Funct. Anal.* **236** 244–264. MR2227134 https://doi.org/10.1016/j.jfa.2005.12.010

[19] Shirikyan, A. (2006). Law of large numbers and central limit theorem for randomly forced PDE's. *Probab. Theory Related Fields* **134** 215–247. MR2222383 https://doi.org/10.1007/s00440-005-0427-6

[20] Wang, F.-Y. (2016). Gradient estimates and applications for SDEs in Hilbert space with multiplicative noise and Dini continuous drift. *J. Differ. Equ.* **260** 2792–2829. MR3427683 https://doi.org/10.1016/j.jde.2015.10.020

[21] Wang, F.-Y. (2023). Exponential ergodicity for non-dissipative McKean-Vlasov SDEs. *Bernoulli* **29** 1035–1062. MR4550214 https://doi.org/10.3150/22-bej1489

[22] Wang, Y., Wu, F. and Zhu, C. (2022). Limit theorems for additive functionals of stochastic functional differential equations with infinite delay. *J. Differ. Equ.* **308** 421–454. MR4340783 https://doi.org/10.1016/j.jde.2021.11.011

[23] Xia, P., Xie, L., Zhang, X. and Zhao, G. (2020). $L^q(L^p)$-theory of stochastic differential equations. *Stochastic Process. Appl.* **130** 5188–5211. MR4108486 https://doi.org/10.1016/j.spa.2020.03.004

Bernoulli **31**(3), 2025, 1735–1758
https://doi.org/10.3150/24-BEJ1757

Random fields on Hilbert spaces with their equivalent Gaussian measures

VINÍCIUS FERREIRA[1,2,a], EMILIO PORCU[1,3,b] and JORGE ZUBELLI[1,2,3,c]

[1]*Department of Mathematics, Khalifa University, Abu Dhabi, UAE,* [a]*vini.ssf@impa.br,* [b]*emilio.porcu@ku.ac.ae,* [c]*zubelli@gmail.com*
[2]*IMPA, Rio de Janeiro, Brazil*
[3]*ADIA Lab, Abu Dhabi, the United Arab Emirates*

We consider the infinite dimensional case for Gaussian fields having a compact set of $\mathbb{R}^d$ as their index set. For Hilbert space valued Gaussian fields, we prove conditions for equivalence of their Gaussian measures, in the sense that the Gaussian measures induced by two given fields are equivalent on their paths. Such conditions are proved to depend on the difference between the correlation operators that characterize the two Gaussian measures. Further, we provide the analogue of the celebrated Bochner's characterisation for correlation operators that represents the infinite dimensional extension of matrix-valued kernels.

Keywords: Functional random field; Gaussian measure; Hilbert-Schmidt operator

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. (2023). Stein's method meets computational statistics: A review of some recent developments. *Statist. Sci.* **38** 120–139. MR4534646 https://doi.org/10.1214/22-sts863

[2] Arafat, A., Porcu, E., Bevilacqua, M. and Mateu, J. (2018). Equivalence and orthogonality of Gaussian measures on spheres. *J. Multivariate Anal.* **167** 306–318. MR3830648 https://doi.org/10.1016/j.jmva.2018.05.005

[3] Bachoc, F., Porcu, E., Bevilacqua, M., Furrer, R. and Faouzi, T. (2022). Asymptotically equivalent prediction in multivariate geostatistics. *Bernoulli* **28** 2518–2545. MR4474552 https://doi.org/10.3150/21-bej1427

[4] Barp, A., Oates, C.J., Porcu, E. and Girolami, M. (2022). A Riemann-Stein kernel method. *Bernoulli* **28** 2181–2208. MR4474540 https://doi.org/10.3150/21-bej1415

[5] 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

[6] Bochner, S. (1933). Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. *Math. Ann.* **108** 378–410. MR1512856 https://doi.org/10.1007/BF01452844

[7] Bochner, S. (1933). Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind. *Fund. Math.* **20** 262–176.

[8] Bochner, S. (1941). Hilbert distances and positive definite functions. *Ann. of Math. (2)* **42** 647–656. MR0005782 https://doi.org/10.2307/1969252

[9] Bolin, D. and Kirchner, K. (2023). Equivalence of measures and asymptotically optimal linear prediction for Gaussian random fields with fractional-order covariance operators. *Bernoulli* **29** 1476–1504. MR4550232 https://doi.org/10.3150/22-bej1507

[10] Brault, R., d'Alché, B.F. and Heinonen, M. (2016). Random Fourier features for operator-valued kernels. arXiv preprint. Available at arXiv:1605.02536.

[11] Caponera, A. (2024). Asymptotics for isotropic Hilbert-valued spherical random fields. *Bernoulli* **30** 1723–1745. MR4746586 https://doi.org/10.3150/23-bej1650

[12] Chen, W.Y., Barp, A., Briol, F.-X., Gorham, J., Girolami, M., Mackey, L. and Oates, C. (2019). Stein point Markov chain Monte Carlo. In *ICML* 1011–1021. PMLR.
[13] 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
[14] Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics* **21** 224–239. MR2842240 https://doi.org/10.1002/env.1003
[15] Faouzi, T., Porcu, E. and Bevilacqua, M. (2022). Space-time estimation and prediction under fixed-domain asymptotics with compactly supported covariance functions. *Statist. Sinica* **32** 1187–1203. MR4449890 https://doi.org/10.5705/ss.202020.0010
[16] Ferreira, V., Porcu, E. and Zubelli, J. (2025). Supplement to "Random fields on Hilbert spaces with their equivalent Gaussian measures." https://doi.org/10.3150/24-BEJ1757SUPP
[17] Ghojogh, B., Ghodsi, A., Karray, F. and Crowley, M. (2021). Reproducing kernel Hilbert Space, Mercer's Theorem, eigenfunctions, Nyström method, and use of kernels in machine learning: Tutorial and survey. arXiv preprint. Available at arXiv:2106.08443.
[18] Gikhman, I.I. and Skorokhod, A.V. (2004). *The Theory of Stochastic Processes. I. Classics in Mathematics*. Berlin: Springer. Translated from the Russian by S. Kotz, Reprint of the 1974 edition. MR2058259
[19] 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
[20] Kirchner, K. and Bolin, D. (2022). Necessary and sufficient conditions for asymptotically optimal linear prediction of random fields on compact metric spaces. *Ann. Statist.* **50** 1038–1065. MR4404928 https://doi.org/10.1214/21-aos2138
[21] Li, D., Tang, W. and Banerjee, S. (2023). Inference for Gaussian processes with Matérn covariogram on compact Riemannian manifolds. *J. Mach. Learn. Res.* **24** Paper No. [101], 26. MR4583262
[22] Matérn, B. (1986). *Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*, 2nd ed. Heidelberg: Springer.
[23] Mikusiński, J. (1978). *The Bochner Integral*. *Lehrbücher und Monographien aus dem Gebiete der Exakten Wissenschaften, Mathematische Reihe, Band 55*. Basel-Stuttgart: Birkhäuser. MR0492147
[24] Neeb, K.-H. (1998). Operator-valued positive definite kernels on tubes. *Monatsh. Math.* **126** 125–160. MR1639387 https://doi.org/10.1007/BF01473583
[25] Nourdin, I. and Peccati, G. (2010). Cumulants on the Wiener space. *J. Funct. Anal.* **258** 3775–3791. MR2606872 https://doi.org/10.1016/j.jfa.2009.10.024
[26] 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
[27] Parthasarathy, T. (1972). *Selection Theorems and Their Applications*. *Lecture Notes in Mathematics* **2263**. Berlin: Springer. MR0417368
[28] Peszat, S. and Zabczyk, J. (2007). *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach*. *Encyclopedia of Mathematics and Its Applications* **113**. Cambridge: Cambridge Univ. Press. MR2356959 https://doi.org/10.1017/CBO9780511721373
[29] Pettis, B.J. (1938). On integration in vector spaces. *Trans. Amer. Math. Soc.* **44** 277–304. MR1501970 https://doi.org/10.2307/1989973
[30] Prakasa Rao, B.L.S. (2014). *Nonparametric Functional Estimation*. *Probability and Mathematical Statistics*. New York: Academic Press [Harcourt Brace Jovanovich, Publishers].
[31] Reed, M. and Simon, B. (1975). *Methods of Modern Mathematical Physics. II. Fourier Analysis, Self-Adjointness*. New York-London: Academic Press [Harcourt Brace Jovanovich, Publishers]. MR0493420
[32] Ruiz-Medina, M. and Porcu, E. (2015). Equivalence of Gaussian measures of multivariate random fields. *Stoch. Environ. Res. Risk Assess.* **29** 325–334.
[33] Skorohod, A.V. and Jadrenko, M.Ĭ. (1973). Absolute continuity of measures that correspond to homogeneous Gaussian fields. *Teor. Veroyatn. Primen.* **18** 27–40. MR0312563
[34] 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. Berkeley, CA: Univ. California Press. MR0402873

[35] 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

[36] 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

[37] 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

[38] Stein, M.L. (1999). *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

[39] Tang, W., Zhang, L. and Banerjee, S. (2021). On identifiability and consistency of the nugget in Gaussian spatial process models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 1044–1070. MR4349127 https://doi.org/10.1111/rssb.12472

[40] 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

[41] Wynne, G., Kasprzak, M.J. and Duncan, A.B. (2022). A spectral representation of kernel Stein discrepancy with application to goodness-of-fit tests for measures on infinite dimensional Hilbert spaces. arXiv preprint. Available at arXiv:2206.04552.

[42] Yaglom, A.M. (2004). *An Introduction to the Theory of Stationary Random Functions*, Courier Corporation.

[43] Yosida, K. (2012). *Functional Analysis*. *Classics in Mathematics*. Berlin: Springer.

[44] 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

Bernoulli **31**(3), 2025, 1759–1788
https://doi.org/10.3150/24-BEJ1758

Laplace and saddlepoint approximations in high dimensions

YANBO TANG[1,a] and NANCY REID[2,b]

[1]*Department of Mathematics, Imperial College London, London, UK,* [a]*yanbo.tang@imperial.ac.uk*
[2]*Department of Statistical Sciences, University of Toronto, Toronto, Canada,* [b]*nancym.reid@utoronto.ca*

We examine the behaviour of the Laplace and saddlepoint approximations in the high-dimensional setting, where the dimension of the model is allowed to increase with the number of observations. Approximations to the joint density, the marginal posterior density and the conditional density are considered. Our results show that under the mildest assumptions on the model, the error of the joint density approximation is $O(p^4/n)$ if $p = o(n^{1/4})$ for the Laplace approximation and saddlepoint approximation, and $O(p^3/n)$ if $p = o(n^{1/3})$ under additional assumptions on the second derivative of the log-likelihood. Stronger results are obtained for the approximation to the marginal posterior density.

Keywords: Integral approximations; Laplace approximation; saddlepoint approximation

References

Barber, R.F., Drton, M. and Tan, K.M. (2016). Laplace approximation in high-dimensional Bayesian regression. In *Statistical Analysis for High-Dimensional Data*. *Abel Symp.* **11** 15–36. Cham: Springer. MR3616262 https://doi.org/10.1007/978-3-319-27099-9_2

Bilodeau, B., Stringer, A. and Tang, Y. (2024). Stochastic convergence rates and applications of adaptive quadrature in Bayesian inference. *J. Amer. Statist. Assoc.* **119** 690–700. MR4713923 https://doi.org/10.1080/01621459.2022.2141635

Bilodeau, B., Tang, Y. and Stringer, A. (2023). On the tightness of the Laplace approximation for statistical inference. *Statist. Probab. Lett.* **198** 109839. MR4568825 https://doi.org/10.1016/j.spl.2023.109839

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

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B* **49** 1–39. MR0893334 https://doi.org/10.1111/j.2517-6161.1987.tb01422.x

Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

Davison, A.C. (1988). Approximate conditional inference in generalized linear models. *J. Roy. Statist. Soc. Ser. B* **50** 445–461. MR0970979 https://doi.org/10.1111/j.2517-6161.1988.tb01740.x

Davison, A.C., Fraser, D.A.S., Reid, N. and Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *J. Amer. Statist. Assoc.* **109** 302–314. MR3180565 https://doi.org/10.1080/01621459.2013.839451

de Boor, C. (1978). *A Practical Guide to Splines*. *Applied Mathematical Sciences* **27**. New York-Berlin: Springer. MR0507062

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

Fraser, D.A.S., Reid, N. and Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103** 625–639. MR3551788 https://doi.org/10.1093/biomet/asw022

Grünwald, P. and van Ommen, T. (2017). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal.* **12** 1069–1103. MR3724979 https://doi.org/10.1214/17-BA1085

Huang, C., Di Caterina, C. and Sartori, N. (2022). Directional testing for high dimensional multivariate normal distributions. *Electron. J. Stat.* **16** 6489–6511. MR4522374 https://doi.org/10.1214/22-ejs2089

Jensen, J.L. (2021). On the use of saddlepoint approximations in high dimensional inference. *Sankhya A* **83** 379–392. MR4227217 https://doi.org/10.1007/s13171-019-00188-x

Jiang, W. and Tanner, M.A. (2008). Gibbs posterior for variable selection in high-dimensional classification and data mining. *Ann. Statist.* **36** 2207–2231. MR2458185 https://doi.org/10.1214/07-AOS547

Kasprzak, M.J., Giordano, R. and Broderick, T. (2022). How good is your Gaussian approximation of the posterior? Finite-sample computable error bounds for a variety of useful divergences. Available at arXiv:2209.14992.

Kass, R., Tierney, L. and Kadane, J. (1990). The validity of posterior expansions based on Laplace's method. In *Bayesian and Likelihood Methods in Statistics and Econometrics* (S. Geisser, J.S. Hodges, S.J. Press and A. Zellner, eds.) 473–488. Amsterdam: Elsevier.

Katsevich, A. (2023a). Tight dimension dependence of the Laplace approximation. Available at arXiv:2305.17604.

Katsevich, A. (2023b). Improved scaling with dimension in the Bernstein-von Mises theorem for two statistical models. Available at arXiv:2308.06899.

Kauermann, G., Krivobokova, T. and Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 487–503. MR2649606 https://doi.org/10.1111/j.1467-9868.2008.00691.x

Kolassa, J.E. (2003). Multivariate saddlepoint tail probability approximations. *Ann. Statist.* **31** 274–286. MR1962507 https://doi.org/10.1214/aos/1046294465

Kolassa, J.E. (2006). *Series Approximation Methods in Statistics*. New York: Springer.

Laplace, P.S. and Stigler, S.M. (1986). Memoir on the probability of the causes of events. *Statist. Sci.* **1** 364–378. https://doi.org/10.1214/ss/1177013621

Ogden, H. (2021). On the error in Laplace approximations of high-dimensional integrals. *Stat* **10** e380. MR4276039 https://doi.org/10.1002/sta4.380

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

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

Reid, N. (2003). Asymptotics and the theory of inference. *Ann. Statist.* **31** 1695–1731. MR2036388 https://doi.org/10.1214/aos/1074290325

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

Shun, Z. and McCullagh, P. (1995). Laplace approximation of high-dimensional integrals. *J. Roy. Statist. Soc. Ser. B* **57** 749–760. MR1354079 https://doi.org/10.1111/j.2517-6161.1995.tb02060.x

Spokoiny, V. (2023). Dimension free nonasymptotic bounds on the accuracy of high-dimensional Laplace approximation. *SIAM/ASA J. Uncertain. Quantificat.* **11** 1044–1068. MR4646387 https://doi.org/10.1137/22M1495688

Tang, Y. and Reid, N. (2020). Modified likelihood root in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 1349–1369. MR4176346 https://doi.org/10.1111/rssb.12389

Tang, Y. and Reid, N. (2025). Supplement to "Laplace and saddlepoint approximations in high dimensions." https://doi.org/10.3150/24-BEJ1758SUPP

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86. MR0830567 https://doi.org/10.1080/01621459.1986.10478240

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 3–36. MR2797734 https://doi.org/10.1111/j.1467-9868.2010.00749.x

Bernoulli **31**(3), 2025, 1789–1810
https://doi.org/10.3150/24-BEJ1762

Short-time expansion of characteristic functions in a rough volatility setting with applications

CARSTEN H. CHONG[1,a] and VIKTOR TODOROV[2,b]

[1]*Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology, Hong Kong,* [a]*carstenchong@ust.hk*
[2]*Department of Finance, Northwestern University, USA,* [b]*v-todorov@kellogg.northwestern.edu*

We derive a higher-order asymptotic expansion of the conditional characteristic function of the increment of an Itô semimartingale over a shrinking time interval. The spot characteristics of the Itô semimartingale are allowed to have dynamics of general form. In particular, their paths can be rough, that is, exhibit local behavior like that of a fractional Brownian motion, while at the same time have jumps with arbitrary degree of activity. The expansion result shows the distinct roles played by the different features of the spot characteristics dynamics. As an application of our result, we construct a nonparametric estimator of the Hurst parameter of the diffusive volatility process from portfolios of short-dated options written on an underlying asset.

Keywords: Asymptotic expansion; characteristic function; fractional Brownian motion; Hurst parameter; infinite variation jumps; Itô semimartingale; options; rough volatility

References

[1] Bayer, C., Friz, P.K., Gulisashvili, A., Horvath, B. and Stemper, B. (2019). Short-time near-the-money skew in rough fractional volatility models. *Quant. Finance* **19** 779–798. MR3939657 https://doi.org/10.1080/14697688.2018.1529420
[2] Bennedsen, M., Lunde, A. and Pakkanen, M.S. (2022). Decoupling the short- and long-term behavior of stochastic volatility. *J. Financ. Econom.* **20** 961–1006.
[3] Carr, P., Geman, H., Madan, D.B. and Yor, M. (2003). Stochastic volatility for Lévy processes. *Math. Finance* **13** 345–382. MR1995283 https://doi.org/10.1111/1467-9965.00020
[4] Carr, P. and Madan, D. (2001). Optimal positioning in derivative securities. *Quant. Finance* **1** 19–37. MR1810014 https://doi.org/10.1088/1469-7688/1/1/301
[5] Chong, C.H., Delerue, T. and Li, G. (2024). When frictions are fractional: Rough noise in high-frequency data. *J. Amer. Statist. Assoc.* Forthcoming. https://doi.org/10.1080/01621459.2024.2428466
[6] Chong, C.H., Delerue, T. and Mies, F. (2025). Rate-optimal estimation of mixed semimartingales. *Ann. Statist.* **53** 219–244. MR4865014 https://doi.org/10.1214/24-aos2461
[7] Chong, C.H. and Todorov, V. (2025). Supplement to "Short-time expansion of characteristic functions in a rough volatility setting with applications." https://doi.org/10.3150/24-BEJ1762SUPP
[8] Duffie, D., Pan, J. and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68** 1343–1376. MR1793362 https://doi.org/10.1111/1468-0262.00164
[9] El Euch, O., Fukasawa, M., Gatheral, J. and Rosenbaum, M. (2019). Short-term at-the-money asymptotics under stochastic volatility models. *SIAM J. Financial Math.* **10** 491–511. MR3945237 https://doi.org/10.1137/18M1167565
[10] El Euch, O., Fukasawa, M. and Rosenbaum, M. (2018). The microstructural foundations of leverage effect and rough volatility. *Finance Stoch.* **22** 241–280. MR3778355 https://doi.org/10.1007/s00780-018-0360-z
[11] El Euch, O. and Rosenbaum, M. (2019). The characteristic function of rough Heston models. *Math. Finance* **29** 3–38. MR3905737 https://doi.org/10.1111/mafi.12173
[12] El Euch, O., Gatheral, J. and Rosenbaum, M. (2019). Roughening Heston. *Risk* 84–89.

1350-7265

[13] Forde, M., Gerhold, S. and Smith, B. (2021). Small-time, large-time, and $H \to 0$ asymptotics for the rough Heston model. *Math. Finance* **31** 203–241. MR4205882 https://doi.org/10.1111/mafi.12290

[14] Forde, M., Smith, B. and Viitasaari, L. (2021). Rough volatility, CGMY jumps with a finite history and the rough Heston model—small-time asymptotics in the $k\sqrt{t}$ regime. *Quant. Finance* **21** 541–563. MR4228324 https://doi.org/10.1080/14697688.2020.1790634

[15] Forde, M. and Zhang, H. (2017). Asymptotics for rough stochastic volatility models. *SIAM J. Financial Math.* **8** 114–145. MR3608743 https://doi.org/10.1137/15M1009330

[16] Friz, P.K., Gassiat, P. and Pigato, P. (2021). Precise asymptotics: Robust stochastic volatility models. *Ann. Appl. Probab.* **31** 896–940. MR4254499 https://doi.org/10.1214/20-aap1608

[17] Friz, P.K., Gassiat, P. and Pigato, P. (2022). Short-dated smile under rough volatility: Asymptotics and numerics. *Quant. Finance* **22** 463–480. MR4401811 https://doi.org/10.1080/14697688.2021.1999486

[18] Fukasawa, M. (2021). Volatility has to be rough. *Quant. Finance* **21** 1–8. MR4188876 https://doi.org/10.1080/14697688.2020.1825781

[19] Fukasawa, M., Takabatake, T. and Westphal, R. (2022). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Math. Finance* **32** 1086–1132. MR4485063 https://doi.org/10.1111/mafi.12354

[20] Gatheral, J., Jaisson, T. and Rosenbaum, M. (2018). Volatility is rough. *Quant. Finance* **18** 933–949. MR3805308 https://doi.org/10.1080/14697688.2017.1393551

[21] Jacod, J. and Protter, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Heidelberg: Springer. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

[22] Jacod, J. and Todorov, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Statist.* **42** 1029–1069. MR3210995 https://doi.org/10.1214/14-AOS1213

[23] Jacod, J. and Todorov, V. (2016). Efficient estimation of integrated volatility in presence of infinite variation jumps with multiple activity indices. In *The Fascination of Probability, Statistics and Their Applications* 317–341. Cham: Springer. MR3495691

[24] Jacod, J. and Todorov, V. (2018). Limit theorems for integrated local empirical characteristic exponents from noisy high-frequency data with application to volatility and jump activity estimation. *Ann. Appl. Probab.* **28** 511–576. MR3770883 https://doi.org/10.1214/17-AAP1311

[25] Livieri, G., Mouti, S., Pallavicini, A. and Rosenbaum, M. (2018). Rough volatility: Evidence from option prices. *IISE Trans.* **50** 767–776.

[26] Qin, L. and Todorov, V. (2019). Nonparametric implied Lévy densities. *Ann. Statist.* **47** 1025–1060. MR3909959 https://doi.org/10.1214/18-AOS1703

[27] Rømer, S.E. (2022). Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets. *Quant. Finance* **22** 1805–1838. MR4480680 https://doi.org/10.1080/14697688.2022.2081592

[28] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. Translated from the 1990 Japanese original, Revised by the author. MR1739520

[29] Todorov, V. (2019). Nonparametric spot volatility from options. *Ann. Appl. Probab.* **29** 3590–3636. MR4047988 https://doi.org/10.1214/19-AAP1488

[30] Todorov, V. (2021). Higher-order small time asymptotic expansion of Itô semimartingale characteristic function with application to estimation of leverage from options. *Stochastic Process. Appl.* **142** 671–705. MR4334188 https://doi.org/10.1016/j.spa.2021.09.005

[31] Wang, X., Xiao, W. and Yu, J. (2023). Modeling and forecasting realized volatility with the fractional Ornstein-Uhlenbeck process. *J. Econometrics* **232** 389–415. MR4539492 https://doi.org/10.1016/j.jeconom.2021.08.001

Bernoulli **31**(3), 2025, 1811–1833
https://doi.org/10.3150/24-BEJ1781

On Lasso estimator for the drift function in diffusion models

GABRIELA CIOŁEK[1,a], DMYTRO MARUSHKEVYCH[2,c] and MARK PODOLSKIJ[1,b]

[1]*Department of Mathematics, University of Luxembourg, Esch-sur-Alzette, Luxembourg,*
[a]*gabriela.ciolek@uni.lu,* [b]*mark.podolskij@uni.lu*
[2]*School of Mathematics and Statistics, UNSW, Sydney, Australia ,* [c]*d.marushkevych@unsw.edu.au*

In this paper we study the properties of the Lasso estimator of the drift component in the diffusion setting. More specifically, we consider a multivariate parametric diffusion model X observed continuously over the interval $[0,T]$ and investigate drift estimation under sparsity constraints. We allow the dimensions of the model and the parameter space to be large. We obtain an oracle inequality for the Lasso estimator and derive an error bound for the L^2-distance using concentration inequalities for linear functionals of diffusion processes. The probabilistic part is based upon elements of empirical processes theory and, in particular, on the chaining method.

Keywords: Concentration inequalities; diffusion models; high dimensional statistics; Lasso; parametric estimation

References

[1] Amorino, C., Belomestny, D., Pilipauskaite, V. and Podolskij, M. (2024). Polynomial rates via deconvolution for nonparametric estimation in McKean-Vlasov SDEs. *Probab. Theory Related Fields*. To appear.

[2] Amorino, C., Heidari, A., Pilipauskaitė, V. and Podolskij, M. (2023). Parameter estimation of discretely observed interacting particle systems. *Stochastic Process. Appl.* **163** 350–386. MR4612302 https://doi.org/10.1016/j.spa.2023.06.011

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

[4] Belomestny, D., Pilipauskaitė, V. and Podolskij, M. (2023). Semiparametric estimation of McKean-Vlasov SDEs. *Ann. Inst. Henri Poincaré Probab. Stat.* **59** 79–96. MR4533721 https://doi.org/10.1214/22-aihp1261

[5] Belomestny, D., Podolskij, M. and Zhou, S.-Y. (2024). On nonparametric estimation of the interaction function in particle system models. ArXiv preprint. Available at arXiv:2402.14419.

[6] Bhattacharya, R.N. (1978). Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *Ann. Probab.* **6** 541–553. MR0494525

[7] Bishwal, J.P.N. (2011). Estimation in interacting diffusions: Continuous and discrete sampling. *Appl. Math. (Irvine)* **2** 1154–1158. MR2924983 https://doi.org/10.4236/am.2011.29160

[8] Bolley, F., Cañizo, J.A. and Carrillo, J.A. (2011). Stochastic mean-field limit: Non-Lipschitz forces and swarming. *Math. Models Methods Appl. Sci.* **21** 2179–2210. MR2860672 https://doi.org/10.1142/S0218202511005702

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

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

[11] Carmona, R. and Zhu, X. (2016). A probabilistic approach to mean field games with major and minor players. *Ann. Appl. Probab.* **26** 1535–1580. MR3513598 https://doi.org/10.1214/15-AAP1125

[12] Chen, X. (2021). Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. *Electron. Commun. Probab.* **26** Paper No. 45. MR4284626 https://doi.org/10.1214/21-ecp416

1350-7265

[13] Ciołek, G., Marushkevych, D. and Podolskij, M. (2020). On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein-Uhlenbeck model. *Electron. J. Stat.* **14** 4395–4420. MR4194266 https://doi.org/10.1214/20-EJS1775

[14] De Gregorio, A. and Iacus, S.M. (2012). Adaptive LASSO-type estimation for multivariate diffusion processes. *Econometric Theory* **28** 838–860. MR2959127 https://doi.org/10.1017/S0266466611000806

[15] Della Maestra, L. and Hoffmann, M. (2022). Nonparametric estimation for interacting particle systems: McKean-Vlasov models. *Probab. Theory Related Fields* **182** 551–613. MR4367954 https://doi.org/10.1007/s00440-021-01044-6

[16] Della Maestra, L. and Hoffmann, M. (2023). The LAN property for McKean-Vlasov models in a mean-field regime. *Stochastic Process. Appl.* **155** 109–146. MR4503434 https://doi.org/10.1016/j.spa.2022.10.002

[17] Dexheimer, N. and Strauch, C. (2024). On Lasso and Slope drift estimators for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli* **30** 88–116. MR4665571 https://doi.org/10.3150/22-bej1574

[18] Dirksen, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.* **20** no. 53. MR3354613 https://doi.org/10.1214/EJP.v20-3760

[19] Faugeras, O., Touboul, J. and Cessac, B. (2009). A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. *Front. Comput. Neurosci.* **3** 1–28.

[20] Fujimori, K. (2019). The Dantzig selector for a linear model of diffusion processes. *Stat. Inference Stoch. Process.* **22** 475–498. MR3996026 https://doi.org/10.1007/s11203-018-9191-y

[21] Gaïffas, S. and Matulewicz, G. (2019). Sparse inference of the drift of a high-dimensional Ornstein-Uhlenbeck process. *J. Multivariate Anal.* **169** 1–20. MR3875583 https://doi.org/10.1016/j.jmva.2018.08.005

[22] Gao, F., Guillin, A. and Wu, L. (2014). Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.* **58** 358–382. MR3403002 https://doi.org/10.1137/S0040585X97986667

[23] Genon-Catalot, V. and Larédo, C. (2021). Parametric inference for small variance and long time horizon McKean-Vlasov diffusion models. *Electron. J. Stat.* **15** 5811–5854. MR4355698 https://doi.org/10.1214/21-ejs1922

[24] Genon-Catalot, V. and Larédo, C. (2024). Parametric inference for ergodic McKean-Vlasov stochastic differential equations. *Bernoulli* **30** 1971–1997. MR4746596 https://doi.org/10.3150/23-bej1660

[25] Guillin, A., Léonard, C., Wu, L. and Yao, N. (2009). Transportation-information inequalities for Markov processes. *Probab. Theory Related Fields* **144** 669–695. MR2496446 https://doi.org/10.1007/s00440-008-0159-5

[26] Jacod, J. and Protter, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Heidelberg: Springer. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

[27] Kasonga, R.A. (1990). Maximum likelihood theory for large interacting systems. *SIAM J. Appl. Math.* **50** 865–875. MR1050917 https://doi.org/10.1137/0150050

[28] Küchler, U. and Sørensen, M. (1997). *Exponential Families of Stochastic Processes*. *Springer Series in Statistics*. New York: Springer. MR1458891 https://doi.org/10.1007/b98954

[29] Kutoyants, Y.A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. *Springer Series in Statistics*. London: Springer London, Ltd. MR2144185 https://doi.org/10.1007/978-1-4471-3866-2

[30] Liu, M. and Qiao, H. (2022). Parameter estimation of path-dependent McKean-Vlasov stochastic differential equations. *Acta Math. Sci. Ser. B Engl. Ed.* **42** 876–886. MR4411008 https://doi.org/10.1007/s10473-022-0304-8

[31] McKean, H.P. Jr. (1966). Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. *Arch. Ration. Mech. Anal.* **21** 343–367. MR0214112 https://doi.org/10.1007/BF00264463

[32] McKean, H.P. Jr. (1967). Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*. *Lecture Series in Differential Equations, Session 7* 41–57. Office of Aerospace Research, United States Air Force, Arlington, VA: Air Force Office of Scientific Research. MR0233437

[33] Periera, J.B.A. and Ibrahimi, M. (2014). Support recovery for the drift coefficient of high-dimensional diffusions. *IEEE Trans. Inf. Theory* **60** 4026–4049. MR3225948 https://doi.org/10.1109/TIT.2014.2317178

[34] Rogers, L.C.G. and Williams, D. (1987). *Diffusions, Markov Processes, and Martingales. Vol. 2: Itô Calculus*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0921238

[35] Saussereau, B. (2012). Transportation inequalities for stochastic differential equations driven by a fractional Brownian motion. *Bernoulli* **18** 1–23. MR2888696 https://doi.org/10.3150/10-BEJ324
[36] Sharrock, L., Kantas, N., Parpas, P. and Pavliotis, G.A. (2023). Online parameter estimation for the McKean-Vlasov stochastic differential equation. *Stochastic Process. Appl.* **162** 481–546. MR4597535 https://doi.org/10.1016/j.spa.2023.05.002
[37] Spokoiny, V. (2017). Penalized maximum likelihood estimation and effective dimension. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 389–429. MR3606746 https://doi.org/10.1214/15-AIHP720
[38] Sznitman, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX—1989*. *Lecture Notes in Math.* **1464** 165–251. Berlin: Springer. MR1108185 https://doi.org/10.1007/BFb0085169
[39] Talagrand, M. (2005). *The Generic Chaining*. *Springer Monographs in Mathematics*. Berlin: Springer. MR2133757
[40] Talagrand, M. (2014). *Upper and Lower Bounds for Stochastic Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **60**. Heidelberg: Springer. MR3184689 https://doi.org/10.1007/978-3-642-54075-2
[41] Trottner, L., Aeckerle-Willems, C. and Strauch, C. (2023). Concentration analysis of multivariate elliptic diffusions. *J. Mach. Learn. Res.* **24** Paper No. [106]. MR4583267
[42] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. MR2724359 https://doi.org/10.1007/b13794
[43] 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
[44] Varvenne, M. (2019). Concentration inequalities for stochastic differential equations with additive fractional noise. *Electron. J. Probab.* **24** Paper No. 124. MR4029427 https://doi.org/10.1214/19-ejp384
[45] Veraar, M. (2012). The stochastic Fubini theorem revisited. *Stochastics* **84** 543–551. MR2966093 https://doi.org/10.1080/17442508.2011.618883
[46] Wen, J., Wang, X., Mao, S. and Xiao, X. (2016). Maximum likelihood estimation of McKean-Vlasov stochastic differential equation and its application. *Appl. Math. Comput.* **274** 237–246. MR3433132 https://doi.org/10.1016/j.amc.2015.11.019

Bernoulli **31**(3), 2025, 1834–1857
https://doi.org/10.3150/24-BEJ1788

Optimal Wasserstein-1 distance between SDEs driven by Brownian motion and stable processes

CHANG-SONG DENG[1,a], RENÉ L. SCHILLING[2,b] and LIHU XU[3,c]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan, China,* [a]*dengcs@whu.edu.cn*
[2]*Institut für Mathematische Stochastik, Fakultät Mathematik, TU Dresden, Dresden, Germany,* [b]*rene.schilling@tu-dresden.de*
[3]*Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau S.A.R., China,* [c]*lihuxu@um.edu.mo*

We are interested in the following two $\mathbb{R}^d$-valued stochastic differential equations (SDEs):

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}L_t, \quad X_0 = x, \\ \mathrm{d}Y_t &= b(Y_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t, \quad Y_0 = y, \end{aligned}$$

where σ is an invertible $d \times d$ matrix, L_t is a rotationally symmetric α-stable Lévy process, and B_t is a d-dimensional standard Brownian motion (note that B_t is a rotationally symmetric α-stable Lévy process with $\alpha = 2$). We show that for any $\alpha_0 \in (1,2)$ the Wasserstein-1 distance W_1 satisfies for $\alpha \in [\alpha_0, 2)$

$$W_1\left(\mathrm{law}(X_t^x), \mathrm{law}(Y_t^y)\right) \le C_1 \mathrm{e}^{-C_2 t}|x-y| + \frac{C}{\alpha_0 - 1}(2-\alpha) d \log(1+d),$$

which implies, in particular,

$$W_1(\mu_\alpha, \mu_2) \le \frac{C}{\alpha_0 - 1}(2-\alpha) d \log(1+d),$$

where μ_α and μ_2 are the ergodic measures of X_t and Y_t respectively. For the special case of a d-dimensional Ornstein–Uhlenbeck system, we show that $W_1(\mu_\alpha, \mu_2) \ge C_d(2-\alpha)$ for all $\alpha \in (1,2)$; this indicates that the convergence rate with respect to α in the previous bound is optimal. The term $d\log(1+d)$ appearing in the bound seems to also be optimal for the dimension d.

Keywords: Optimal convergence rate; SDE driven by α-stable processes; SDE driven by Brownian motions; Wasserstein-1 distance

References

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

Chen, Z.-Q., Hao, Z. and Zhang, X. (2020). Hölder regularity and gradient estimates for SDEs driven by cylindrical α-stable processes. *Electron. J. Probab.* **25** Paper No. 137, 23. MR4179301 https://doi.org/10.1214/20-ejp542

Chen, P., Shao, Q.-M. and Xu, L. (2023). A probability approximation framework: Markov process approach. *Ann. Appl. Probab.* **33** 1419–1459. MR4564436 https://doi.org/10.1214/22-aap1853

Chen, Z.-Q., Zhang, X. and Zhao, G. (2021). Supercritical SDEs driven by multiplicative stable-like Lévy processes. *Trans. Amer. Math. Soc.* **374** 7621–7655. MR4328678 https://doi.org/10.1090/tran/8343

Chen, P., Deng, C.-S., Schilling, R.L. and Xu, L. (2023). Approximation of the invariant measure of stable SDEs by an Euler-Maruyama scheme. *Stochastic Process. Appl.* **163** 136–167. MR4610124 https://doi.org/10.1016/j.spa.2023.06.001

1350-7265

Deng, C.-S. and Schilling, R.L. (2019). Exact asymptotic formulas for the heat kernels of space and time-fractional equations. *Fract. Calc. Appl. Anal.* **22** 968–989. MR4023103 https://doi.org/10.1515/fca-2019-0052

Deng, C.-S., Schilling, R.L. and Xu, L. (2022). Singular integrals of subordinators with applications to structural properties of SPDEs. *Trans. Amer. Math. Soc.* **375** 6043–6073. MR4474884 https://doi.org/10.1090/tran/8595

Dong, Z., Xu, L. and Zhang, X. (2014). Exponential ergodicity of stochastic Burgers equations driven by α-stable processes. *J. Stat. Phys.* **154** 929–949. MR3164597 https://doi.org/10.1007/s10955-013-0881-y

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

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

Engel, K.-J. and Nagel, R. (2000). *One-Parameter Semigroups for Linear Evolution Equations*. *Graduate Texts in Mathematics* **194**. New York: Springer. MR1721989

Grobys, K. (2021). What do we know about the second moment of financial markets? *Int. Rev. Financ. Anal.* **78** 101891. https://doi.org/10.1016/j.irfa.2021.101891

Hairer, M. (2021). The convergence of Markov processes. (accessed June 21, 2024). https://www.hairer.org/notes/Convergence.pdf.

Janicki, A., Michna, Z. and Weron, A. (1996). Approximation of stochastic differential equations driven by α-stable Lévy motion. *Appl. Math. (Warsaw)* **24** 149–168. MR1424549 https://doi.org/10.4064/am-24-2-149-168

Kühn, F. and Schilling, R.L. (2019). Strong convergence of the Euler-Maruyama approximation for a class of Lévy-driven SDEs. *Stochastic Process. Appl.* **129** 2654–2680. MR3980140 https://doi.org/10.1016/j.spa.2018.07.018

Kwaśnicki, M. (2017). Ten equivalent definitions of the fractional Laplace operator. *Fract. Calc. Appl. Anal.* **20** 7–51. MR3613319 https://doi.org/10.1515/fca-2017-0002

Liang, M. and Wang, J. (2020). Gradient estimates and ergodicity for SDEs driven by multiplicative Lévy noises via coupling. *Stochastic Process. Appl.* **130** 3053–3094. MR4080736 https://doi.org/10.1016/j.spa.2019.09.001

Liu, X. (2022a). On the α-dependence of stochastic differential equations with Hölder drift and driven by α-stable Lévy processes. *J. Math. Anal. Appl.* **506** Paper No. 125642, 23. MR4314064 https://doi.org/10.1016/j.jmaa.2021.125642

Liu, X. (2022b). Limits of invariant measures of stochastic Burgers equations driven by two kinds of α-stable processes. *Stochastic Process. Appl.* **146** 1–21. MR4365030 https://doi.org/10.1016/j.spa.2021.12.016

Liu, X. (2022c). The α-dependence of the invariant measure of stochastic real Ginzburg-Landau equation driven by α-stable Lévy processes. *J. Differ. Equ.* **314** 418–445. MR4367871 https://doi.org/10.1016/j.jde.2022.01.024

Liu, W., Song, R. and Xie, L. (2020). Gradient estimates for the fundamental solution of Lévy type operator. *Adv. Nonlinear Anal.* **9** 1453–1462. MR4078178 https://doi.org/10.1515/anona-2020-0062

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

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

Peligrad, M., Sang, H., Xiao, Y. and Yang, G. (2022). Limit theorems for linear random fields with innovations in the domain of attraction of a stable law. *Stochastic Process. Appl.* **150** 596–621. MR4433071 https://doi.org/10.1016/j.spa.2022.05.003

Priola, E. and Zabczyk, J. (2011). Structural properties of semilinear SPDEs driven by cylindrical stable processes. *Probab. Theory Related Fields* **149** 97–137. MR2773026 https://doi.org/10.1007/s00440-009-0243-5

Priola, E., Shirikyan, A., Xu, L. and Zabczyk, J. (2012). Exponential ergodicity and regularity for equations with Lévy noise. *Stochastic Process. Appl.* **122** 106–133. MR2860444 https://doi.org/10.1016/j.spa.2011.10.003

Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (Reprint of the 10th Edition)*. New York: Dover Publications. MR0167642

Wang, J. (2016). L^p-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli* **22** 1598–1616. MR3474827 https://doi.org/10.3150/15-BEJ705

Wang, R., Xiong, J. and Xu, L. (2017). Irreducibility of stochastic real Ginzburg-Landau equation driven by α-stable noises and applications. *Bernoulli* **23** 1179–1201. MR3606763 https://doi.org/10.3150/15-BEJ773

Xu, L. (2013). Ergodicity of the stochastic real Ginzburg-Landau equation driven by α-stable noises. *Stochastic Process. Appl.* **123** 3710–3736. MR3084156 https://doi.org/10.1016/j.spa.2013.05.002

Zhang, X. (2013). Derivative formulas and gradient estimates for SDEs driven by α-stable processes. *Stochastic Process. Appl.* **123** 1213–1228. MR3016221 https://doi.org/10.1016/j.spa.2012.11.012

Bernoulli **31**(3), 2025, 1858–1888
https://doi.org/10.3150/24-BEJ1789

From dense to sparse design: Optimal rates under the supremum norm for estimating the mean function in functional data analysis

MAX BERGER[a], PHILIPP HERMANN[b] and HAJO HOLZMANN[c]

Philipps-Universität, Marburg, Germany, [a]*mberger@mathematik.uni-marburg.de,* [b]*herm@mathematik.uni-marburg.de,* [c]*holzmann@mathematik.uni-marburg.de*

We derive optimal rates of convergence in the supremum norm for estimating the Hölder-smooth mean function of a stochastic process which is repeatedly and discretely observed with additional errors at fixed, multivariate, synchronous design points, the typical scenario for machine recorded functional data. Similarly to the optimal rates in L_2 obtained in (Cai and Yuan, *Ann. Statist.* **39** (2011) 2330–2355), for sparse design a discretization term dominates, while in the dense case the parametric $\sqrt{n}$ rate can be achieved as if the n processes were continuously observed without errors. The supremum norm is of practical interest since it corresponds to the visualization of the estimation error, and forms the basis for the construction of uniform confidence bands. We show that in contrast to the analysis in L_2, there is an intermediate regime between the sparse and dense cases dominated by the contribution of the observation errors. Furthermore, under the supremum norm interpolation estimators which suffice in L_2 turn out to be sub-optimal in the presence of observation errors in the dense setting, which helps to explain their poor empirical performance. In contrast to previous contributions involving the supremum norm, we discuss optimality even in the multivariate setting, and for dense design obtain the $\sqrt{n}$ rate of convergence without additional logarithmic factors. We also derive a central limit theorem in the supremum norm, and provide simulations and real data applications to illustrate our results.

Keywords: Asymptotic confidence sets; dense and sparse design; functional data; minimax optimality; supremum norm

References

Berger, M., Hermann, P. and Holzmann, H. (2025). Supplement to "From dense to sparse design: Optimal rates under the supremum norm for estimating the mean function in functional data analysis." https://doi.org/10.3150/24-BEJ1789SUPP

Cai, T.T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. MR2906870 https://doi.org/10.1214/11-AOS898

Chang, C., Lin, X. and Ogden, R.T. (2017). Simultaneous confidence bands for functional regression models. *J. Statist. Plann. Inference* **188** 67–81. MR3648318 https://doi.org/10.1016/j.jspi.2017.03.002

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference* **147** 1–23. MR3151843 https://doi.org/10.1016/j.jspi.2013.04.002

Degras, D.A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21** 1735–1765. MR2895997 https://doi.org/10.5705/ss.2009.207

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

Kalogridis, I. and Van Aelst, S. (2023). Robust optimal estimation of location from discretely sampled functional data. *Scand. J. Stat.* **50** 411–451. MR4599920

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

Liebl, D. and Reimherr, M. (2023). Fast and fair simultaneous confidence bands for functional parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 842–868. MR4745945 https://doi.org/10.1093/jrsssb/qkad026

Morris, J.S. (2015). Functional regression. *Annu. Rev. Stat. Appl.* **2** 321–359.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **2**. Hayward, CA: IMS. MR1089429

Ramsay, J. and Silvermann, B. (1998). *Functional Data Analysis*. Springer.

Tsybakov, A.B. (2004). *Introduction à L'estimation Non-paramétrique*. *Mathématiques & Applications (Berlin) [Mathematics & Applications]* **41**. Berlin: Springer. MR2013911

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes—with Applications to Statistics*. Springer Science & Business Media.

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

Wang, Y., Wang, G., Wang, L. and Ogden, R.T. (2020). Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics* **76** 427–437. MR4125268 https://doi.org/10.1111/biom.13156

Xiao, L. (2020). Asymptotic properties of penalized splines for functional data. *Bernoulli* **26** 2847–2875. MR4140531 https://doi.org/10.3150/20-BEJ1209

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. MR3546451 https://doi.org/10.1214/16-AOS1446

Bernoulli **31**(3), 2025, 1889–1907
https://doi.org/10.3150/24-BEJ1790

Optimal scaling results for Moreau-Yosida Metropolis-adjusted Langevin algorithms

FRANCESCA R. CRUCINIO[1,a], ALAIN DURMUS[2,b], PABLO JIMÉNEZ[3,c] and GARETH O. ROBERTS[4,d]

[1]*King's College London, London, UK,* [a]*francesca_romana.crucinio@kcl.ac.uk*
[2]*Centre de Mathématiques Appliquées, Ecole Polytechnique, Institut Polytechnique de Paris, Paris, France,* [b]*alain.durmus@polytechnique.edu*
[3]*Sorbonne Université and Université Paris Cité, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, F-75005 Paris, France,* [c]*pablo.jimenez-moreno@polytechnique.edu*
[4]*Department of Statistics, University of Warwick, Coventry, UK,* [d]*gareth.o.roberts@warwick.ac.uk*

We consider a recently proposed class of MCMC methods which uses proximity maps instead of gradients to build proposal mechanisms which can be employed for both differentiable and non-differentiable targets. These methods have been shown to be stable for a wide class of targets, making them a valuable alternative to Metropolis-adjusted Langevin algorithms (MALA), and have found wide application in imaging contexts. The wider stability properties are obtained by building the Moreau-Yosida envelope for the target of interest, which depends on a parameter λ. In this work, we investigate the optimal scaling problem for this class of algorithms, which encompasses MALA, and provide practical guidelines for the implementation of these methods.

Keywords: Laplace distribution; Markov chain Monte Carlo; proximity map

References

[1] Agrawal, S., Vats, D., Łatuszyński, K. and Roberts, G.O. (2023). Optimal scaling of MCMC beyond Metropolis. *Adv. in Appl. Probab.* **55** 492–509. MR4583786 https://doi.org/10.1017/apr.2022.37

[2] Atchadé, Y.F. (2015). A Moreau-Yosida approximation scheme for a class of high-dimensional posterior distributions. arXiv preprint, arXiv:1505.07072.

[3] Barbero, A. and Sra, S. (2011). Fast Newton-type methods for total variation regularization. In *ICML* (L. Getoor and T. Scheffer, eds.) 313–320. Omnipress.

[4] Barbero, Á. and Sra, S. (2018). Modular proximal optimization for multidimensional total-variation regularization. *J. Mach. Learn. Res.* **19** Paper No. 56, 82. MR3899758

[5] Bauschke, H.H. and Combettes, P.L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC*. New York: Springer. With a foreword by Hédy Attouch. MR2798533 https://doi.org/10.1007/978-1-4419-9467-7

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

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

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

[10] 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

1350-7265 This research was funded, in whole or in part, by EPSRC, EP/R034710/1.

[11] Brooks, S., Gelman, A., Jones, G.L. and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. Boca Raton, FL: CRC Press. MR2742422 https://doi.org/10.1201/b10905

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

[13] Combettes, P.L. and Pesquet, J.-C. (2011). Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. (R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke and H. Wolkowicz, eds.). *Springer Optim. Appl.* **49** 185–212. New York: Springer. MR2858838 https://doi.org/10.1007/978-1-4419-9569-8_10

[14] Crucinio, F.R., Durmus, A., Jiménez, P. and Roberts, G.O. (2025). Supplement to "Optimal scaling results for Moreau-Yosida Metropolis-adjusted Langevin algorithms." https://doi.org/10.3150/24-BEJ1790SUPP

[15] Durmus, A., Le Corff, S., Moulines, E. and Roberts, G.O. (2017). Optimal scaling of the random walk Metropolis algorithm under L^p mean differentiability. *J. Appl. Probab.* **54** 1233–1260. MR3731293 https://doi.org/10.1017/jpr.2017.61

[16] Durmus, A. and Moulines, E. (2022). On the geometric convergence for MALA under verifiable conditions. arXiv preprint, arXiv:2201.01951.

[17] Durmus, A., Moulines, É. and Pereyra, M. (2018). Efficient Bayesian computation by proximal Markov chain Monte Carlo: When Langevin meets Moreau. *SIAM J. Imaging Sci.* **11** 473–506. MR3763089 https://doi.org/10.1137/16M1108340

[18] Dwivedi, R., Chen, Y., Wainwright, M.J. and Yu, B. (2019). Log-concave sampling: Metropolis-Hastings algorithms are fast. In *Proceedings of the 31st Conference on Learning Theory* (S. Bubeck, V. Perchet and P. Rigollet, eds.). *Proceedings of Machine Learning Research* **75** 793–797. PMLR. MR4048994

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

[20] Grenander, U. and Miller, M.I. (1994). Representations of knowledge in complex systems. *J. Roy. Statist. Soc. Ser. B* **56** 549–603. With discussion and a reply by the authors. MR1293234

[21] Jarner, S.F. and Roberts, G.O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.* **34** 781–815. MR2396939 https://doi.org/10.1111/j.1467-9469.2007.00557.x

[22] Jourdain, B., Lelièvre, T. and Miasojedow, B. (2014). Optimal scaling for the transient phase of Metropolis Hastings algorithms: The longtime behavior. *Bernoulli* **20** 1930–1978. MR3263094 https://doi.org/10.3150/13-BEJ546

[23] Livingstone, S. and Zanella, G. (2022). The Barker proposal: Combining robustness and efficiency in gradient-based MCMC. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 496–523. MR4412995 https://doi.org/10.1111/rssb.12482

[24] Louchet, C. and Moisan, L. (2008). Total variation denoising using posterior expectation. In *2008 16th European Signal Processing Conference* 1–5. IEEE.

[25] Ma, Y.-A., Chen, T. and Fox, E. (2015). A complete recipe for stochastic gradient MCMC. *NeurIPS* **28**.

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

[27] Neal, P. and Roberts, G. (2008). Optimal scaling for random walk Metropolis on spherically constrained target densities. *Methodol. Comput. Appl. Probab.* **10** 277–297. MR2399684 https://doi.org/10.1007/s11009-007-9046-2

[28] Neal, P., Roberts, G. and Yuen, W.K. (2012). Optimal scaling of random walk Metropolis algorithms with discontinuous target densities. *Ann. Appl. Probab.* **22** 1880–1927. MR3025684 https://doi.org/10.1214/11-AAP817

[29] Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Found. Trends Optim.* **1** 127–239.

[30] Pereyra, M. (2016). Proximal Markov chain Monte Carlo algorithms. *Stat. Comput.* **26** 745–760. MR3515019 https://doi.org/10.1007/s11222-015-9567-4

[31] Pillai, N.S. (2022). Optimal scaling for the proximal Langevin algorithm in high dimensions. arXiv preprint, arXiv:2204.10793.

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

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

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

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

[36] 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

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

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

[39] Rossky, P.J., Doll, J.D. and Friedman, H.L. (1978). Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.* **69** 4628–4633.

[40] Schmidt, M., Roux, N. and Bach, F. (2011). Convergence rates of inexact proximal-gradient methods for convex optimization. *NeurIPS* **24**.

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

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

[43] Vono, M., Dobigeon, N. and Chainais, P. (2019). Bayesian image restoration under Poisson noise and log-concave prior. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1712–1716. IEEE.

[44] Vorstrup Goldman, J., Sell, T. and Singh, S.S. (2022). Gradient-based Markov chain Monte Carlo for Bayesian inference with non-differentiable priors. *J. Amer. Statist. Assoc.* **117** 2182–2193. MR4528497 https://doi.org/10.1080/01621459.2021.1909600

[45] Yang, J., Roberts, G.O. and Rosenthal, J.S. (2020). Optimal scaling of random-walk Metropolis algorithms on general target distributions. *Stochastic Process. Appl.* **130** 6094–6132. MR4140028 https://doi.org/10.1016/j.spa.2020.05.004

[46] Zhou, X., Heng, Q., Chi, E.C. and Zhou, H. (2024). Proximal MCMC for Bayesian inference of constrained and regularized estimation. *Amer. Statist.* just-accepted 1–24.

Bernoulli **31**(3), 2025, 1908–1928
https://doi.org/10.3150/24-BEJ1791

Lower bounds on the rate of convergence for accept-reject-based Markov chains in Wasserstein and total variation distances

AUSTIN BROWN[1,a] and GALIN JONES[2,b]

[1]*Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada,* [a]*ad.brown@utoronto.ca*
[2]*School of Statistics, University of Minnesota, Minneapolis, MN, U.S.A.,* [b]*galin@umn.edu*

To avoid poor empirical performance in Metropolis-Hastings and other accept-reject-based algorithms practitioners often tune them by trial and error. Lower bounds on the convergence rate are developed in both total variation and Wasserstein distances in order to identify how the simulations will fail so these settings can be avoided, providing guidance on tuning. Particular attention is paid to using the lower bounds to study the convergence complexity of accept-reject-based Markov chains and to constrain the rate of convergence for geometrically ergodic Markov chains. The theory is applied in several settings. For example, if the target density concentrates with a parameter n (e.g. posterior concentration, Laplace approximations), it is demonstrated that the convergence rate of a Metropolis-Hastings chain can be arbitrarily slow if the tuning parameters do not depend carefully on n. This is demonstrated with Bayesian logistic regression with Zellner's g-prior when the dimension and sample increase together and flat prior Bayesian logistic regression as n tends to infinity.

Keywords: Bayesian statistics; convergence complexity; Markov chain Monte Carlo; Metropolis-Hastings; Wasserstein distance

References

[1] Agrawal, S., Vats, D., Łatuszyński, K. and Roberts, G.O. (2023). Optimal scaling of MCMC beyond Metropolis. *Adv. in Appl. Probab.* **55** 492–509. MR4583786 https://doi.org/10.1017/apr.2022.37

[2] Andrieu, C., Lee, A., Power, S. and Wang, A.Q. (2024). Explicit convergence bounds for Metropolis Markov chains: Isoperimetry, spectral gaps and profiles. *Ann. Appl. Probab.* **34** 4022–4071. MR4783036 https://doi.org/10.1214/24-aap2058

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

[4] Barker, A.A. (1964). Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Aust. J. Phys.* **18** 119–132.

[5] Belloni, A. and Chernozhukov, V. (2009). On the computational complexity of MCMC-based estimators in large samples. *Ann. Statist.* **37** 2011–2055. MR2533478 https://doi.org/10.1214/08-AOS634

[6] Bezener, M., Hughes, J. and Jones, G. (2018). Bayesian spatiotemporal modeling using hierarchical spatial priors, with applications to functional magnetic resonance imaging (with discussion). *Bayesian Anal.* **13** 1261–1313. MR3882358 https://doi.org/10.1214/18-BA1108

[7] Bhattacharya, R. and Jones, G.L. (2023). Explicit constraints on the geometric rate of convergence of random walk Metropolis-Hastings. Preprint. Available at arXiv:2307.11644.

[8] Bierkens, J. (2016). Non-reversible Metropolis-Hastings. *Stat. Comput.* **26** 1213–1228. MR3538633 https://doi.org/10.1007/s11222-015-9598-x

[9] Bou-Rabee, N. and Eberle, A. (2020). Markov chain Monte Carlo methods.

[10] Brooks, S.P., Gelman, A., Jones, G.L. and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman & Hall.

[11] Brown, A. and Jones, G. (2025). Supplement to "Lower bounds on the rate of convergence for accept-reject-based Markov chains in Wasserstein and total variation distances." https://doi.org/10.3150/24-BEJ1791SUPP
[12] Brown, A. and Jones, G.L. (2024). Exact convergence analysis for Metropolis-Hastings independence samplers in Wasserstein distances. *J. Appl. Probab.* **61** 33–54. MR4706676 https://doi.org/10.1017/jpr.2023.21
[13] Chen, M.-H. and Shao, Q.-M. (2001). Propriety of posterior distribution for dichotomous quantal response models. *Proc. Amer. Math. Soc.* **129** 293–302. MR1694452 https://doi.org/10.1090/S0002-9939-00-05513-1
[14] Chewi, S., Lu, C., Ahn, K., Cheng, X., Gouic, T.L. and Rigollet, P. (2021). Optimal dimension dependence of the Metropolis-adjusted Langevin algorithm. In *Proceedings of Thirty Fourth Conference on Learning Theory* **134** 1260–1300.
[15] Dalalyan, A.S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 https://doi.org/10.1111/rssb.12183
[16] Durmus, A. and Moulines, É. (2015). Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis adjusted Langevin algorithm. *Stat. Comput.* **25** 5–19. MR3304898 https://doi.org/10.1007/s11222-014-9511-z
[17] Dwivedi, R., Chen, Y., Wainwright, M.J. and Yu, B. (2019). Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.* **20** Paper No. 183. MR4048994
[18] Ekvall, K.O. and Jones, G.L. (2021). Convergence analysis of a collapsed Gibbs sampler for Bayesian vector autoregressions. *Electron. J. Stat.* **15** 691–721. MR4202497 https://doi.org/10.1214/21-EJS1800
[19] Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13** 342–368. MR0773172 https://doi.org/10.1214/aos/1176346597
[20] Fort, G., Moulines, E., Roberts, G.O. and Rosenthal, J.S. (2003). On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.* **40** 123–146. MR1953771 https://doi.org/10.1239/jap/1044476831
[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] Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504 https://doi.org/10.2307/3318737
[23] Hairer, M. and Mattingly, J.C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability* **63** 109–117. Basel: Birkhäuser/Springer Basel AG. MR2857021 https://doi.org/10.1007/978-3-0348-0021-1_7
[24] Hairer, M., Stuart, A.M. and Vollmer, S.J. (2014). Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.* **24** 2455–2490. MR3262508 https://doi.org/10.1214/13-AAP982
[25] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97
[26] Herbei, R. and McKeague, I.W. (2009). Hybrid samplers for ill-posed inverse problems. *Scand. J. Stat.* **36** 839–853. MR2573311 https://doi.org/10.1111/j.1467-9469.2009.00649.x
[27] Jarner, S.F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4
[28] Jarner, S.F. and Yuen, W.K. (2004). Conductance bounds on the L^2 convergence rate of Metropolis algorithms on unbounded state spaces. *Adv. in Appl. Probab.* **36** 243–266. MR2035782 https://doi.org/10.1239/aap/1077134472
[29] Johndrow, J.E., Smith, A., Pillai, N. and Dunson, D.B. (2019). MCMC for imbalanced categorical data. *J. Amer. Statist. Assoc.* **114** 1394–1403. MR4011787 https://doi.org/10.1080/01621459.2018.1505626
[30] Johnson, A.A., Jones, G.L. and Neath, R.C. (2013). Component-wise Markov chain Monte Carlo: Uniform and geometric ergodicity under mixing and composition. *Statist. Sci.* **28** 360–375. MR3135537 https://doi.org/10.1214/13-STS423
[31] Jones, G.L. (2004). On the Markov chain central limit theorem. *Probab. Surv.* **1** 299–320. MR2068475 https://doi.org/10.1214/154957804100000051
[32] Jones, G.L. and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447 https://doi.org/10.1214/ss/1015346317
[33] Jones, G.L. and Hobert, J.P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Ann. Statist.* **32** 784–817. MR2060178 https://doi.org/10.1214/009053604000000184

[34] Jones, G.L., Roberts, G.O. and Rosenthal, J.S. (2014). Convergence of conditional Metropolis-Hastings samplers. *Adv. in Appl. Probab.* **46** 422–445. MR3215540 https://doi.org/10.1239/aap/1401369701
[35] Joulin, A. and Ollivier, Y. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** 2418–2442. MR2683634 https://doi.org/10.1214/10-AOP541
[36] Kantorovič, L.V. and Rubinšteĭn, G.Š. (1958). On a space of completely additive functions. *Vestn. Leningr. Univ.* **13** 52–59. MR0102006
[37] Kass, R.E., Tierney, L. and Kadane, J.B. (1990). The validity of posterior expansions based on Laplace's method. In *Essays in Honor of George Bernard* 473–488.
[38] Łatuszyński, K., Miasojedow, B. and Niemiro, W. (2013). Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli* **19** 2033–2066. MR3129043 https://doi.org/10.3150/12-BEJ442
[39] Łatuszyński, K. and Roberts, G.O. (2013). CLTs and asymptotic variance of time-sampled Markov chains. *Methodol. Comput. Appl. Probab.* **15** 237–247. MR3030220 https://doi.org/10.1007/s11009-011-9237-8
[40] Lawler, G.F. and Sokal, A.D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309** 557–580. MR0930082 https://doi.org/10.2307/2000925
[41] Lee, Y.T., Shen, R. and Tian, K. (2021). Lower bounds on metropolized sampling methods for well-conditioned distributions. In *Advances in Neural Information Processing Systems* **34** 18812–18824.
[42] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882 https://doi.org/10.1214/aos/1033066201
[43] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machine. *J. Chem. Phys.* **21**.
[44] Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2509253 https://doi.org/10.1017/CBO9780511626630
[45] Miller, J.W. (2021). Asymptotic normality, concentration, and coverage of generalized posteriors. *J. Mach. Learn. Res.* **22** Paper No. 168. MR4318524
[46] Peskun, P.H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** 607–612. MR0362823 https://doi.org/10.1093/biomet/60.3.607
[47] Qin, Q. and Hobert, J.P. (2019). Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression. *Ann. Statist.* **47** 2320–2347. MR3953453 https://doi.org/10.1214/18-AOS1749
[48] Qin, Q. and Jones, G.L. (2022). Convergence rates of two-component MCMC samplers. *Bernoulli* **28** 859–885. MR4388922 https://doi.org/10.3150/21-bej1369
[49] Rajaratnam, B. and Sparks, D. (2015). MCMC-based inference in the era of big data: A fundamental analysis of the convergence complexity of high-dimensional chains. Preprint. Available at arXiv:1508.00947.
[50] Roberts, G.O., Gelman, A. and Gilks, W.R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254
[51] Roberts, G.O. and Rosenthal, J.S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. MR1448322 https://doi.org/10.1214/ECP.v2-981
[52] Roberts, G.O. and Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123
[53] Roberts, G.O. and Rosenthal, J.S. (2011). Quantitative non-geometric convergence bounds for independence samplers. *Methodol. Comput. Appl. Probab.* **13** 391–403. MR2788864 https://doi.org/10.1007/s11009-009-9157-z
[54] Roberts, G.O. and Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158 https://doi.org/10.1093/biomet/83.1.95
[55] 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
[56] Rosenthal, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509
[57] Rudolf, D. (2012). Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math.* **485** 1–93. MR2977521 https://doi.org/10.4064/dm485-0-1
[58] Schmidler, S.C. and Woodard, D.B. (2011). Lower bounds on the convergence rates of adaptive MCMC methods. Tech. Rep., Duke Univ.

[59] Shun, Z. and McCullagh, P. (1995). Laplace approximation of high-dimensional integrals. *J. Roy. Statist. Soc. Ser. B* **57** 749–760. MR1354079

[60] Tang, Y. and Reid, N. (2021). Laplace and saddlepoint approximations in high dimensions. Preprint. Available at arXiv:2107.10885.

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

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

[63] Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86. MR0830567

[64] Tierney, L., Kass, R.E. and Kadane, J.B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76** 425–433. MR1040637 https://doi.org/10.1093/biomet/76.3.425

[65] 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

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

[67] Vats, D., Gonçalves, F.B., Łatuszyński, K. and Roberts, G.O. (2022). Efficient Bernoulli factory Markov chain Monte Carlo for intractable posteriors. *Biometrika* **109** 369–385. MR4430963 https://doi.org/10.1093/biomet/asab031

[68] Villani, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Providence, RI: Amer. Math. Soc. MR1964483 https://doi.org/10.1090/gsm/058

[69] Villani, C. (2009). *Optimal Transport: Old and New*. Berlin: Springer.

[70] Wang, G. (2022). Exact convergence analysis of the independent Metropolis-Hastings algorithms. *Bernoulli* **28** 2012–2033. MR4411520 https://doi.org/10.3150/21-bej1409

[71] Wang, X. and Roy, V. (2018). Geometric ergodicity of Pólya-gamma Gibbs sampler for Bayesian logistic regression with a flat prior. *Electron. J. Stat.* **12** 3295–3311. MR3861283 https://doi.org/10.1214/18-EJS1481

[72] Woodard, D.B. (2015). A lower bound on the mixing time of uniformly ergodic Markov chains in terms of the spectral radius. Preprint. Available at arXiv:1405.0028.

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

[74] Wu, K., Schmidler, S. and Chen, Y. (2022). Minimax mixing time of the Metropolis-adjusted Langevin algorithm for log-concave sampling. *J. Mach. Learn. Res.* **23** Paper No. [270]. MR4577709

[75] Yang, J. and Rosenthal, J.S. (2023). Complexity results for MCMC derived from quantitative bounds. *Ann. Appl. Probab.* **33** 1259–1300. MR4564431 https://doi.org/10.1214/22-aap1846

[76] Yang, Y., Wainwright, M.J. and Jordan, M.I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *Ann. Statist.* **44** 2497–2532. MR3576552 https://doi.org/10.1214/15-AOS1417

[77] Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques*. *Stud. Bayesian Econometrics Statist.* **6** 233–243. Amsterdam: North-Holland. MR0881437

[78] Zhuo, B. and Gao, C. (2021). Mixing time of Metropolis-Hastings for Bayesian community detection. *J. Mach. Learn. Res.* **22** Paper No. 10. MR4253703

Bernoulli **31**(3), 2025, 1929–1954
https://doi.org/10.3150/24-BEJ1792

Convergence of empirical optimal transport in unbounded settings

THOMAS STAUDT[a] and SHAYAN HUNDRIESER[b]

Institute for Mathematical Stochastics, University of Göttingen, Göttingen, Germany,
[a]*thomas.staudt@uni-goettingen.de,* [b]*s.hundrieser@math.uni-goettingen.de*

In compact settings, the convergence rate of the empirical optimal transport cost to its population value is well understood for a wide class of spaces and cost functions. In unbounded settings, however, hitherto available results require strong assumptions on the ground costs and the concentration of the involved measures. In this work, we pursue a decomposition-based approach to generalize the convergence rates found in compact spaces to unbounded settings under generic moment assumptions that are sharp up to an arbitrarily small $\epsilon > 0$. Hallmark properties of empirical optimal transport on compact spaces, like the recently established adaptation to lower complexity, are shown to carry over to the unbounded case.

Keywords: Convergence rate; curse of dimensionality; metric entropy; Wasserstein distance

References

Ajtai, M., Komlós, J. and Tusnády, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264. MR0779885 https://doi.org/10.1007/BF02579135

Ambrosio, L., Brué, E. and Semola, D. (2021). *Lectures on Optimal Transport*. *Unitext* **130**. Cham: Springer. MR4294651 https://doi.org/10.1007/978-3-030-72162-6

Ambrosio, L., Goldman, M. and Trevisan, D. (2022). On the quadratic random matching problem in two-dimensional domains. *Electron. J. Probab.* **27** Paper No. 54, 35. MR4416128 https://doi.org/10.1214/22-ejp784

Barthe, F. and Bordenave, C. (2013). Combinatorial optimization over two random point sets. In *Séminaire de Probabilités XLV*. *Lecture Notes in Math.* **2078** 483–535. Cham: Springer. MR3185927 https://doi.org/10.1007/978-3-319-00321-4_19

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

Bobkov, S.G. and Ledoux, M. (2021). A simple Fourier analytic proof of the AKT optimal matching theorem. *Ann. Appl. Probab.* **31** 2567–2584. MR4350968 https://doi.org/10.1214/20-aap1656

Bogachev, V.I. (2007). *Measure Theory. Vol. I, II*. Berlin: Springer. MR2267655 https://doi.org/10.1007/978-3-540-34514-5

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

Chao, M.-T. and Strawderman, W. (1972). Negative moments of positive random variables. *J. Amer. Statist. Assoc.* **67** 429–431. https://doi.org/10.2307/2284399

Chizat, L., Roussillon, P., Léger, F., Vialard, F.-X. and Peyré, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. In *Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS'20* Curran Associates Inc.

Deb, N., Ghosal, P. and Sen, B. (2021). Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. In *Proceedings of the 35th International Conference on Neural Information Processing Systems. NIPS'21* Curran Associates Inc.

Dedecker, J. and Merlevède, F. (2019). Behavior of the empirical Wasserstein distance in $\mathbb{R}^d$ under moment conditions. *Electron. J. Probab.* **24** Paper No. 6, 32. MR3916326 https://doi.org/10.1214/19-EJP266 del Barrio, E., González Sanz, A. and Loubes, J.-M. (2024). Central limit theorems for semi-discrete Wasserstein distances. *Bernoulli* **30** 554–580. MR4665589 https://doi.org/10.3150/23-bej1608

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

Divol, V. (2021). A short proof on the rate of convergence of the empirical measure for the Wasserstein distance. Preprint. Available at arXiv:2101.08126. https://doi.org/10.48550/arXiv.2101.08126

Dobrić, V. and Yukich, J.E. (1995). Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8** 97–118. MR1308672 https://doi.org/10.1007/BF02213456

Dudley, R.M. (1969). The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Stat.* **40** 40–50. MR0236977 https://doi.org/10.1214/aoms/1177697802

Forrow, A., Hütter, J.-C., Nitzan, M., Rigollet, P., Schiebinger, G. and Weed, J. (2019). Statistical optimal transport via factored couplings. In *Proceedings of the 22nd International Conference on Artificial*. *Intelligence and Statistics* **89** 2454–2465.

Fournier, N. (2023). Convergence of the empirical measure in expected Wasserstein distance: Non-asymptotic explicit bounds in $\mathbb{R}^d$. *ESAIM Probab. Stat.* **27** 749–775. MR4624322 https://doi.org/10.1051/ps/2023011

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

Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton, NJ: Princeton Univ. Press. MR3586373 https://doi.org/10.1515/9781400883592

Goldman, M. and Trevisan, D. (2021). Convergence of asymptotic costs for random Euclidean matching problems. *Probab. Math. Phys.* **2** 341–362. MR4408015 https://doi.org/10.2140/pmp.2021.2.341

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. and Courville, A. (2017). Improved training of Wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17* 5769–5779. Hook, NY, USA: Curran Associates Inc., Red.

Hundrieser, S., Staudt, T. and Munk, A. (2024). Empirical optimal transport between different measures adapts to lower complexity. *Ann. Inst. Henri Poincaré Probab. Stat.* **60** 824–846. MR4757509 https://doi.org/10.1214/23-aihp1369

Hundrieser, S., Klatt, M., Munk, A. and Staudt, T. (2024). A unifying approach to distributional limits for empirical optimal transport. *Bernoulli* **30** 2846–2877. MR4779851 https://doi.org/10.3150/23-bej1697

Kantorovitch, L. (1942). On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37** 199–201. MR0009619

Kloeckner, B.R. (2020). Empirical measures: Regularity is a counter-curse to dimensionality. *ESAIM Probab. Stat.* **24** 408–434. MR4153634 https://doi.org/10.1051/ps/2019025

Kolkin, N., Salavon, J. and Shakhnarovich, G. (2019). Style transfer by relaxed optimal transport and self-similarity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 10051–10060.

Kolmogorov, A.N. and Tihomirov, V.M. (1961). ε-entropy and ε-capacity of sets in functional space. *Amer. Math. Soc. Transl. Ser. 2* **17** 277–364. MR0124720

Larsson, M., Park, J. and Wiesel, J. (2024). On concentration of the empirical measure for radial transport costs. *Stochastic Process. Appl.* **178** Paper No. 104466. MR4792529 https://doi.org/10.1016/j.spa.2024.104466

Ledoux, M. (2018). On optimal matching of Gaussian samples II. Preprint.

Ledoux, M. (2019). On optimal matching of Gaussian samples. *J. Math. Sci.* **238** 495–522. https://doi.org/10.1007/s10958-019-04253-6

Ledoux, M. and Zhu, J.-X. (2021). On optimal matching of Gaussian samples III. *Probab. Math. Statist.* **41** 237–265. MR4359822 https://doi.org/10.37190/0208-4147.41.2.3

Lei, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798. MR4036051 https://doi.org/10.3150/19-BEJ1151

Manole, T. and Niles-Weed, J. (2024). Sharp convergence rates for empirical optimal transport with smooth costs. *Ann. Appl. Probab.* **34** 1108–1135. MR4700254 https://doi.org/10.1214/23-aap1986

Manole, T., Balakrishnan, S., Niles-Weed, J. and Wasserman, L. (2024). Plugin estimation of smooth optimal transport maps. *Ann. Statist.* **52** 966–998. MR4784066 https://doi.org/10.1214/24-aos2379

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. Histoire de L'Académie Royale des Sciences de Paris 666–704.

Niles-Weed, J. and Rigollet, P. (2022). Estimation of Wasserstein distances in the spiked transport model. *Bernoulli* **28** 2663–2688. MR4474558 https://doi.org/10.3150/21-bej1433

Nolan, J.P. (2020). *Univariate Stable Distributions: Models for Heavy Tailed Data*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR4230105 https://doi.org/10.1007/978-3-030-52915-4

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

Peyré, G. and Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **11** 355–607. https://doi.org/10.1561/2200000073

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Cham: Birkhäuser/Springer. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

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.e22. https://doi.org/10.1016/j.cell.2019.01.006

Singh, S. and Póczos, B. (2018). Minimax distribution estimation in Wasserstein distance. Preprint. Available at arXiv:1802.08855. https://doi.org/10.48550/arXiv:1802.08855.

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

Staudt, T. and Hundrieser, S. (2025). Supplement to "Convergence of empirical optimal transport in unbounded settings." https://doi.org/10.3150/24-BEJ1792SUPP

Staudt, T., Hundrieser, S. and Munk, A. (2022). On the uniqueness of Kantorovich potentials. Preprint. Available at arXiv:2201.08316. https://doi.org/2201.08316.10.48550/arXiv.

Talagrand, M. (1992). The Ajtai-Komlós-Tusnády matching theorem for general measures. In *Probability in Banach Spaces, 8 (Brunswick, ME, 1991)*. *Progress in Probability* **30** 39–54. Boston, MA: Birkhäuser. MR1227608

Talagrand, M. (1994a). Matching theorems and empirical discrepancy computations using majorizing measures. *J. Amer. Math. Soc.* **7** 455–537. MR1227476 https://doi.org/10.2307/2152764

Talagrand, M. (1994b). The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.* **22** 919–959. MR1288137

Tameling, C., Stoldt, S., Stephan, T., Naas, J., Jakobs, S. and Munk, A. (2021). Colocalization for super-resolution microscopy via optimal transport. *Nat. Comput. Sci.* **1** 199–211. https://doi.org/10.1038/s43588-021-00050-x

Villani, C. (2008). *Optimal Transport: Old and New*. *A Series of Comprehensive Studies in Mathematics*. Berlin: Springer. https://doi.org/10.1007/978-3-540-71050-9

von Bahr, B. and Esseen, C.-G. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Stat.* **36** 299–303. MR0170407 https://doi.org/10.1214/aoms/1177700291

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

Yukich, J. (1992). Some generalizations of the Euclidean two-sample matching problem. In *Probability in Banach Spaces, 8 (Brunswick, ME, 1991)*. *Progress in Probability* **30** 55–66. Boston, MA: Birkhäuser. MR1227609

Bernoulli **31**(3), 2025, 1955–1990
https://doi.org/10.3150/24-BEJ1793

Topological signatures of periodic-like signals

FRÉDÉRIC CHAZAL[1,a], BERTRAND MICHEL[2,c] and WOJCIECH REISE[1,b]

[1]*DataShape, Inria Saclay, Institut Mathématique d'Orsay, 307 Rue Michel Magat, 91400 Orsay, France,* [a]*frederic.chazal@inria.fr,* [b]*reisewojciech@gmail.com*
[2]*Nantes Université, Centrale Nantes, Laboratoire de Mathématiques Jean Leray, CNRS UMR 6629, 1 Rue de La Noe, 44321 Nantes, France,* [c]*bertrand.michel@ec-nantes.fr*

We present a method to construct signatures of periodic-like data. Based on topological considerations, our construction encodes information about the order and values of local extrema. Its main strength is robustness to reparametrization of the observed signal, so that it depends only on the form of the periodic function. The signature converges as the observation contains increasingly many periods. We show that it can be estimated from the observation of a single time series using bootstrap techniques.

Keywords: Dependent data; functional data; limit theorems; persistent homology; time series

References

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F. and Ziegelmeier, L. (2017). Persistence images: A stable vector representation of persistent homology. *J. Mach. Learn. Res.* **18** Paper No. 8. MR3625712

Azaïs, J.-M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. Hoboken, NJ: Wiley. MR2478201 https://doi.org/10.1002/9780470434642

Berry, E., Chen, Y.-C., Cisewski-Kehe, J. and Fasy, B.T. (2020). Functional summaries of persistence diagrams. *J. Appl. Comput. Topol.* **4** 211–262. MR4096338 https://doi.org/10.1007/s41468-020-00048-w

Bogachev, V.I. (2007). *Measure Theory. Vol. I, II*. Berlin: Springer. MR2267655 https://doi.org/10.1007/978-3-540-34514-5

Bois, A., Tervil, B., Moreau, A., Vienne-Jumeau, A., Ricard, D. and Oudre, L. (2022). A topological data analysis-based method for gait signals with an application to the study of multiple sclerosis. *PLoS ONE* **17** 1–23. https://doi.org/10.1371/journal.pone.0268475

Bonis, T., Chazal, F., Michel, B. and Reise, W. (2024). Topological phase estimation method for reparameterized periodic functions. *Adv. Comput. Math.* **50** Paper No. 66. MR4770431 https://doi.org/10.1007/s10444-024-10157-0

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

Bühlmann, P. (1995). The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Process. Appl.* **58** 247–265. MR1348377 https://doi.org/10.1016/0304-4149(95)00019-4

Bühlmann, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074 https://doi.org/10.1214/ss/1023798998

Carrière, M., Chazal, F., Ike, Y., Lacombe, T., Royer, M. and Umeda, Y. (2020). PersLay: A neural network nayer for persistence diagrams and new graph topological signatures. In *AISTATS'2017* **108** 2786–2796. PMLR.

Chazal, F. and Michel, B. (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Front. Artif. Intell.* **4**. https://doi.org/10.3389/frai.2021.667963

Chazal, F., Michel, B. and Reise, W. (2025). Supplement to "Topological signatures of periodic-like signals." https://doi.org/10.3150/24-BEJ1793SUPP

Chazal, F., Fasy, B.T., Lecci, F., Rinaldo, A. and Wasserman, L. (2014). Stochastic convergence of persistence landscapes and silhouettes. In *Computational Geometry (SoCG'14)* 474–483. New York: ACM. MR3382329

Chazal, F., de Silva, V., Glisse, M. and Oudot, S. (2016). *The Structure and Stability of Persistence Modules*. *SpringerBriefs in Mathematics*. Cham: Springer. MR3524869 https://doi.org/10.1007/978-3-319-42545-0

Chung, Y.-M. and Lawson, A. (2022). Persistence curves: A canonical framework for summarizing persistence diagrams. *Adv. Comput. Math.* **48** Paper No. 6. MR4368950 https://doi.org/10.1007/s10444-021-09893-4

Cohen-Steiner, D., Edelsbrunner, H. and Harer, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.* **37** 103–120. MR2279866 https://doi.org/10.1007/s00454-006-1276-5

Cohen-Steiner, D., Edelsbrunner, H., Harer, J. and Mileyko, Y. (2010). Lipschitz functions have L_p-stable persistence. *Found. Comput. Math.* **10** 127–139. MR2594441 https://doi.org/10.1007/s10208-010-9060-6

Corcoran, P. and Jones, C. (2017). Modelling topological features of swarm behaviour in space and time with persistence landscapes. *IEEE Access* **5** 18534–18544. https://doi.org/10.1109/ACCESS.2017.2749319

Dedecker, J., Doukhan, P., Lang, G., José Rafael., L.R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*. *Lecture Notes in Statistics* **190**. New York: Springer. MR2338725

Divol, V. and Polonik, W. (2019). On the choice of weight functions for linear representations of persistence diagrams. *J. Appl. Comput. Topol.* **3** 249–283. MR3996956 https://doi.org/10.1007/s41468-019-00032-z

Doukhan, P. (1994). *Mixing: Properties and Examples*. *Lecture Notes in Statistics* **85**. New York: Springer. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0

Fernández, X. and Mateos, D. (2022). Topological biomarkers for real-time detection of epileptic seizures. ArXiv preprint. Available at arXiv:2211.02523. https://doi.org/10.1214/aos/1069362747

Ghil, M. and Sciamarella, D. (2023). Review article: Dynamical systems, algebraic topology, and the climate sciences. *EGUsphere*. Preprint. https://doi.org/10.5194/egusphere-2023-216

Gidea, M. and Katz, Y. (2018). Topological data analysis of financial time series: Landscapes of crashes. *Phys. A* **491** 820–834. MR3721543 https://doi.org/10.1016/j.physa.2017.09.028

Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K. and Stanley, H.E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **101** E215–220. https://doi.org/10.1161/01.cir.101.23.e215

Hu, Y. and Le, K. (2013). A multiparameter Garsia-Rodemich-Rumsey inequality and some applications. *Stochastic Process. Appl.* **123** 3359–3377. MR3071383 https://doi.org/10.1016/j.spa.2013.04.019

Khasawneh, F.A. and Munch, E. (2016). Chatter detection in turning using persistent homology. *Mech. Syst. Signal Process.* **70–71** 527–541. https://doi.org/10.1016/j.ymssp.2015.09.046

Khorram, S., McInnis, M.G. and Provost, E.M. (2019). Trainable time warping: Aligning time-series in the continuous-time domain. In *ICASSP'2019* 3502–3506. https://doi.org/10.1109/ICASSP.2019.8682322

Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. New York: Springer. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

Krebs, J. (2021). On limit theorems for persistent Betti numbers from dependent data. *Stochastic Process. Appl.* **139** 139–174. MR4260054 https://doi.org/10.1016/j.spa.2021.04.013

Marron, J.S., Ramsay, J.O., Sangalli, L.M. and Srivastava, A. (2015). Functional data analysis of amplitude and phase variation. *Statist. Sci.* **30** 468–484. MR3432837 https://doi.org/10.1214/15-STS524

Merlevède, F. and Peligrad, M. (2020). Functional CLT for nonstationary strongly mixing processes. *Statist. Probab. Lett.* **156** 108581. MR4003491 https://doi.org/10.1016/j.spl.2019.108581

Meyn, S.P. and Tweedie, R.L. (2012). *Markov Chains and Stochastic Stability*. Springer Science & Business Media.

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

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

Perez, D. (2022). On C^0-persistent homology and trees. ArXiv preprint. Available at arXiv:2012.02634v3.

Perng, C.S., Wang, H., Zhang, S.R. and Parker, D.S. (2000). Landmarks: A new model for similarity-based pattern querying in time series databases. In *ICDE'00* 33–42. San Diego, CA, USA: IEEE Comput. Soc. https://doi.org/10.1109/ICDE.2000.839385

Pettis, B.J. (1938). On integration in vector spaces. *Trans. Amer. Math. Soc.* **44** 277–304. MR1501970 https://doi.org/10.2307/1989973

Plonka, G. and Zheng, Y. (2016). Relation between total variation and persistence distance and its application in signal processing. *Adv. Comput. Math.* **42** 651–674. MR3502107 https://doi.org/10.1007/s10444-015-9438-8

Radulović, D. (1996). The bootstrap for empirical processes based on stationary observations. *Stochastic Process. Appl.* **65** 259–279. MR1425360 https://doi.org/10.1016/S0304-4149(96)00102-0

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

Reise, W. (2023). Topological techniques for inference on periodic functions with phase variation Ph.D. thesis Université Paris-Saclay.

Rio, E. (1997). About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Stat.* **1** 35–61. MR1382517 https://doi.org/10.1051/ps:1997102

Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes*. *Probability Theory and Stochastic Modelling* **80**. Berlin: Springer. MR3642873 https://doi.org/10.1007/978-3-662-54323-8

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

Shevchenko, G. (2017). Kolmogorov continuity theorem and Holder norm. MathOverflow. Available at https://mathoverflow.net/q/279085 (version: 2017-08-19).

Su, J., Kurtek, S., Klassen, E. and Srivastava, A. (2014). Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *Ann. Appl. Stat.* **8** 530–552. MR3192001 https://doi.org/10.1214/13-AOAS701

Tang, R. and Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889. MR2461217 https://doi.org/10.1093/biomet/asn047

Utev, S.A. (1990). Central limit theorem for dependent random variables. In *Probability Theory and Mathematical Statistics, Vol. II (Vilnius, 1989)* 519–528. Vilnius: "Mokslas". MR1153906

Wang, K. and Gasser, T. (1997). Alignment of curves by dynamic time warping. *Ann. Statist.* **25** 1251–1276. MR1447750 https://doi.org/10.1214/aos/1069362747

Bernoulli **31**(3), 2025, 1991–2017
https://doi.org/10.3150/24-BEJ1794

LAMN property for stable-Lévy SDEs with constant scale coefficient

ALEXANDRE BROUSTE[a], LAURENT DENIS[b] and THI-BAO-TRÂM NGÔ[c]

Le Mans Université, Laboratoire Manceau de Mathématiques, Avenue Olivier Messiaen, 72085 Le Mans Cedex 09, France, [a]*Alexandre.Brouste@univ-lemans.fr,* [b]*Laurent.Denis@univ-lemans.fr,* [c]*Thi_Bao_Tram.Ngo@univ-lemans.fr*

The joint parametric estimation of the drift coefficient, the scale coefficient and the jump activity in stochastic differential equations driven by a symmetric stable Lévy process is considered, based on high-frequency observations. Firstly, the LAMN property for the corresponding Euler-type scheme is proven and lower bounds for the estimation risk in this setting are deduced. When the approximation scheme experiment is asymptotically equivalent to the original one, these bounds can be transferred. Secondly, a one-step procedure is proposed which is shown to be fast and asymptotically normal and even asymptotically efficient when the scale coefficient is constant. The performances in terms of asymptotical variance and computation time on samples of finite size are illustrated with simulations.

Keywords: LAMN property; Lévy process; one-step procedure; parametric estimation; stable process; stochastic differential equation

References

Bayraktar, E. and Clément, E. (2024). Estimation of a pure-jump stable Cox-Ingersoll-Ross process. *Bernoulli*. in press.

Brouste, A., Denis, L. and Ngô, T.-B.-T. (2025). Supplement to “LAMN property for stable-Lévy SDEs with constant scale coefficient.” https://doi.org/10.3150/24-BEJ1794SUPP

Brouste, A. and Fukasawa, M. (2018). Local asymptotic normality property for fractional Gaussian noise under high-frequency observations. *Ann. Statist.* **46** 2045–2061. MR3845010 https://doi.org/10.1214/17-AOS1611

Brouste, A. and Masuda, H. (2018). Efficient estimation of stable Lévy process with symmetric jumps. *Stat. Inference Stoch. Process.* **21** 289–307. MR3824969 https://doi.org/10.1007/s11203-018-9181-0

Clément, E. (2023). Hellinger and total variation distance in approximating Lévy driven SDEs. *Ann. Appl. Probab.* **33** 2176–2209. MR4583668 https://doi.org/10.1214/22-aap1863

Clément, E. and Gloter, A. (2015). Local asymptotic mixed normality property for discretely observed stochastic differential equations driven by stable Lévy processes. *Stochastic Process. Appl.* **125** 2316–2352. MR3322866 https://doi.org/10.1016/j.spa.2015.01.002

Clément, E. and Gloter, A. (2019). Estimating functions for SDE driven by stable Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 1316–1348. MR4010937 https://doi.org/10.1214/18-aihp920

Clément, E. and Gloter, A. (2020). Joint estimation for SDE driven by locally stable Lévy processes. *Electron. J. Stat.* **14** 2922–2956. MR4134348 https://doi.org/10.1214/20-EJS1737

Clément, E., Gloter, A. and Nguyen, H. (2019). LAMN property for the drift and volatility parameters of a SDE driven by a stable Lévy process. *ESAIM Probab. Stat.* **23** 136–175. MR3945580 https://doi.org/10.1051/ps/2018007

Genon-Catelot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **29** 119–151. MR1204521

Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: A Malliavin calculus approach. *Bernoulli* **7** 899–912. MR1873834 https://doi.org/10.2307/3318625

1350-7265

Höpfner, R. (2014). *Asymptotic Statistics: With a View to Stochastic Processes*. *De Gruyter Graduate*. Berlin: de Gruyter. MR3185373 https://doi.org/10.1515/9783110250282

Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A* **44** 173–212. MR0688800

Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypotheses. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 129–156. Berkeley-Los Angeles, Calif.: Univ. California Press. MR0084918

Le Cam, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Stat.* **35** 1419–1455. MR0207093 https://doi.org/10.1214/aoms/1177700372

Masuda, H. (2015). Parametric estimation of Lévy processes. In *Lévy Matters. IV*. *Lecture Notes in Math.* **2128** 179–286. Cham: Springer. MR3364258 https://doi.org/10.1007/978-3-319-12373-8_3

Masuda, H. (2023). Optimal stable Ornstein-Uhlenbeck regression. *Jpn. J. Stat. Data Sci.* **6** 573–605. MR4594508 https://doi.org/10.1007/s42081-023-00197-z

Matsui, M. and Takemura, A. (2006). Some improvements in numerical evaluation of symmetric stable density and its derivatives. *Comm. Statist. Theory Methods* **35** 149–172. MR2274041 https://doi.org/10.1080/03610920500439729

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. Translated from the 1990 Japanese original, Revised by the author. MR1739520

Shiryaev, A.N. and Spokoiny, V.G. (2000). *Statistical Experiments and Decisions: Asymptotic Theory*. *Advanced Series on Statistical Science & Applied Probability* **8**. River Edge, NJ: World Scientific Co., Inc. MR1791434 https://doi.org/10.1142/9789812779243

Strasser, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. *De Gruyter Studies in Mathematics* **7**. Berlin: de Gruyter. MR0812467 https://doi.org/10.1515/9783110850826

Sweeting, T.J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8** 1375–1381. MR0594652

Todorov, V. (2013). Power variation from second order differences for pure jump semimartingales. *Stochastic Process. Appl.* **123** 2829–2850. MR3054547 https://doi.org/10.1016/j.spa.2013.04.005

Bernoulli **31**(3), 2025, 2018–2041
https://doi.org/10.3150/24-BEJ1795

Edgeworth expansion by Stein's method

XIAO FANG[1,a] and SONG-HAO LIU[2,b]

[1]*Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China,* [a]*xfang@sta.cuhk.edu.hk*
[2]*Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, China,* [b]*liusonghao_cuhk@link.cuhk.edu.hk*

Edgeworth expansion provides higher-order corrections to the normal approximation for a probability distribution. The classical proof of Edgeworth expansion is via characteristic functions. As a powerful method for distributional approximations, Stein's method has also been used to prove Edgeworth expansion results. However, these results assume that either the test function is smooth (which excludes indicator functions of the half line) or that the random variables are continuous (which excludes random variables having only a continuous component). Thus, how to recover the classical Edgeworth expansion result using Stein's method has remained an open problem. In this paper, we develop Stein's method for two-term Edgeworth expansions in a general case. Our approach involves repeated use of Stein equations, Stein identities via Stein kernels, and a replacement argument.

Keywords: Central limit theorem; Edgeworth expansion; Stein kernel; Stein's method

References

Angst, J. and Poly, G. (2017). A weak Cramér condition and application to Edgeworth expansions. *Electron. J. Probab.* **22** Paper No. 59, 24. MR3683368 https://doi.org/10.1214/17-EJP77

Bally, V., Caramellino, L. and Poly, G. (2018). Convergence in distribution norms in the CLT for non identical distributed random variables. *Electron. J. Probab.* **23** Paper No. 45, 51. MR3814239 https://doi.org/10.1214/18-ejp174

Barbour, A.D. (1986). Asymptotic expansions based on smooth functions in the central limit theorem. *Probab. Theory Related Fields* **72** 289–303. MR0836279 https://doi.org/10.1007/BF00699108

Berry, A.C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. MR0003498 https://doi.org/10.2307/1990053

Bhattacharya, R.N. and Rao, R.R. (1976). *Normal Approximation and Asymptotic Expansions*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0436272

Biermé, H., Bonami, A., Nourdin, I. and Peccati, G. (2012). Optimal Berry-Esseen rates on the Wiener space: The barrier of third and fourth cumulants. *ALEA Lat. Am. J. Probab. Math. Stat.* **9** 473–500. MR3069374

Braverman, A., Dai, J.G. and Fang, X. (2024). High-order steady-state diffusion approximations. *Oper. Res.* **72** 604–616. MR4736779

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

Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. *Probability and Its Applications (New York)*. Heidelberg: Springer. MR2732624 https://doi.org/10.1007/978-3-642-15007-4

Courtade, T.A., Fathi, M. and Pananjady, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 777–790. MR3949953 https://doi.org/10.1214/18-aihp898

Esseen, C.-G. (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astron. Fys.* **28A** 19. MR0011909

Fang, X. (2019). Wasserstein-2 bounds in normal approximation under local dependence. *Electron. J. Probab.* **24** Paper No. 35, 14. MR3940765 https://doi.org/10.1214/19-EJP301

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

Fathi, M. (2019). Stein kernels and moment maps. *Ann. Probab.* **47** 2172–2185. MR3980918 https://doi.org/10.1214/18-AOP1305

Fathi, M. (2021). Higher-order Stein kernels for Gaussian approximation. *Studia Math.* **256** 241–258. MR4172696 https://doi.org/10.4064/sm190415-28-10

Kim, Y.T. and Park, H.S. (2018). An Edgeworth expansion for functionals of Gaussian fields and its applications. *Stochastic Process. Appl.* **128** 3967–3999. MR3906975 https://doi.org/10.1016/j.spa.2018.01.006

Ledoux, M., Nourdin, I. and Peccati, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25** 256–306. MR3320893 https://doi.org/10.1007/s00039-015-0312-0

Nourdin, I. and Peccati, G. (2009a). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122 https://doi.org/10.1007/s00440-008-0162-x

Nourdin, I. and Peccati, G. (2009b). Noncentral convergence of multiple integrals. *Ann. Probab.* **37** 1412–1426. MR2546749 https://doi.org/10.1214/08-AOP435

Petrov, V.V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete [Results in Mathematics and Related Areas], Band 82*. New York: Springer. Translated from the Russian by A. A. Brown. MR0388499

Privault, N. (2015). Stein approximation for Itô and Skorohod integrals by Edgeworth type expansions. *Electron. Commun. Probab.* **20** No. 35, 10. MR3342169 https://doi.org/10.1214/ECP.v20-3827

Prokhorov, Y.V. (1952). A local theorem for densities. *Dokl. Akad. Nauk SSSR* **83** 797–800. MR0049501

Rinott, Y. and Rotar, V. (2003). On Edgeworth expansions for dependency-neighborhoods chain structures and Stein's method. *Probab. Theory Related Fields* **126** 528–570. MR2001197 https://doi.org/10.1007/s00440-003-0271-5

Saumard, A. (2019). Weighted Poincaré inequalities, concentration inequalities and tail bounds related to Stein kernels in dimension one. *Bernoulli* **25** 3978–4006. MR4010979 https://doi.org/10.3150/19-bej1117

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. Berkeley, CA: Univ. California Press. MR0402873

Stein, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. Hayward, CA: IMS. MR0882007

Bernoulli **31**(3), 2025, 2042–2076
https://doi.org/10.3150/24-BEJ1796

Explicit constraints on the geometric rate of convergence of random walk Metropolis-Hastings

RIDDHIMAN BHATTACHARYA[1,a] and GALIN JONES[2,b]

[1] *Duke University, Durham, United States of America,* [a] *riddhiman.bhattacharya@duke.edu*
[2] *School of Statistics, University of Minnesota, Minneapolis, United States of America,* [b] *galin@umn.edu*

Convergence rate analyses of random walk Metropolis-Hastings Markov chains on general state spaces have largely focused on establishing sufficient conditions for geometric ergodicity or on analysis of mixing times. Geometric ergodicity is a key sufficient condition for the Markov chain Central Limit Theorem and allows rigorous approaches to assessing Monte Carlo error. The sufficient conditions for geometric ergodicity of the random walk Metropolis-Hastings Markov chain are refined and extended, which allows the analysis of previously inaccessible settings such as Bayesian Poisson regression. The key technical innovation is the development of explicit drift and minorization conditions for random walk Metropolis-Hastings, which allows explicit upper and lower bounds on the geometric rate of convergence. Alternative lower bounds on the geometric rate of convergence are developed using spectral theory. The previous sufficient conditions for geometric ergodicity have not provided explicit constraints on the rate of geometric rate of convergence because the method used only implies the existence of drift and minorization conditions. The theoretical results are applied to random walk Metropolis-Hastings algorithms for a class of exponential families and generalized linear models that address Bayesian regression problems.

Keywords: Geometric ergodicity; Markov chain Monte Carlo; Metropolis-Hastings; random walk

References

[1] Andrieu, C., Lee, A., Power, S. and Wang, A.Q. (2022). Explicit convergence bounds for Metropolis Markov chains: Isoperimetry, spectral gaps and profiles. arXiv preprint arXiv:2211.08959.
[2] Barndorff-Nielsen, O. (2014). *Information and Exponential Families: In Statistical Theory*. New York: Wiley.
[3] Baxendale, P.H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738.
[4] Belloni, A. and Chernozhukov, V. (2009). On the computational complexity of MCMC-based estimators in large samples. *Ann. Statist.* **37** 2011–2055.
[5] Bhattacharya, R. and Jones, G. (2025). Supplement to "Explicit constraints on the geometric rate of convergence of random walk Metropolis-Hastings." https://doi.org/10.3150/24-BEJ1796SUPP
[6] Brofos, J.A., Roy, V. and Lederman, R.R. (2023). Geometric ergodicity in modified variations of Riemannian manifold and Lagrangian Monte Carlo. arXiv:2301.01409.
[7] Brooks, S., Gelman, A., Jones, G.L. and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: CRC Press.
[8] Brown, A. and Jones, G.L. (2024). Lower bounds for the convergence rate of accept-reject-based Markov chains. *Bernoulli*, to appear.
[9] Chan, K.S. and Geyer, C.J. (1994). Discussion: Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1747–1758.
[10] Christensen, O.F., Moller, J. and Waagepetersen, R.P. (2001). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models. *Methodol. Comput. Appl. Probab.* **3** 309–327.

[11] Dwivedi, R., Chen, Y., Wainwright, M.J. and Yu, B. (2018). Log-concave sampling: Metropolis-Hastings algorithms are fast! In *Conference on Learning Theory* 793–797. PMLR.
[12] Flegal, J.M., Hughes, J., Vats, D. and Dai, N. (2020). mcmcse: Monte Carlo standard errors for MCMC. R package version 1.4-1.
[13] Gelman, A., Gilks, W. and Roberts, G. (1996). Efficient Metropolis jumping rules. In *Bayesian Statistics 5* (J.M. Bernardo, J.O. Berger, A.P. Dawid, D.V. Lindley and A.F.M. Smith, eds.). Oxford: Oxford Univ. Press.
[14] Geyer, C.J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.* **7** 473–511.
[15] Geyer, C.J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **56** 261–274.
[16] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.
[17] Jarner, S.F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361.
[18] Jerison, D.C. (2019). Quantitative convergence rates for reversible Markov chains via strong random times. arXiv:1908.06459.
[19] Johndrow, J. and Smith, A. (2018). Fast mixing of Metropolis-Hastings with unimodal targets. *Electron. Commun. Probab.* **23** 1–9.
[20] Johnson, A.A., Jones, G.L. and Neath, R.C. (2013). Component-wise Markov chain Monte Carlo: Uniform and geometric ergodicity under mixing and composition. *Statist. Sci.* **28** 360–375.
[21] Johnson, L.T. and Geyer, C.J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076.
[22] Jones, G.L. (2004). On the Markov chain central limit theorem. *Probab. Surv.* **1** 299–320.
[23] Jones, G.L., Haran, M., Caffo, B.S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547.
[24] Jones, G.L. and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334.
[25] Jones, G.L. and Hobert, J.P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Ann. Statist.* **32** 784–817.
[26] Jones, G.L., Roberts, G.O. and Rosenthal, J.S. (2014). Convergence of conditional Metropolis-Hastings samplers. *Adv. in Appl. Probab.* **46** 422–445.
[27] Lawler, G.F. and Sokal, A.D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309** 557–580.
[28] Livingstone, S., Betancourt, M., Byrne, S. and Girolami, M. (2019). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli* **25** 3109–3138.
[29] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121.
[30] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machine. *J. Chem. Phys.* **21** 1087–1092.
[31] Meyn, S.P. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability* **2**. Cambridge: Cambridge Univ. Press.
[32] Roberts, G.O. and Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110.
[33] Roberts, G.O. and Tweedie, R.L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* **80** 211–229.
[34] Robertson, N., Flegal, J.M., Vats, D. and Jones, G.L. (2021). Assessing and visualizing simultaneous simulation error. *J. Comput. Graph. Statist.* **30** 324–334.
[35] Rosenthal, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566.
[36] Rosenthal, J.S. (1996). Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Stat. Comput.* **6** 269–275.
[37] Roy, V. and Zhang, L. (2021). Convergence of position-dependent MALA with application to conditional simulation in GLMMs. arXiv:2108.12662.

[38] Rudolf, D. and Ullrich, M. (2013). Positivity of hit-and-run and related algoprithms. *Electron. Commun. Probab.* **18** 1–8.

[39] Vats, D., Flegal, J.M. and Jones, G.L. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli* **24** 1860–1909.

[40] Vats, D., Flegal, J.M. and Jones, G.L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337.

[41] Zhou, Q., Yang, J., Vats, D., Roberts, G.O. and Rosenthal, J.S. (2022). Dimension-free mixing for high-dimensional Bayesian variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1751–1784.

Bernoulli **31**(3), 2025, 2077–2104
https://doi.org/10.3150/24-BEJ1797

PDE characterization of geometric distribution functions and quantiles

DIMITRI KONEN[a]

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK,
[a]*dk738@cam.ac.uk*

We show that, in any Euclidean space, an arbitrary probability measure can be reconstructed explicitly by its geometric (or spatial) distribution function. The reconstruction takes the form of a (potentially fractional) linear PDE, where the differential operator is given in closed form. This result implies that, contrary to a common belief in the statistical depth community, geometric cdfs in principle provide exact control over the probability content of all depth regions. We present a comprehensive study of the regularity of the geometric distribution function, and show that, contrary to the univariate case, a continuous density in general does not give rise to a geometric cdf with enough regularity to reconstruct the density pointwise. Surprisingly, we prove that the reconstruction displays different behaviours in odd and even dimension: it is local in odd dimension and completely nonlocal in even dimension. We investigate this issue and provide a partial counterpart for even dimensions, and establish a general representation formula of the geometric cdf of spherically symmetric probability laws in odd dimensions. We finally provide explicit density reconstructions from their corresponding geometric distribution function in dimension 2 and 3.

Keywords: Centrality regions; deconvolution; fractional Laplacian; multivariate quantiles; spatial quantiles; statistical depth

References

Caffarelli, L. and Silvestre, L. (2007). An extension problem related to the fractional Laplacian. *Comm. Partial Differential Equations* **32** 1245–1260. MR2354493 https://doi.org/10.1080/03605300600987306

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

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

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

Di Nezza, E., Palatucci, G. and Valdinoci, E. (2012). Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.* **136** 521–573. MR2944369 https://doi.org/10.1016/j.bulsci.2011.12.004

Fernández-Real, X. and Ros-Oton, X. (2022). *Regularity Theory for Elliptic PDE. Zurich Lectures in Advanced Mathematics* **28**. Berlin: EMS Press. MR4560756 https://doi.org/10.4171/zlam/28

Gijbels, I. and Nagy, S. (2016). On smoothness of Tukey depth contours. *Statistics* **50** 1075–1085. MR3522389 https://doi.org/10.1080/02331888.2016.1145680

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

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

Grafakos, L. (2014). *Classical Fourier Analysis*, 3rd ed. *Graduate Texts in Mathematics* **249**. New York: Springer. MR3243734 https://doi.org/10.1007/978-1-4939-1194-3

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

1350-7265

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

Kemperman, J.H.B. (1987). The median of a finite measure on a Banach space. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods (Neuchâtel, 1987)* 217–230. Amsterdam: North-Holland. MR0949228

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

Konen, D. (2025). Supplement to "PDE characterization of geometric distribution functions and quantiles." https://doi.org/10.3150/24-BEJ1797SUPP

Konen, D. and Paindaveine, D. (2022). Multivariate ρ-quantiles: A spatial approach. *Bernoulli* **28** 1912–1934. MR4411516 https://doi.org/10.3150/21-bej1404

Kwaśnicki, M. (2019). Fractional Laplace operator and its properties. In *Handbook of Fractional Calculus with Applications. Vol. 1* 159–193. Berlin: de Gruyter. MR3888401

Lischke, A., Pang, G., Gulian, M. et al. (2020). What is the fractional Laplacian? A comparative review with new results. *J. Comput. Phys.* **404** 109009, 62 pp. MR4043885 https://doi.org/10.1016/j.jcp.2019.109009

Liu, R.Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400 https://doi.org/10.1214/aos/1176347507

Liu, R.Y., Parelius, J.M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. MR1724033 https://doi.org/10.1214/aos/1018031260

Möttönen, J., Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552. MR1439313 https://doi.org/10.1214/aos/1031833663

Nagy, S. (2017). Monotonicity properties of spatial depth. *Statist. Probab. Lett.* **129** 373–378. MR3688558 https://doi.org/10.1016/j.spl.2017.06.025

Nagy, S. (2020). The halfspace depth characterization problem. In *Nonparametric Statistics*. *Springer Proc. Math. Stat.* **339** 379–389. Cham: Springer. MR4178663 https://doi.org/10.1007/978-3-030-57306-5_34

Nagy, S. (2021). Halfspace depth does not characterize probability distributions. *Statist. Papers* **62** 1135–1139. MR4262188 https://doi.org/10.1007/s00362-019-01130-x

Nagy, S., Schütt, C. and Werner, E.M. (2019). Halfspace depth and floating body. *Stat. Surv.* **13** 52–118. MR3973130 https://doi.org/10.1214/19-ss123

Oberhettinger, F. (1972). *Tables of Bessel Transforms*. New York: Springer. MR0352888

Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. New York: Springer. MR2598854 https://doi.org/10.1007/978-1-4419-0468-3

Olver, F., Lozier, D., Boisvert, R. and Clark, C., eds. (2010). *NIST Handbook of Mathematical Functions*. National Institute of Standards and Technology, and Cambridge University Press.

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

Passeggeri, R. and Reid, N. (2022). A universal robustification procedure. Preprint. Available at arXiv:2206.06998.

Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. (1990). *Integrals and Series. Vol. 3: More Special Functions*. New York: Gordon and Breach Science Publishers. MR1054647

Romon, G. (2022). Statistical properties of approximate geometric quantiles in infinite-dimensional Banach spaces. Preprint. Available at arXiv:2211.00035.

Serfling, R. (2002). Quantile functions for multivariate analysis: Approaches and applications. *Stat. Neerl.* **56** 214–232. MR1916321 https://doi.org/10.1111/1467-9574.00195

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

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

Serfling, R. and Zuo, Y. (2010). Discussion of "Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth", by M. Hallin, D. Paindaveine, and M. Šiman. *Ann. Statist.* **38** 676–684.

Silvestre, L. (2007). Regularity of the obstacle problem for a fractional power of the Laplace operator. *Comm. Pure Appl. Math.* **60** 67–112. MR2270163 https://doi.org/10.1002/cpa.20153

Singha, S., Kratz, M. and Vadlamani, S. (2023). From geometric quantiles to halfspace depth: A geometric approach for extremal behaviour. ESSEC Working Paper 2307.

Stinga, P.R. (2019). User's guide to the fractional Laplacian and the method of semigroups. In *Handbook of Fractional Calculus with Applications. Vol. 2* 235–265. Berlin: de Gruyter. MR3965397

Struyf, A. and Rousseeuw, P.J. (1999). Halfspace depth and regression depth characterize the empirical distribution. *J. Multivariate Anal.* **69** 135–153. MR1701410 https://doi.org/10.1006/jmva.1998.1804

Tukey, J.W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974), Vol. 2* 523–531. Montreal, QC: Canad. Math. Congr. MR0426989

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

Zuo, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. MR2012822 https://doi.org/10.1214/aos/1065705115

Bernoulli **31**(3), 2025, 2105–2126
https://doi.org/10.3150/24-BEJ1798

Clustering a mixture of Gaussians with unknown covariance

DAMEK DAVIS[1,a], MATEO DÍAZ[2,b] and KAIZHENG WANG[3,c]

[1]*Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA, USA,* [a]*damek@wharton.upenn.edu*

[2]*Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA,* [b]*mateodd@jhu.edu*

[3]*Department of IEOR and Data Science Institute, Columbia University, New York, NY, USA,* [c]*kaizheng.wang@columbia.edu*

We investigate a clustering problem with data from a mixture of Gaussians that share a common but unknown, and potentially ill-conditioned, covariance matrix. We start by considering Gaussian mixtures with two equally-sized components and derive a Max-Cut integer program based on maximum likelihood estimation. We prove its solutions achieve the optimal misclassification rate when the number of samples grows linearly in the dimension, up to a logarithmic factor. However, solving the Max-Cut problem appears to be computationally intractable. To overcome this, we develop an efficient spectral algorithm that attains the optimal rate but requires a quadratic sample size. Although this sample complexity is worse than that of the Max-Cut problem, we conjecture that no polynomial-time method can perform better. Furthermore, we gather numerical and theoretical evidence that supports the existence of a statistical-computational gap. Finally, we generalize the Max-Cut program to a k-means program that handles multi-component mixtures with possibly unequal weights.

Keywords: Clustering; k-means; maximum cut; mixture models; statistical-computational tradeoff

References

[1] Abbe, E., Fan, J. and Wang, K. (2022). An ℓ_p theory of PCA and spectral clustering. *Ann. Statist.* **50** 2359–2385. MR4474494 https://doi.org/10.1214/22-aos2196

[2] Anandkumar, A., Ge, R., Hsu, D., Kakade, S.M. and Telgarsky, M. (2014). Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **15** 2773–2832. MR3270750

[3] Azizyan, M., Singh, A. and Wasserman, L. (2015). Efficient sparse clustering of high-dimensional non-spherical Gaussian mixtures. In *Artificial Intelligence and Statistics* 37–45.

[4] Bakshi, A., Diakonikolas, I., Jia, H., Kane, D.M., Kothari, P.K. and Vempala, S.S. (2022). Robustly learning mixtures of k arbitrary Gaussians. In *STOC '22—Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* 1234–1247. New York: ACM. MR4490075 https://doi.org/10.1145/3519935.3519953

[5] Bakshi, A. and Kothari, P. (2020). Outlier-robust clustering of non-spherical mixtures. Preprint. Available at arXiv:2005.02970. https://doi.org/10.48550/arXiv.2005.02970

[6] Balakrishnan, S., Wainwright, M.J. and Yu, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* **45** 77–120. MR3611487 https://doi.org/10.1214/16-AOS1435

[7] Banks, J., Moore, C., Vershynin, R., Verzelen, N. and Xu, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inf. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

[8] Barak, B., Kelner, J.A. and Steurer, D. (2014). Rounding sum-of-squares relaxations. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* 31–40. https://doi.org/10.1145/3583680

[9] Barak, B. and Steurer, D. (2014). Sum-of-squares proofs and the quest toward optimal algorithms. *Electron. Colloq. Comput. Complex.* **21** 59.

[10] Belkin, M. and Sinha, K. (2010). Toward learning Gaussian mixtures with arbitrary separation. 407–419.

[11] 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

[12] Blanchard, G., Kawanabe, M., Sugiyama, M., Spokoiny, V. and Müller, K.-R. (2006). In search of non-Gaussian components of a high-dimensional distribution. *J. Mach. Learn. Res.* **7** 247–282. MR2274368

[13] Brennan, M. and Bresler, G. (2019). Average-case lower bounds for learning sparse mixtures, robust estimation and semirandom adversaries. Preprint. Available at arXiv:1908.06130. https://doi.org/10.48550/arXiv.1908.06130

[14] Brennan, M. and Bresler, G. (2020). Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory* 648–847. PMLR.

[15] Brubaker, S.C. and Vempala, S.S. (2008). Isotropic PCA and affine-invariant clustering. In *Building Bridges*. *Bolyai Soc. Math. Stud.* **19** 241–281. Berlin: Springer. MR2484643 https://doi.org/10.1007/978-3-540-85221-6_8

[16] Cai, T.T., Ma, J. and Zhang, L. (2019). CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality. *Ann. Statist.* **47** 1234–1267. MR3911111 https://doi.org/10.1214/18-AOS1711

[17] Cardoso, J.-F. (1989). Source separation using higher order moments. In *International Conference on Acoustics, Speech, and Signal Processing,* 2109–2112. IEEE. https://doi.org/10.1109/ICASSP.1989.266878

[18] Chen, X. and Yang, Y. (2021). Cutoff for exact recovery of Gaussian mixture models. *IEEE Trans. Inf. Theory* **67** 4223–4238. MR4289373 https://doi.org/10.1109/TIT.2021.3063155

[19] Chen, X. and Yang, Y. (2021). Hanson-Wright inequality in Hilbert spaces with application to K-means clustering for non-Euclidean data. *Bernoulli* **27** 586–614. MR4177382 https://doi.org/10.3150/20-BEJ1251

[20] Chen, X. and Zhang, A.Y. (2021). Optimal clustering in anisotropic Gaussian mixture models. Preprint. Available at arXiv:2101.05402. https://doi.org/10.48550/arXiv.2101.05402

[21] Cherapanamjeri, Y., Hopkins, S.B., Kathuria, T., Raghavendra, P. and Tripuraneni, N. (2020). Algorithms for heavy-tailed statistics: Regression, covariance estimation, and beyond. In *STOC '20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* 601–609. New York: ACM. MR4141785 https://doi.org/10.1145/3357713.3384329

[22] Daskalakis, C., Tzamos, C. and Zampetakis, M. (2017). Ten steps of EM suffice for mixtures of two Gaussians. In *Conference on Learning Theory* 704–710. PMLR.

[23] 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

[24] Davis, D., Díaz, M. and Wang, K. Clustering a mixture of Gaussians with unknown covariance. Preprint. Available at arXiv:2110.01602. https://doi.org/10.48550/arXiv.2110.01602

[25] Davis, D., Díaz, M. and Wang, K. (2025). Supplement to "Clustering a mixture of Gaussians with unknown covariance." https://doi.org/10.3150/24-BEJ1798SUPP

[26] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

[27] Deshpande, Y. and Montanari, A. (2015). Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory* 523–562. PMLR.

[28] Deza, M.M. and Laurent, M. (2010). *Geometry of Cuts and Metrics*. *Algorithms and Combinatorics* **15**. Heidelberg: Springer. MR2841334 https://doi.org/10.1007/978-3-642-04295-9

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

[30] 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

[31] Dudeja, R. and Hsu, D. (2021). Statistical query lower bounds for tensor PCA. *J. Mach. Learn. Res.* **22** Paper No. 83, 51 pp. MR4253776

[32] Dwivedi, R., Ho, N., Khamaru, K., Wainwright, M.J., Jordan, M.I. and Yu, B. (2020). Singularity, misspecification and the convergence rate of EM. *Ann. Statist.* **48** 3161–3182. MR4185804 https://doi.org/10.1214/19-AOS1924

[33] Fei, Y. and Chen, Y. (2018). Hidden integrality of SDP relaxations for sub-Gaussian mixture models. In *Conference on Learning Theory* 1931–1965. PMLR.

[34] Feldman, V., Grigorescu, E., Reyzin, L., Vempala, S.S. and Xiao, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** Art. 8, 37 pp. MR3664576 https://doi.org/10.1145/3046674
[35] Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7** 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x
[36] Flammarion, N., Palaniappan, B. and Bach, F. (2017). Robust discriminitive clustering with sparse regularizers. *J. Mach. Learn. Res.* **18** Paper No. 80, 50 pp. MR3714243
[37] Friedman, J.H. (1989). Regularized discriminant analysis. *J. Amer. Statist. Assoc.* **84** 165–175. MR0999675
[38] Gamarnik, D., Jagannath, A. and Wein, A.S. (2020). Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science* 131–140. Los Alamitos, CA: IEEE Computer Soc. MR4232029 https://doi.org/10.1109/FOCS46700.2020.00021
[39] Ge, R., Huang, Q. and Kakade, S.M. (2015). Learning mixtures of Gaussians in high dimensions [extended abstract]. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing* 761–770. New York: ACM. MR3388256 https://doi.org/10.1145/2746539.2746616
[40] Ghosh, M., Jeronimo, F.G., Jones, C., Potechin, A. and Rajendran, G. (2020). Sum-of-squares lower bounds for Sherrington-Kirkpatrick via planted affine planes. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science* 954–965. Los Alamitos, CA: IEEE Computer Soc. MR4232101
[41] Giraud, C. and Verzelen, N. (2018). Partial recovery bounds for clustering with the relaxed K-means. *Math. Stat. Learn.* **1** 317–374. MR4059724
[42] Goemans, M.X. and Williamson, D.P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.* **42** 1115–1145. MR1412228 https://doi.org/10.1145/227683.227684
[43] Gurobi Optimization, LLC (2021). Gurobi optimizer reference manual.
[44] Hopkins, S.B. (2020). Mean estimation with sub-Gaussian rates in polynomial time. *Ann. Statist.* **48** 1193–1213. MR4102693 https://doi.org/10.1214/19-AOS1843
[45] Hopkins, S.B., Kothari, P.K., Potechin, A., Raghavendra, P., Schramm, T. and Steurer, D. (2017). The power of sum-of-squares for detecting hidden structures. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017* 720–731. Los Alamitos, CA: IEEE Computer Soc. MR3734275 https://doi.org/10.1109/FOCS.2017.72
[46] Hopkins, S.B., Shi, J., Schramm, T. and Steurer, D. (2016). Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 178–191. New York: ACM. MR3536564 https://doi.org/10.1145/2897518.2897529
[47] Jin, J., Ke, Z.T. and Wang, W. (2017). Phase transitions for high dimensional clustering and related problems. *Ann. Statist.* **45** 2151–2189. MR3718165 https://doi.org/10.1214/16-AOS1522
[48] Kearns, M. (1998). Efficient noise-tolerant learning from statistical queries. *J. ACM* **45** 983–1006. MR1678849 https://doi.org/10.1145/293347.293351
[49] Kothari, P.K., Steinhardt, J. and Steurer, D. (2018). Robust moment estimation and improved clustering via sum of squares. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* 1035–1046. New York: ACM. MR3826315
[50] Kunisky, D., Wein, A.S. and Bandeira, A.S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, Its Applications and Computation*. *Springer Proc. Math. Stat.* **385** 1–50. Cham: Springer. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1
[51] Kwon, J. and Caramanis, C. (2020). The EM algorithm gives sample-optimality for learning mixtures of well-separated Gaussians. In *Conference on Learning Theory* 2425–2487. PMLR.
[52] Lasserre, J.B. (2001). Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* **11** 796–817. MR1814045 https://doi.org/10.1137/S1052623400366802
[53] Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry and Applications*.
[54] Löffler, M., Zhang, A.Y. and Zhou, H.H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *Ann. Statist.* **49** 2506–2530. MR4338373 https://doi.org/10.1214/20-aos2044
[55] Lu, Y. and Zhou, H.H. (2016). Statistical and computational guarantees of Lloyd's algorithm and its variants. Preprint. Available at arXiv:1612.02099. https://doi.org/10.48550/arXiv.1612.02099

[56] Luo, Y. and Zhang, A.R. (2022). Tensor clustering with planted structures: Statistical optimality and computational limits. *Ann. Statist.* **50** 584–613. MR4382029 https://doi.org/10.1214/21-aos2123

[57] Mao, C. and Wein, A.S. (2021). Optimal spectral recovery of a planted vector in a subspace. Preprint. Available at arXiv:2105.15081. https://doi.org/10.48550/arXiv.2105.15081

[58] Meka, R., Potechin, A. and Wigderson, A. (2015). Sum-of-squares lower bounds for planted clique [extended abstract]. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing* 87–96. New York: ACM. MR3388186

[59] Mixon, D.G., Villar, S. and Ward, R. (2017). Clustering subgaussian mixtures by semidefinite programming. *Inf. Inference* **6** 389–415. MR3764529 https://doi.org/10.1093/imaiai/iax001

[60] Mohanty, S., Raghavendra, P. and Xu, J. (2020). Lifting sum-of-squares lower bounds: Degree-2 to degree-4. In *STOC '20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* 840–853. New York: ACM. MR4141804 https://doi.org/10.1145/3357713.3384319

[61] Moitra, A. and Valiant, G. (2010). Settling the polynomial learnability of mixtures of Gaussians. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010* 93–102. Los Alamitos, CA: IEEE Computer Soc. MR3024779

[62] Ndaoud, M. (2022). Sharp optimal recovery in the two component Gaussian mixture model. *Ann. Statist.* **50** 2096–2126. MR4474484 https://doi.org/10.1214/22-aos2178

[63] Nesterov, Y. and Nemirovskii, A. (1994). *Interior-Point Polynomial Algorithms in Convex Programming*. *SIAM Studies in Applied Mathematics* **13**. Philadelphia, PA: SIAM. MR1258086 https://doi.org/10.1137/1.9781611970791

[64] Parrilo, P.A. (2000). Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology.

[65] Royer, M. (2017). Adaptive clustering through semidefinite programming. In *Advances in Neural Information Processing Systems* 1795–1803.

[66] Schoenebeck, G. (2008). Linear level Lasserre lower bounds for certain k-CSPs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science* 593–602. IEEE. https://doi.org/10.1109/FOCS.2008.74

[67] Tan, Y.S. and Vershynin, R. (2018). Polynomial time and sample complexity for non-Gaussian component analysis: Spectral methods. In *Conference on Learning Theory* 498–534. PMLR.

[68] Vempala, S. and Wang, G. (2004). A spectral algorithm for learning mixture models. *J. Comput. System Sci.* **68** 841–860. MR2059647 https://doi.org/10.1016/j.jcss.2003.11.008

[69] Verzelen, N. and Arias-Castro, E. (2017). Detection and feature selection in sparse mixture models. *Ann. Statist.* **45** 1920–1950. MR3718157 https://doi.org/10.1214/16-AOS1513

[70] Wang, K., Yan, Y. and Díaz, M. (2020). Efficient clustering for stretched mixtures: Landscape and optimality. *Adv. Neural Inf. Process. Syst.* **33**.

[71] Wu, Y. and Zhou, H.H. (2021). Randomly initialized EM algorithm for two-component Gaussian mixture achieves near optimality in $O(\sqrt{n})$ iterations. *Math. Stat. Learn.* **4** 143–220. MR4383733 https://doi.org/10.4171/msl/29

[72] Xiao, H., Rasul, K. and Vollgraf, R. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. Preprint. Available at arXiv:1708.07747. https://doi.org/10.48550/arXiv.1708.07747

[73] Ye, J., Zhao, Z. and Wu, M. (2007). Discriminative k-means for clustering. *Adv. Neural Inf. Process. Syst.* **20** 1649–1656.

[74] Zadik, I., Song, M.J., Wein, A.S. and Bruna, J. (2022). Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory* 1247–1248. PMLR.

Bernoulli **31**(3), 2025, 2127–2151
https://doi.org/10.3150/24-BEJ1799

On the convergence of PINNs

NATHAN DOUMÈCHE[1,a], GÉRARD BIAU[1,2,b] and CLAIRE BOYER[1,2,c]

[1]*Sorbonne Université, CNRS, LPSM, F-75005 Paris, France,* [a]*nathan.doumeche@sorbonne-universite.fr,* [b]*gerard.biau@sorbonne-universite.fr,* [c]*claire.boyer@sorbonne-universite.fr*
[2]*Institut universitaire de France (IUF), Paris, France*

Physics-informed neural networks (PINNs) are a promising approach that combines the power of neural networks with the interpretability of physical modeling. PINNs have shown good practical performance in solving partial differential equations (PDEs) and in hybrid modeling scenarios, where physical models enhance data-driven approaches. However, it is essential to establish their theoretical properties in order to fully understand their capabilities and limitations. In this study, we highlight that classical training of PINNs can suffer from systematic overfitting. This problem can be addressed by adding a ridge regularization to the empirical risk, which ensures that the resulting estimator is risk-consistent for both linear and nonlinear PDE systems. However, the strong convergence of PINNs to a solution satisfying the physical constraints requires a more involved analysis using tools from functional analysis and calculus of variations. In particular, for linear PDE systems, an implementable Sobolev-type regularization allows to reconstruct a solution that not only achieves statistical accuracy but also maintains consistency with the underlying physics.

Keywords: Consistency; hybrid modeling; PDE solver; physics-informed neural networks

References

Agranovich, M.S. (2015). *Sobolev Spaces, Their Generalizations and Elliptic Problems in Smooth and Lipschitz Domains*. *Springer Monographs in Mathematics*. Cham: Springer. Revised translation of the 2013 Russian original. MR3287270 https://doi.org/10.1007/978-3-319-14648-5_2

Arnone, E., Kneip, A., Nobile, F. and Sangalli, L.M. (2022). Some first results on the consistency of spatial regression with partial differential equation regularization. *Statist. Sinica* **32** 209–238. MR4359630 https://doi.org/10.5705/ss.202019.0346

Arzani, A., Wang, J.X. and D'Souza, R.M. (2021). Uncovering near-wall blood flow from sparse data with physics-informed neural networks. *Phys. Fluids* **33** 071905. https://doi.org/10.1063/5.0055600

Azzimonti, L., Sangalli, L.M., Secchi, P., Domanin, M. and Nobile, F. (2015). Blood flow velocity field estimation via spatial regression with PDE penalization. *J. Amer. Statist. Assoc.* **110** 1057–1071. MR3420684 https://doi.org/10.1080/01621459.2014.946036

Brezis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. *Universitext*. New York: Springer. MR2759829

Cai, S., Wang, Z., Wang, S., Perdikaris, P. and Karniadakis, G.E. (2021). Physics-informed neural networks for heat transfer problems. *J. Heat Transf.* **143** 060801. https://doi.org/10.1115/1.4050542

Chandrajit, B., McLennan, L., Andeen, T. and Roy, A. (2023). Recipes for when physics fails: Recovering robust learning of physics informed neural networks. *Mach. Learn.: Sci. Technol.* **4** 015013. https://doi.org/10.1088/2632-2153/acb416

Costabal, F.S., Yang, Y., Perdikaris, P., Hurtado, D.E. and Kuhl, E. (2020). Physics-informed neural networks for cardiac activation mapping. *AIP Conf. Proc.* **8** 42. https://doi.org/10.3389/fphy.2020.00042

Cunha, B., Droz, C., Zine, A., Foulard, S. and Ichchou, M. (2023). A review of machine learning methods applied to structural dynamics and vibroacoustic. *Mech. Syst. Signal Process.* **110535**. https://doi.org/10.1016/j.ymssp.2023.110535

Cuomo, S., Schiano Di Cola, V., Giampaolo, F., Rozza, G., Raissi, M. and Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. *J. Sci. Comput.* **92** Paper No. 88, 62. MR4457972 https://doi.org/10.1007/s10915-022-01939-z

1350-7265

Davini, D., Samineni, B., Thomas, B., Tran, A.H., Zhu, C., Ha, K., Dasika, G. and White, L. (2021). Using physics-informed regularization to improve extrapolation capabilities of neural networks. In *Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021)*.

Daw, A., Karpatne, A., Watkins, W.D., Read, J.S. and Kumar, V. (2022). Physics-guided neural networks (PGNN): An application in lake temperature modeling. In *Knowledge Guided Machine Learning: Accelerating Discovery Using Scientific Knowledge and Data* (A. Karpatne, R. Kannan and V. Kumar, eds.) 352–372. New York: CRC Press/CRC. https://doi.org/10.1201/9781003143376-15

de Bézenac, E., Pajot, A. and Gallinari, P. (2019). Deep learning for physical processes: Incorporating prior scientific knowledge. *J. Stat. Mech. Theory Exp.* **2019** 124009, 21. MR4063607 https://doi.org/10.1088/1742-5468/ab3195

De Ryck, T. and Mishra, S. (2022). Error analysis for physics-informed neural networks (PINNs) approximating Kolmogorov PDEs. *Adv. Comput. Math.* **48** Paper No. 79, 40. MR4510503 https://doi.org/10.1007/s10444-022-09985-9

De Ryck, T., Lanthaler, S. and Mishra, S. (2021). On the approximation of rough functions with deep neural networks. *Neural Netw.* **143** 732–750.

de Wolff, T., Carrillo, H., Martí, L. and Sanchez-Pi, N. (2021). Towards optimally weighted physics-informed neural networks in ocean modelling. arXiv:2106.08747.

Doumèche, N., Biau, G. and Boyer, C. (2025). Supplement to "On the convergence of PINNs." https://doi.org/10.3150/24-BEJ1799SUPPA, https://doi.org/10.3150/24-BEJ1799SUPPB

Esfahani, I.C. (2023). A data-driven physics-informed neural network for predicting the viscosity of nanofluids. *AIP Adv.* **13** 025206. https://doi.org/10.1063/5.0132846

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

Ferraccioli, F., Sangalli, L.M. and Finos, L. (2022). Some first inferential tools for spatial regression with differential regularization. *J. Multivariate Anal.* **189** Paper No. 104866, 12. MR4384110 https://doi.org/10.1016/j.jmva.2021.104866

Fischer, S. and Steinwart, I. (2020). Sobolev norm learning rates for regularized least-squares algorithms. *J. Mach. Learn. Res.* **21** Paper No. 205, 38. MR4209491

Gokhale, G., Claessens, B. and Develder, C. (2022). Physics informed neural networks for control oriented thermal modeling of buildings. *Appl. Energy* **314** 118852. https://doi.org/10.1016/j.apenergy.2022.118852

Grisvard, P. (2011). *Elliptic Problems in Nonsmooth Domains*. *Classics in Applied Mathematics* **69**. Philadelphia, PA: SIAM. Reprint of the 1985 original [MR0775683], With a foreword by Susanne C. Brenner. MR3396210 https://doi.org/10.1137/1.9781611972030.ch1

Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning* (D. Precup and Y.W. Teh, eds.). *Proceedings of Machine Learning Research* **70** 1321–1330. PMLR. arXiv:1706.04599.

Hao, Z., Liu, S., Zhang, Y., Ying, C., Feng, Y., Su, H. and Zhu, J. (2022). Physics-informed machine learning: A survey on problems, methods and applications. arXiv:2211.08064.

He, Q., Barajas-Solano, D., Tartakovsky, G. and Tartakovsky, A.M. (2020). Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Adv. Water Resour.* **141** 103610. https://doi.org/10.1016/j.advwatres.2020.103610

Jagtap, A.D., Kawaguchi, K. and Karniadakis, G.E. (2020). Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *J. Comput. Phys.* **404** 109136, 23. MR4051868 https://doi.org/10.1016/j.jcp.2019.109136

Kapusuzoglu, B. and Mahadevan, S. (2020). Physics-informed and hybrid machine learning in additive manufacturing: Application to fused filament fabrication. *JOM-US* **72** 4695–4705. https://doi.org/10.1007/s11837-020-04438-4

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021). Physics-informed machine learning. *Nat. Rev. Phys.* **3** 422–440. https://doi.org/10.1038/s42254-021-00314-5

Krishnapriyan, A., Gholami, A., Zhe, S., Kirby, R. and Mahoney, M.W. (2021). Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems* (M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang and J.W. Vaughan, eds.) **34** 26548–26560. Curran Associates, Inc. arXiv:2109.01050.

Krogh, A. and Hertz, J. (1991). A simple weight decay can improve generalization. In *Advances in Neural Information Processing Systems* (J. Moody, S. Hanson and R.P. Lippmann, eds.) **4** 950–957. Morgan-Kaufmann.

Li, S., Wang, G., Di, Y., Wang, L., Wang, H. and Zhou, Q. (2023). A physics-informed neural network framework to predict 3D temperature field without labeled data in process of laser metal deposition. *Eng. Appl. Artif. Intell.* **120** 105908. https://doi.org/10.1016/j.engappai.2023.105908

Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. *Entropy* **23** 18. https://doi.org/10.3390/e23010018

Loshchilov, I. and Hutter, F. (2019). Decoupled weight decay regularization. In *7th International Conference on Learning Representations*. arXiv:1711.05101.

Mishra, S. and Molinaro, R. (2023). Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA J. Numer. Anal.* **43** 1–43. MR4565573 https://doi.org/10.1093/imanum/drab093

Nabian, M.A. and Meidani, H. (2020). Physics-driven regularization of deep neural networks for enhanced engineering design and analysis. *J. Comput. Inf. Sci. Eng.* **20** 011006. https://doi.org/10.1115/1.4044507

Pannell, J.J., Rigby, S.E. and Panoutsos, G. (2022). Physics-informed regularisation procedure in neural networks: An application in blast protection engineering. *Int. J. Prot. Struct.* **13** 555–578. https://doi.org/10.1177/20414196211073501

Qian, Y., Zhang, Y., Huang, Y. and Dong, S. (2023). Physics-informed neural networks for approximating dynamic (hyperbolic) PDEs of second order in time: Error analysis and algorithms. *J. Comput. Phys.* **495** Paper No. 112527, 30. MR4653320 https://doi.org/10.1016/j.jcp.2023.112527

Rai, R. and Sahu, C.K. (2020). Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber-physical system (CPS) focus. *IEEE Access* **8** 71050–71073. https://doi.org/10.1109/ACCESS.2020.2987324

Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378** 686–707. MR3881695 https://doi.org/10.1016/j.jcp.2018.10.045

Ramezankhani, M., Nazemi, A., Narayan, A., Voggenreiter, H., Harandi, M., Seethaler, R. and Milani, A.S. (2022). A data-driven multi-fidelity physics-informed learning framework for smart manufacturing: A composites processing case study. In *2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS)* 01–07. IEEE. https://doi.org/10.1109/ICPS51978.2022.9816983

Riel, B., Minchew, B. and Bischoff, T. (2021). Data-driven inference of the mechanics of slip along glacier beds using physics-informed neural networks: Case study on Rutford Ice Stream, Antarctica. *J. Adv. Model. Earth Syst.* **13** e2021MS002621. https://doi.org/10.1029/2021MS002621

Sangalli, L.M. (2021). Spatial regression with partial differential equation regularisation. *Int. Stat. Rev.* **89** 505–531. MR4411916 https://doi.org/10.1111/insr.12444

Shin, Y., Darbon, J. and Karniadakis, G.E. (2020). On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type PDEs. *Commun. Comput. Phys.* **28** 2042–2074. MR4188529 https://doi.org/10.4208/cicp.oa-2020-0193

Shin, Y., Zhang, Z. and Karniadakis, G.E. (2023). Error estimates of residual minimization using neural networks for linear PDEs. *JMLMC* **4** 73–101. https://doi.org/10.1615/JMachLearnModelComput.2023050411

Song, C., Alkhalifah, T. and Waheed, U.B. (2021). Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks. *Geophys. J. Int.* **225** 846–859. https://doi.org/10.1093/gji/ggab010

Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642

von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Walczak, M., Pfrommer, J., Pick, A., Ramamurthy, R., Garcke, J., Bauckhage, C. and Schuecker, J. (2023). Informed machine learning – a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Trans. Knowl. Data Eng.* **35** 614–633. https://doi.org/10.1109/TKDE.2021.3079836

Wang, S., Yu, X. and Perdikaris, P. (2022). When and why PINNs fail to train: A neural tangent kernel perspective. *J. Comput. Phys.* **449** Paper No. 110768, 28. MR4337814 https://doi.org/10.1016/j.jcp.2021.110768

Wang, C., Bentivegna, E., Zhou, W., Klein, L. and Elmegreen, B. (2020a). Physics-informed neural network super resolution for advection-diffusion models. In *Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020)*.

Wang, R., Kashinath, K., Mustafa, M., Albert, A. and Yu, R. (2020b). Towards physics-informed deep learning for turbulent flow prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 1457–1466. Association for Computing Machinery. arXiv:1911.08655.

Willard, J., Jia, X., Xu, S., Steinbach, M. and Kumar, V. (2023). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput. Surv.* **55**. https://doi.org/10.1145/3514228

Wu, S., Zhu, A., Tang, Y. and Lu, B. (2023). Convergence of physics-informed neural networks applied to linear second-order elliptic interface problems. *Commun. Comput. Phys.* **33** 596–627. MR4567824 https://doi.org/10.4208/cicp.oa-2022-0218

Xu, K., Zhang, M., Li, J., Du, S.S., Kawarabayashi, K.I. and Jegelka, S. (2021). How neural networks extrapolate: From feedforward to graph neural networks. In *International Conference on Learning Representations*. arXiv:2009.11848.

Zhang, R., Liu, Y. and Sun, H. (2020). Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling. *Eng. Struct.* **215** 110704. https://doi.org/10.1016/j.engstruct.2020.110704

Bernoulli **31**(3), 2025, 2152–2176
https://doi.org/10.3150/24-BEJ1800

A model-based approach to density estimation in sup-norm

GUILLAUME MAILLARD[a]

Department of Mathematics, University of Luxembourg, Esch-sur-Alzette, Grand Duchy of Luxembourg,
[a]*guillaume.maillard@ensai.fr*

We define a general method for finding a quasi-best approximant in sup-norm to a target density belonging to a given model, based on independent samples drawn from distributions which average to the target (which does not necessarily belong to the model). We also provide a general method for selecting among a countable family of such models. These estimators satisfy oracle inequalities in the general setting. The quality of the bounds depends on the volume of sets on which $|p-q|$ is close to its maximum, where p, q belong to the model (or possibly to two different models, in the case of model selection). This leads to optimal results in a number of settings, including piecewise polynomials on a given partition and anisotropic smoothness classes. Particularly interesting is the case of the single index model with fixed smoothness β, where we recover the one-dimensional rate: this was an open problem.

Keywords: Density estimation; minimax theory; model selection; robust estimation

References

Baraud, Y. (2016). Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electron. J. Stat.* **10** 1709–1728. MR3522658 https://doi.org/10.1214/15-EJS1055

Baraud, Y. (2021). Tests and estimation strategies associated to some loss functions. *Probab. Theory Related Fields* **180** 799–846. MR4288332 https://doi.org/10.1007/s00440-021-01065-1

Baraud, Y., Birgé, L. and Sart, M. (2017). A new method for estimation and model selection: ρ-estimation. *Invent. Math.* **207** 425–517. MR3595933 https://doi.org/10.1007/s00222-016-0673-5

Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028 https://doi.org/10.1007/s004400050210

Bertin, K. (2004). Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes. *Bernoulli* **10** 873–888. MR2093615 https://doi.org/10.3150/bj/1099579160

Dahmen, W., DeVore, R. and Scherer, K. (1980). Multidimensional spline approximation. *SIAM J. Numer. Anal.* **17** 380–402. MR0581486 https://doi.org/10.1137/0717033

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. *Springer Series in Statistics*. New York: Springer. MR1843146 https://doi.org/10.1007/978-1-4613-0125-7

Lepski, O. (2013). Multivariate density estimation under sup-norm loss: Oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. MR3099129 https://doi.org/10.1214/13-AOS1109

Lepski, O.V. and Rebelles, G. (2020). Structural adaptation in the density model. *Math. Stat. Learn.* **3** 345–386. MR4362042 https://doi.org/10.4171/msl/24

Massart, P. (2014). A non asymptotic walk in probability and statistics. In *Past, Present, and Future of Statistical Science* (Lin, X., Genest, C., Banks, D.L., Molenberghs, G., Scott, D.W. and Wang, J.-L., eds.). Chapman and Hall: London.

Maillard, G. (2025). Supplement to "A model-based approach to density estimation in sup-norm." https://doi.org/10.3150/24-BEJ1800SUPP

Massart, P. (2007). *Concentration Inequalities and Model Selection*. *Lecture Notes in Math.* **1896**. Berlin: Springer. MR2319879

1350-7265

Nikol'skiĭ, S.M. (1977). *Priblizhenie Funktsiĭ Mnogikh Peremennykh i Teoremy Vlozheniya*, 2nd ed. Moscow: "Nauka". MR0506247

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

Samarov, A. and Tsybakov, A. (2007). Aggregation of density estimators and dimension reduction. In *Advances in Statistical Modeling and Inference*. *Ser. Biostat.* **3** 233–251. Hackensack, NJ: World Sci. Publ. MR2416118 https://doi.org/10.1142/9789812708298_0012

von Luxburg, U. and Schölkopf, B. (2011). Statistical learning theory: Models, concepts, and results. In *Inductive Logic* (D.M. Gabbay, S. Hartmann and J. Woods, eds.). *Hand. Hist. Log.* **10** 651–706. North-Holland. https://doi.org/10.1016/B978-0-444-52936-7.50016-1

Bernoulli **31**(3), 2025, 2177–2200
https://doi.org/10.3150/24-BEJ1801

Goodness-of-fit tests for high-dimensional parametric multiresponse regressions

RAN LIU[1,a], JIAQI HUANG[2,b] and LIXING ZHU[3,4,c]

[1]*School of Mathematics and Statistics, Beijing Jiaotong University, Beijing, China,* [a]*ranl2@bjtu.edu.cn*
[2]*School of Statistics, Beijing Normal University, Beijing, China,* [b]*jhuang@mail.bnu.edu.cn*
[3]*Department of Statistics, Beijing Normal University at Zhuhai, Zhuhai, China,* [c]*lzhu@bnu.edu.cn*
[4]*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*

This study investigates goodness-of-fit testing in the context of high-dimensional parametric multiresponse regression, where the numbers of responses and parameters may diverge as the sample size increases. Particular attention is given to ordinary differential equation models, whose solutions can be interpreted as examples of multiresponse regression. To address the challenges posed by these models, we introduce two novel tests: a new global smoothing test and a new local smoothing test. The former exhibits a normal weak limit under the null hypothesis, representing a significant improvement over classical counterparts with intractable limiting null distributions in fixed-dimensional scenarios. The latter exhibits a dimension-agnostic property under certain conditions. Moreover, an increase in the number of responses can enhance the sensitivity of both tests to alternative models. Under specific regularity conditions, these tests can detect local alternatives distinct from the null hypothesis at rates faster than the fastest possible rates achievable by classical tests in fixed-dimensional cases. Numerical studies are conducted to evaluate the performance of the proposed tests and demonstrate their efficacy in real data analysis.

Keywords: Global smoothing test; high dimensionality; least squares estimation; local smoothing test; model specification

References

Bierens, H.J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65** 1129–1151. MR1475077 https://doi.org/10.2307/2171881

Chen, S., Shojaie, A. and Witten, D.M. (2017). Network reconstruction from high-dimensional ordinary differential equations. *J. Amer. Statist. Assoc.* **112** 1697–1707. MR3750892 https://doi.org/10.1080/01621459.2016.1229197

Chen, S.X. and Van Keilegom, I. (2009). A goodness-of-fit test for parametric and semi-parametric models in multiresponse regression. *Bernoulli* **15** 955–976. MR2597579 https://doi.org/10.3150/09-BEJ208

Currie, J. and Wilson, D.I. (2012). *OPTI: Lowering the Barrier Between Open Source Optimizers and the Industrial MATLAB User*. Savannah, Georgia, USA: Foundations of Computer-Aided Process Operations.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27** 1012–1040. MR1724039 https://doi.org/10.1214/aos/1018031266

Easton, R.W. (1991). Ordinary differential equations: An introduction to nonlinear analysis (Herbert Amann). *SIAM Rev.* **33** 669–670. https://doi.org/10.1137/1033154

Escanciano, J.C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22** 1030–1051. MR2328527 https://doi.org/10.1017/S0266466606060506

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. MR2065194 https://doi.org/10.1214/009053604000000256

Goel, N.S., Maitra, S.C. and Montroll, E.W. (1971). On the Volterra and other nonlinear models of interacting populations. *Rev. Modern Phys.* **43** 231–276. MR0484546 https://doi.org/10.1103/RevModPhys.43.231

González-Manteiga, W. and Crujeiras, R.M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22** 361–411. MR3093195 https://doi.org/10.1007/s11749-013-0327-5

Guo, X., Wang, T. and Zhu, L. (2016). Model checking for parametric single-index models: A dimension reduction model-adaptive approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1013–1035. MR3557187 https://doi.org/10.1111/rssb.12147

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096 https://doi.org/10.1016/0047-259X(84)90044-7

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774 https://doi.org/10.1214/aos/1176349403

Hooker, G. (2009). Forcing function diagnostics for nonlinear dynamics. *Biometrics* **65** 928–936. MR2649866 https://doi.org/10.1111/j.1541-0420.2008.01172.x

Hooker, G., Ellner, S.P., Roditi, L.D.V. and Earn, D.J.D. (2011). Parameterizing state-space models for infectious disease dynamics by generalized profiling: Measles in Ontario. *J. R. Soc. Interface* **8** 961–974. https://doi.org/10.1098/rsif.2010.0412

Khuri, A.I. (2001). An overview of the use of generalized linear models in response surface methodology. *Nonlinear Anal., Theory Methods Appl.* **47** 2023–2034. https://doi.org/10.1016/S0362-546X(01)00330-3

Lavergne, P. and Patilea, V. (2012). One for all and all for one: Regression checks with many regressors. *J. Bus. Econom. Statist.* **30** 41–52. MR2899183 https://doi.org/10.1198/jbes.2011.07152

Liang, H., Miao, H. and Wu, H. (2010). Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model. *Ann. Appl. Stat.* **4** 460–483. MR2758180 https://doi.org/10.1214/09-AOAS290

Liu, R., Fang, Y. and Zhu, L. (2023). Model checking for parametric ordinary differential equations systems. *Statist. Sinica* **33** 1853–1878. MR4607191 https://doi.org/10.5705/ss.202020.0429

Liu, R., Huang, J. and Zhu, L. (2025). Supplement to "Goodness-of-fit tests for high-dimensional parametric multiresponse regressions." https://doi.org/10.3150/24-BEJ1801SUPP

Liu, H., Palatucci, M. and Zhang, J. (2009). Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery. In *Proc. 26th Int. Conf. Mach. Learn.* 649–656.

Miao, H., Xia, X., Perelson, A.S. and Wu, H. (2011). On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* **53** 3–39. MR2785878 https://doi.org/10.1137/090757009

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316 https://doi.org/10.1214/aos/1031833666

Stute, W., González-Manteiga, W. and Presedo Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149. MR1614600 https://doi.org/10.2307/2669611

Stute, W., Thies, S. and Zhu, L.-X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.* **26** 1916–1934. MR1673284 https://doi.org/10.1214/aos/1024691363

Tan, F. and Zhu, L. (2019). Adaptive-to-model checking for regressions with diverging number of predictors. *Ann. Statist.* **47** 1960–1994. MR3953441 https://doi.org/10.1214/18-AOS1735

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76** 419–433. MR0624344

Wu, L., Qiu, X., Yuan, Y. and Wu, H. (2019). Parameter estimation and variable selection for big systems of linear ordinary differential equations: A matrix-based approach. *J. Amer. Statist. Assoc.* **114** 657–667. MR3963170 https://doi.org/10.1080/01621459.2017.1423074

Xue, H., Miao, H. and Wu, H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Ann. Statist.* **38** 2351–2387. MR2676892 https://doi.org/10.1214/09-AOS784

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289. MR1413644 https://doi.org/10.1016/0304-4076(95)01760-7

Zhu, L. and Li, R. (1998). Dimension-reduction type test for linearity of a stochastic regression model. *Acta Math. Appl. Sin. Engl. Ser.* **14** 165–175. MR1620882 https://doi.org/10.1007/BF02677423

Bernoulli **31**(3), 2025, 2201–2225
https://doi.org/10.3150/24-BEJ1802

A new adaptive local polynomial density estimation procedure on complicated domains

KARINE BERTIN[1,a], NICOLAS KLUTCHNIKOFF[2,b] and FRÉDÉRIC OUIMET[3,4,c]

[1] *CIMFAV-INGEMAT, Universidad de Valparaíso, Valparaíso, Chile,* [a]*karine.bertin@uv.cl*
[2] *Univ Rennes, CNRS, IRMAR-UMR 6625, F-35000 Rennes, France,* [b]*nicolas.klutchnikoff@univ-rennes2.fr*
[3] *Centre de recherches mathématiques, Université de Montréal, Montreal, Canada*
[4] *Department of Mathematics and Statistics, McGill University, Montreal, Canada,* [c]*frederic.ouimet2@mcgill.ca*

This paper presents a novel approach for pointwise estimation of multivariate density functions on known domains of arbitrary dimensions using nonparametric local polynomial estimators. Our method is highly flexible, as it applies to both simple domains, such as open connected sets, and more complicated domains that are not star-shaped around the point of estimation. This enables us to handle domains with sharp concavities, holes, and local pinches, such as polynomial sectors. Additionally, we introduce a data-driven selection rule based on the general ideas of Goldenshluger and Lepski. Our results demonstrate that the local polynomial estimators are minimax under a L^2 risk across a wide range of Hölder-type functional classes. In the adaptive case, we provide oracle inequalities and explicitly determine the convergence rate of our statistical procedure. Simulations on polynomial sectors show that our oracle estimates outperform those of the most popular alternative method, found in the `sparr` package for the `R` software. Our statistical procedure is implemented in an online `R` package which is readily accessible.

Keywords: Adaptive estimation; complicated domain; concave domain; local polynomial; minimax; nonparametric density estimation; oracle inequality; pinched domain; pointwise risk; polynomial sector

References

Aitchison, J. and Lauder, I.J. (1985). Kernel density estimation for compositional data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **34** 129–137. https://doi.org/10.2307/2347365

Ammous, S., Dedecker, J. and Duval, C. (2024). Adaptive directional estimator of the density in $\mathbb{R}^d$ for independent and mixing sequences. *J. Multivariate Anal.* **203** Paper No. 105332, 20 pp. MR4756038 https://doi.org/10.1016/j.jmva.2024.105332

Arnone, E., Ferraccioli, F., Pigolotti, C. and Sangalli, L.M. (2022). A roughness penalty approach to estimate densities over two-dimensional manifolds. *Comput. Statist. Data Anal.* **174** Paper No. 107527, 14 pp. MR4437662 https://doi.org/10.1016/j.csda.2022.107527

Azzimonti, L., Nobile, F., Sangalli, L.M. and Secchi, P. (2014). Mixed finite elements for spatial regression with PDE penalization. *SIAM/ASA J. Uncertain. Quantificat.* **2** 305–335. MR3283911 https://doi.org/10.1137/130925426

Azzimonti, L., Sangalli, L.M., Secchi, P., Domanin, M. and Nobile, F. (2015). Blood flow velocity field estimation via spatial regression with PDE penalization. *J. Amer. Statist. Assoc.* **110** 1057–1071. MR3420684 https://doi.org/10.1080/01621459.2014.946036

Babu, G.J. and Chaubey, Y.P. (2006). Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors. *Statist. Probab. Lett.* **76** 959–969. MR2270097 https://doi.org/10.1016/j.spl.2005.10.031

Bakka, H., Vanhatalo, J., Illian, J.B., Simpson, D. and Rue, H. (2019). Non-stationary Gaussian models with physical barriers. *Spat. Stat.* **29** 268–288. MR3903698 https://doi.org/10.1016/j.spasta.2019.01.002

Barry, R. (2021). latticeDensity: Density estimation and nonparametric regression on irregular regions. R package version 1.2.6. Available at https://cran.r-project.org/src/contrib/Archive/latticeDensity [This link was accessed on 2024–08–21].

Barry, R.P. and McIntyre, J. (2011). Estimating animal densities and home range in regions with irregular boundaries and holes: A lattice-based alternative to the kernel density estimator. *Ecol. Model.* **222** 1666–1672. https://doi.org/10.1016/j.ecolmodel.2011.02.016

Barry, R.P. and McIntyre, J. (2020). Lattice-based methods for regression and density estimation on complicated multidimensional regions. *Environ. Ecol. Stat.* **27** 571–589. https://doi.org/10.1007/s10651-020-00459-z

Bertin, K., El Kolei, S. and Klutchnikoff, N. (2019). Adaptive density estimation on bounded domains. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 1916–1947. MR4029144 https://doi.org/10.1214/18-AIHP938

Bertin, K. and Klutchnikoff, N. (2014). Adaptive estimation of a density function using beta kernels. *ESAIM Probab. Stat.* **18** 400–417. MR3333996 https://doi.org/10.1051/ps/2014010

Bertin, K. and Klutchnikoff, N. (2017). Pointwise adaptive estimation of the marginal density of a weakly dependent process. *J. Statist. Plann. Inference* **187** 115–129. MR3638047 https://doi.org/10.1016/j.jspi.2017.03.003

Bertin, K., Klutchnikoff, N. and Ouimet, F. (2024a). densityLocPoly: Density estimation on complicated domains; implementation of the local polynomial density estimator in 2d. R package version 0.0.2. Available at https://github.com/klutchnikoff/densityLocPoly [This link was accessed on 2024–08–21].

Bertin, K., Klutchnikoff, N. and Ouimet, F. (2025b). Supplement to "A new adaptive local polynomial density estimation procedure on complicated domains." https://doi.org/10.3150/24-BEJ1802SUPP

Bertin, K., Klutchnikoff, N., Léon, J.R. and Prieur, C. (2020). Adaptive density estimation on bounded domains under mixing conditions. *Electron. J. Stat.* **14** 2198–2237. MR4097810 https://doi.org/10.1214/20-EJS1682

Botev, Z.I., Grotowski, J.F. and Kroese, D.P. (2010). Kernel density estimation via diffusion. *Ann. Statist.* **38** 2916–2957. MR2722460 https://doi.org/10.1214/10-AOS799

Brunel, V.-E. (2018). Methods for estimation of convex sets. *Statist. Sci.* **33** 615–632. MR3881211 https://doi.org/10.1214/18-STS669

Cattaneo, M.D., Jansson, M. and Ma, X. (2020). Simple local polynomial density estimators. *J. Amer. Statist. Assoc.* **115** 1449–1455. MR4143477 https://doi.org/10.1080/01621459.2019.1635480

Cattaneo, M.D., Jansson, M. and Ma, X. (2022). lpdensity: Local polynomial density estimation and inference. *J. Stat. Softw.* **101** 1–25. https://doi.org/10.18637/jss.v101.i02

Cattaneo, M.D., Chandak, R., Jansson, M. and Ma, X. (2024). Boundary adaptive local polynomial conditional density estimators. *Bernoulli* **30** 3193–3223. MR4779862 https://doi.org/10.3150/23-bej1711

Chen, S.X. (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* **31** 131–145. MR1718494 https://doi.org/10.1016/S0167-9473(99)00010-9

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

Cline, D.B.H. and Hart, J.D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* **22** 69–84. MR1097362 https://doi.org/10.1080/02331889108802286

Davies, T.M. and Marshall, J.C. (2024). sparr: Spatial and spatiotemporal relative risk. R package version 2.3-15. Available at https://cran.r-project.org/web/packages/sparr/index.html [This link was accessed on 2024–08–21].

Davies, T.M., Marshall, J.C. and Hazelton, M.L. (2018). Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. *Stat. Med.* **37** 1191–1221. MR3777968 https://doi.org/10.1002/sim.7577

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20** 2008–2036. MR1193323 https://doi.org/10.1214/aos/1176348900

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. London: CRC Press. MR1383587 https://doi.org/10.1007/978-1-4899-3150-4

Feng, O.Y., Guntuboyina, A., Kim, A.K.H. and Samworth, R.J. (2021). Adaptation in multivariate log-concave density estimation. *Ann. Statist.* **49** 129–153. MR4206672 https://doi.org/10.1214/20-AOS1950

Ferraccioli, F., Arnone, E., Finos, L., Ramsay, J.O. and Sangalli, L.M. (2021). Nonparametric density estimation over complicated domains. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 346–368. MR4250279 https://doi.org/10.1111/rssb.12415

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation (Proc. Workshop, Heidelberg, 1979). Lecture Notes in Math.* **757** 23–68. Berlin: Springer. MR0564251 https://doi.org/10.1007/BFb0098489

Gasser, T., Müller, H.-G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252. MR0816088 https://doi.org/10.1111/j.2517-6161.1985.tb01350.x

Gawronski, W. and Stadtmüller, U. (1981). Smoothing histograms by means of lattice and continuous distributions. *Metrika* **28** 155–164. MR0638651 https://doi.org/10.1007/BF01902889

Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29** 1264–1280. MR1873330 https://doi.org/10.1214/aos/1013203453

Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. MR2543590 https://doi.org/10.3150/08-BEJ144

Goldenshluger, A. and Lepski, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields* **143** 41–71. MR2449122 https://doi.org/10.1007/s00440-007-0119-5

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214 https://doi.org/10.1214/11-AOS883

Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields* **159** 479–543. MR3230001 https://doi.org/10.1007/s00440-013-0512-1

Gu, C. (1993). Smoothing spline density estimation: A dimensionless automatic algorithm. *J. Amer. Statist. Assoc.* **88** 495–504. MR1224374 https://doi.org/10.1080/01621459.1993.10476300

Gu, C. and Qiu, C. (1993). Smoothing spline density estimation: Theory. *Ann. Statist.* **21** 217–234. MR1212174 https://doi.org/10.1214/aos/1176349023

Guillas, S. and Lai, M.-J. (2010). Bivariate splines for spatial functional regression models. *J. Nonparametr. Stat.* **22** 477–497. MR2662608 https://doi.org/10.1080/10485250903323180

Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Stat. Comput.* **3** 135–146. https://doi.org/10.1007/BF00147776

Jones, M.C. and Foster, P.J. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statist. Sinica* **6** 1005–1013. MR1422417

Klemelä, J. (2009). Multivariate histograms with data-dependent partitions. *Statist. Sinica* **19** 159–176. MR2487883

Klutchnikoff, N. (2014). Pointwise adaptive estimation of a multivariate function. *Math. Methods Statist.* **23** 132–150. MR3224636 https://doi.org/10.3103/S1066530714020045

Lai, M.-J. and Schumaker, L.L. (2007). *Spline Functions on Triangulations*. *Encyclopedia of Mathematics and Its Applications* **110**. Cambridge: Cambridge Univ. Press. MR2355272 https://doi.org/10.1017/CBO9780511721588

Lai, M.-J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statist. Sinica* **23** 1399–1417. MR3114719 https://doi.org/10.5705/ss.2010.278

Lepskii, O.V. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466. https://doi.org/10.1137/1135065

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

Liu, Y. and Wu, C. (2019). Point-wise estimation for anisotropic densities. *J. Multivariate Anal.* **171** 112–125. MR3892029 https://doi.org/10.1016/j.jmva.2018.11.014

Marron, J.S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **56** 653–671. MR1293239 https://doi.org/10.1111/j.2517-6161.1994.tb02006.x

McIntyre, J. and Barry, R.P. (2018). A lattice-based smoother for regions with irregular boundaries and holes. *J. Comput. Graph. Statist.* **27** 360–367. MR3816271 https://doi.org/10.1080/10618600.2017.1375935

McSwiggan, G., Baddeley, A. and Nair, G. (2017). Kernel density estimation on a linear network. *Scand. J. Stat.* **44** 324–345. MR3658517 https://doi.org/10.1111/sjos.12255

Miller, D.L. and Wood, S.N. (2014). Finite area smoothing with generalized distance splines. *Environ. Ecol. Stat.* **21** 715–731. MR3279587 https://doi.org/10.1007/s10651-014-0277-4

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* **78** 521–530. MR1130920 https://doi.org/10.2307/2337021

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

Petrone, S. (1999). Bayesian density estimation using Bernstein polynomials. *Canad. J. Statist.* **27** 105–126. MR1703623 https://doi.org/10.2307/3315494

Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 79–100. MR1881846 https://doi.org/10.1111/1467-9868.00326

Ramsay, T. (2002). Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 307–319. MR1904707 https://doi.org/10.1111/1467-9868.00339

Rebelles, G. (2015). Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli* **21** 1984–2023. MR3378457 https://doi.org/10.3150/14-BEJ633

Robinson, D. (2015). View package downloads over time with Shiny. Blog post. Available at http://varianceexplained.org/r/cran-view [This link was accessed on 2024–08–21].

Ruppert, D. and Cline, D.B.H. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *Ann. Statist.* **22** 185–210. MR1272080 https://doi.org/10.1214/aos/1176325365

Sangalli, L.M. (2021). Spatial regression with partial differential equation regularisation. *Int. Stat. Rev.* **89** 505–531. MR4411916 https://doi.org/10.1111/insr.12444

Sangalli, L.M., Ramsay, J.O. and Ramsay, T.O. (2013). Spatial spline regression models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 681–703. MR3091654 https://doi.org/10.1111/rssb.12009

Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. Theory Methods* **14** 1123–1136. MR0797636 https://doi.org/10.1080/03610928508828965

Scott-Hayward, L.A.S., Mackenzie, M.L., Ashe, E. and Williams, R. (2015). Modelling killer whale feeding behaviour using a spatially adaptive complex region spatial smoother (CReSS) and generalised estimating equations (GEEs). *J. Agric. Biol. Environ. Stat.* **20** 305–322. MR3396553 https://doi.org/10.1007/s13253-015-0209-2

Shiryaev, A.N. (1996). *Probability*, 2nd ed. *Graduate Texts in Mathematics* **95**. New York: Springer. MR1368405 https://doi.org/10.1007/978-1-4757-2539-1

Tsybakov, A.B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469. MR1700239 https://doi.org/10.1214/aos/1024691478

Varet, S., Lacour, C., Massart, P. and Rivoirard, V. (2023). Numerical performance of penalized comparison to overfitting for multivariate kernel density estimation. *ESAIM Probab. Stat.* **27** 621–667. MR4602387 https://doi.org/10.1051/ps/2022018

Vitale, R.A. (1975). Bernstein polynomial approach to density function estimation. In *Statistical Inference and Related Topics (Proc. Summer Res. Inst. Statist. Inference for Stochastic Processes, Indiana Univ., Bloomington, Ind., 1974, Vol. 2; Dedicated to Z. W. Birnbaum)* 87–99. New York: Academic Press. MR0397977 https://doi.org/10.1016/B978-0-12-568002-8.50011-2

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. *Monographs on Statistics and Applied Probability* **60**. London: CRC Press. MR1319818 https://doi.org/10.1007/978-1-4899-4493-1

Wang, H. and Ranalli, M.G. (2007). Low-rank smoothing splines on complicated domains. *Biometrics* **63** 209–217, 314. MR2345591 https://doi.org/10.1111/j.1541-0420.2006.00674.x

Wood, S.N., Bravington, M.V. and Hedley, S.L. (2008). Soap film smoothing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 931–955. MR2530324 https://doi.org/10.1111/j.1467-9868.2008.00665.x

Xu, M. and Samworth, R.J. (2021). High-dimensional nonparametric density estimation via symmetry and shape constraints. *Ann. Statist.* **49** 650–672. MR4255102 https://doi.org/10.1214/20-aos1972

Yu, S., Wang, G., Wang, L. and Yang, L. (2021). Multivariate spline estimation and inference for image-on-scalar regression. *Statist. Sinica* **31** 1463–1487. MR4297712 https://doi.org/10.5705/ss.202019.0188

Zhang, S. and Karunamuni, R.J. (1998). On kernel density estimation near endpoints. *J. Statist. Plann. Inference* **70** 301–316. MR1649872 https://doi.org/10.1016/S0378-3758(97)00187-0

Zhang, S. and Karunamuni, R.J. (2000). On nonparametric density estimation at the boundary. *J. Nonparametr. Stat.* **12** 197–221. MR1752313 https://doi.org/10.1080/10485250008832805

Zhang, S., Karunamuni, R.J. and Jones, M.C. (1999). An improved estimator of the density function at the boundary. *J. Amer. Statist. Assoc.* **94** 1231–1241. MR1731485 https://doi.org/10.2307/2669937

Zhou, L. and Pan, H. (2014). Smoothing noisy data for irregular regions using penalized bivariate splines on triangulations. *Comput. Statist.* **29** 263–281. MR3260122 https://doi.org/10.1007/s00180-013-0448-z

Bernoulli **31**(3), 2025, 2226–2250
https://doi.org/10.3150/24-BEJ1803

The functional central limit theorem with mean uncertainty under the sublinear expectation

XIAOFAN GUO[1,2,b] and XINPENG LI[1,a]

[1]*Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China,*
[a]*lixinpeng@sdu.edu.cn*
[2]*Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, China,*
[b]*guoxf@sustech.edu.cn*

We introduce a fundamental model for independent and identically distributed sequences with model uncertainty on the canonical space $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. Thanks to the well-defined upper and lower variances, we obtain a functional central limit theorem with mean uncertainty on the canonical space using a method based on the martingale central limit theorem and the stability of stochastic integrals in classical probability theory. We then extend it to the general sublinear expectation space through a representation theorem. Our results generalize Peng's central limit theorem with zero mean to include the mean uncertainty and provide a purely probabilistic proof instead of the existing nonlinear partial differential equation approach. As an application, we consider the two-armed bandit problem and generalize the corresponding central limit theorem from mean certainty to mean uncertainty.

Keywords: Canonical space; central limit theorem; independence and identical distribution; mean uncertainty; sublinear expectation; upper and lower variances

References

Brown, B.M. (1971). Martingale central limit theorems. *Ann. Math. Stat.* **42** 59–66. MR0290428 https://doi.org/10.1214/aoms/1177693494

Chen, Z. and Epstein, L.G. (2022). A central limit theorem for sets of probability measures. *Stochastic Process. Appl.* **152** 424–451. MR4456844 https://doi.org/10.1016/j.spa.2022.07.003

Chen, Z., Epstein, L.G. and Zhang, G. (2023). A central limit theorem, loss aversion and multi-armed bandits. *J. Econom. Theory* **209** 105645. MR4571225 https://doi.org/10.1016/j.jet.2023.105645

Denis, L., Hu, M. and Peng, S. (2011). Function spaces and capacity related to a sublinear expectation: Application to G-Brownian motion paths. *Potential Anal.* **34** 139–161. MR2754968 https://doi.org/10.1007/s11118-010-9185-x

Dolinsky, Y., Nutz, M. and Soner, H.M. (2012). Weak approximation of G-expectations. *Stochastic Process. Appl.* **122** 664–675. MR2868935 https://doi.org/10.1016/j.spa.2011.09.009

Fang, X., Peng, S., Shao, Q.-M. and Song, Y. (2019). Limit theorems with rate of convergence under sublinear expectations. *Bernoulli* **25** 2564–2596. MR4003558 https://doi.org/10.3150/18-BEJ1063

Guo, X. and Li, X. (2021). On the laws of large numbers for pseudo-independent random variables under sublinear expectation. *Statist. Probab. Lett.* **172** 109042. MR4207368 https://doi.org/10.1016/j.spl.2021.109042

Guo, X., Li, S. and Li, X. (2022). On the laws of the iterated logarithm with mean-uncertainty under sublinear expectations. *Probab. Uncertain. Quant. Risk* **7** 1–12. MR4405347 https://doi.org/10.3934/puqr.2022001

Guo, X., Li, S. and Li, X. (2023). Notes on Peng's independence in sublinear expectation theory. *Statist. Probab. Lett.* **193** 109719. MR4503431 https://doi.org/10.1016/j.spl.2022.109719

Hu, M. and Li, X. (2014). Independence under the G-expectation framework. *J. Theoret. Probab.* **27** 1011–1020. MR3245996 https://doi.org/10.1007/s10959-012-0471-y

Hu, M., Wang, F. and Zheng, G. (2016). Quasi-continuous random variables and processes under the G-expectation framework. *Stochastic Process. Appl.* **126** 2367–2387. MR3505230 https://doi.org/10.1016/j.spa.2016.02.003

1350-7265

Krylov, N.V. (1987). *Nonlinear Elliptic and Parabolic Equations of the Second Order*. *Mathematics and Its Applications (Soviet Series)* **7**. Dordrecht: D. Reidel Publishing Co. Translated from the Russian by P. L. Buzytsky. MR0901759 https://doi.org/10.1007/978-94-010-9557-0

Krylov, N.V. (2020). On Shige Peng's central limit theorem. *Stochastic Process. Appl.* **130** 1426–1434. MR4058278 https://doi.org/10.1016/j.spa.2019.05.005

Kurtz, T.G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070. MR1112406

Li, X. (2015). A central limit theorem for m-dependent random variables under sublinear expectations. *Acta Math. Appl. Sin. Engl. Ser.* **31** 435–444. MR3357911 https://doi.org/10.1007/s10255-015-0477-1

Li, S., Li, X. and Yuan, G.X. (2022). Upper and lower variances under model uncertainty and their applications in finance. *Int. J. Financ. Eng.* **9** 2250007. MR4398528 https://doi.org/10.1142/S2424786322500074

Li, X. and Lin, Y. (2017). Generalized Wasserstein distance and weak convergence of sublinear expectations. *J. Theoret. Probab.* **30** 581–593. MR3647070 https://doi.org/10.1007/s10959-015-0651-7

Li, M. and Shi, Y. (2010a). A general central limit theorem under sublinear expectations. *Sci. China Math.* **53** 1989–1994. MR2679081 https://doi.org/10.1007/s11425-010-3156-y

Li, X., Yu, M. and Zheng, S. (2023). On the calculation of upper variance under multiple probabilities. Available at arXiv:2311.12413.

Pardoux, É. and Peng, S.G. (1990). Adapted solution of a backward stochastic differential equation. *Systems Control Lett.* **14** 55–61. MR1037747 https://doi.org/10.1016/0167-6911(90)90082-6

Peng, S. (1997). Backward SDE and related g-expectation. In *Backward Stochastic Differential Equations (Paris, 1995–1996)*. *Pitman Res. Notes Math. Ser.* **364** 141–159. Harlow: Longman. MR1752680

Peng, S. (2008). A new central limit theorem under sublinear expectations. Available at arXiv:0803.2656.

Peng, S. (2009). Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations. *Sci. China Ser. A* **52** 1391–1411. MR2520583 https://doi.org/10.1007/s11425-009-0121-8

Peng, S. (2010). Tightness, weak compactness of nonlinear expectations and application to CLT. Available at arXiv:1006.2541.

Peng, S. (2019a). Law of large numbers and central limit theorem under nonlinear expectations. *Probab. Uncertain. Quant. Risk* **4** 4. MR3974670 https://doi.org/10.1186/s41546-019-0038-2

Peng, S. (2019b). *Nonlinear Expectations and Stochastic Calculus Under Uncertainty*. *Probability Theory and Stochastic Modelling* **95**. Berlin: Springer. MR3970247 https://doi.org/10.1007/978-3-662-59903-7

Peng, S., Yang, S. and Yao, J. (2020). Improving value-at-risk prediction under model uncertainty. *J. Financ. Econ.* **21** 228–259. https://doi.org/10.1093/jjfinec/nbaa022

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

Song, Y. (2020). Normal approximation by Stein's method under sublinear expectations. *Stochastic Process. Appl.* **130** 2838–2850. MR4080731 https://doi.org/10.1016/j.spa.2019.08.005

Terán, P. (2018). Sublinear expectations: On large sample behaviours, Monte Carlo method, and coherent upper previsions. In *The Mathematics of the Uncertain*. *Stud. Syst. Decis. Control* **142** 375–385. Cham: Springer. MR3793220

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. *Monographs on Statistics and Applied Probability* **42**. London: CRC Press. MR1145491 https://doi.org/10.1007/978-1-4899-3472-7

Zhang, L.-X. (2015). Donsker's invariance principle under the sub-linear expectation with an application to Chung's law of the iterated logarithm. *Commun. Math. Stat.* **3** 187–214. MR3357024 https://doi.org/10.1007/s40304-015-0055-0

Zhang, L.X. (2020). The convergence of the sums of independent random variables under the sub-linear expectations. *Acta Math. Sin. (Engl. Ser.)* **36** 224–244. MR4072701 https://doi.org/10.1007/s10114-020-8508-0

Zhang, L.-X. (2021a). Lindeberg's central limit theorems for martingale-like sequences under sub-linear expectations. *Sci. China Math.* **64** 1263–1290. MR4268893 https://doi.org/10.1007/s11425-018-9556-7

Zhang, L.-X. (2021b). On the laws of the iterated logarithm under sub-linear expectations. *Probab. Uncertain. Quant. Risk* **6** 409–460. MR4399897 https://doi.org/10.3934/puqr.2021020

Bernoulli **31**(3), 2025, 2251–2275
https://doi.org/10.3150/24-BEJ1804

Uniform error bound for PCA matrix denoising

XIN T. TONG[1,a], WANJIE WANG[2,b] and YUGUAN WANG[3,c]

[1]*Department of Mathematics, National University of Singapore, Singapore, Singapore,* [a]*xin.t.tong@nus.edu.sg*
[2]*Department of Statistics and Data Science, National University of Singapore, Singapore, Singapore,* [b]*wanjie.wang@nus.edu.sg*
[3]*Department of Statistics, University of Chicago, IL, Chicago, USA,* [c]*yuguanw@uchicago.edu*

Principal component analysis (PCA) is a simple and popular tool for processing high-dimensional data. We investigate its effectiveness for matrix denoising. We consider the clean data are generated from a low-dimensional subspace, but masked by independent high-dimensional sub-Gaussian noises with standard deviation σ. Under the low-rank assumption on the clean data with a mild spectral gap assumption, we prove that the distance between each pair of PCA-denoised data point and the clean data point is uniformly bounded by $O(\sigma \log n)$. To illustrate the spectral gap assumption, we show it can be satisfied when the clean data are independently generated with a non-degenerate covariance matrix. We then provide a general lower bound for the error of the denoised data matrix, which indicates PCA denoising gives a uniform error bound that is rate-optimal. Furthermore, we examine how the error bound impacts downstream applications such as clustering and manifold learning. Numerical results validate our theoretical findings and reveal the importance of the uniform error.

Keywords: High-dimensional data; matrix denoising; PCA; rate-optimal; uniform error

References

Abbe, E. and Sandon, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 670–688. Los Alamitos, CA: IEEE Computer Soc. MR3473334 https://doi.org/10.1109/FOCS.2015.47

Abbe, E., Fan, J., Wang, K. and Zhong, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

Abdi, H. and Williams, L.J. (2010). Principal component analysis. *Wiley Interdiscip. Rev.: Comput. Stat.* **2** 433–459. https://doi.org/10.1002/wics.101

Bao, Z., Ding, X. and Wang, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960

Bhatia, R. (2013). *Matrix Analysis* **169**. Berlin: Springer. https://doi.org/10.1007/978-1-4612-0653-8

Brunton, S.L. and Kutz, J.N. (2022). *Data-Driven Science and Engineering—Machine Learning, Dynamical Systems, and Control*. Cambridge: Cambridge Univ. Press. MR4394302 https://doi.org/10.1017/9781009089517

Bubeck, S. and von Luxburg, U. (2009). Nearest neighbor clustering: A baseline method for consistent clustering with arbitrary objective functions. *J. Mach. Learn. Res.* **10** 657–698. MR2491753

Cai, T.T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541

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

Cape, J., Tang, M. and Priebe, C.E. (2019b). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070

Chen, Y., Chi, Y., Fan, J., Ma, C. et al. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806. https://doi.org/10.1561/2200000079

1350-7265

Cressie, N. and Wikle, C.K. (2015). *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2848400

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

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

Donoho, D. and Gavish, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257

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

Donoho, D.L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math. Chall. Lect.* **1** 1–32. Available at https://dl.icdst.org/pdfs/files/236e636d7629c1a53e6ed4cce1019b6e.pdf.

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

Hartigan, J.A. and Wong, M.A. (1979). Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **28** 100–108. https://doi.org/10.2307/2346830

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

Hu, Y. and Wang, W. (2024). Network-adjusted covariates for community detection. *Biometrika* asae011. https://doi.org/10.1093/biomet/asae011

Jin, J., Ke, Z.T. and Wang, W. (2017). Phase transitions for high dimensional clustering and related problems. *Ann. Statist.* **45** 2151–2189. MR3718165 https://doi.org/10.1214/16-AOS1522

Jin, J. and Wang, W. (2016). Influential features PCA for high dimensional clustering. *Ann. Statist.* **44** 2323–2359. MR3576543 https://doi.org/10.1214/15-AOS1423

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

Jolliffe, I.T. (1972). Discarding variables in a principal component analysis. I. Artificial data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **21** 160–173. MR0311034 https://doi.org/10.2307/2346488

Jolliffe, I. (2005). Principal component analysis. In *Encyclopedia of Statistics in Behavioral Science*. https://doi.org/10.1002/0470013192.bsa501

Khoo, Y., Tong, X.T., Wang, W. and Wang, Y. (2024). Temporal label recovery from noisy dynamical data. Preprint. Available at arXiv:2406.13635. https://doi.org/10.48550/arXiv.2406.13635

Leeb, W.E. (2021). Matrix denoising for weighted loss functions and heterogeneous signals. *SIAM J. Math. Data Sci.* **3** 987–1012. MR4315482 https://doi.org/10.1137/20M1319577

Mei, S., Bai, Y. and Montanari, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637

Montanari, A., Ruan, F. and Yan, J. (2018). Adapting to unknown noise distribution in matrix denoising. Preprint. Available at arXiv:1810.02954. https://doi.org/10.48550/arXiv.1810.02954

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

O'Rourke, S., Vu, V. and Wang, K. (2024). Matrices with Gaussian noise: Optimal estimates for singular subspace perturbation. *IEEE Trans. Inf. Theory* **70** 1978–2002. MR4709772 https://doi.org/10.1109/tit.2023.3331010

Omran, M.G., Engelbrecht, A.P. and Salman, A. (2007). An overview of clustering methods. *Intell. Data Anal.* **11** 583–605. https://doi.org/10.3233/IDA-2007-11602

Pavlidis, T. and Horowitz, S.L. (1974). Segmentation of plane curves. *IEEE Trans. Comput.* **C-23** 860–870. MR0388733 https://doi.org/10.1109/t-c.1974.224041

Pollard, D. (1981). Strong consistency of k-means clustering. *Ann. Statist.* **9** 135–140. MR0600539

Radhakrishna Rao, C. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā Ser. A* **26** 329–358. MR0184375

Rakhlin, A. and Caponnetto, A. (2006). Stability of k-means clustering. *NeurIPS* **19**. Available at https://proceedings.neurips.cc/paper_files/paper/2006/file/58191d2a914c6dae66371c9dcdc91b41-Paper.pdf.

Reiß, M. and Wahl, M. (2020). Nonasymptotic upper bounds for the reconstruction error of PCA. *Ann. Statist.* **48** 1098–1123. MR4102689 https://doi.org/10.1214/19-AOS1839

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9 pp. MR3125258 https://doi.org/10.1214/ECP.v18-2865

Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* **27** 125–140. MR0140376 https://doi.org/10.1007/BF02289630

Singer, A. (2006). From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.* **21** 128–134. MR2238670 https://doi.org/10.1016/j.acha.2006.03.004

Singh, A. and Harrison, A. (1985). Standardized principal components. *Int. J. Remote Sens.* **6** 883–896. https://doi.org/10.1080/01431168508948511

Stewart, G.W. and Sun, J.G. (1990). *Matrix Perturbation Theory*. *Computer Science and Scientific Computing*. Boston, MA: Academic Press. MR1061154

Stone, H. (1961). Approximation of curves by line segments. *Math. Comput.* **15** 40–47. Available at https://www.ams.org/journals/mcom/1961-15-073/S0025-5718-1961-0119390-6/S0025-5718-1961-0119390-6.pdf.

Tao, T. (2012). *Topics in Random Matrix Theory*. *Graduate Studies in Mathematics* **132**. Providence, RI: Amer. Math. Soc. MR2906465 https://doi.org/10.1090/gsm/132

Tong, X., Wang, W. and Wang, Y. (2025). Supplement to "Uniform error bound for PCA matrix denoising." https://doi.org/10.3150/24-BEJ1804SUPP

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. Preprint. Available at arXiv:1011.3027. https://doi.org/10.48550/arXiv.1011.3027

von Luxburg, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416. MR2409803 https://doi.org/10.1007/s11222-007-9033-z

von Luxburg, U., Belkin, M. and Bousquet, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. MR2396807 https://doi.org/10.1214/009053607000000640

Vu, V. (2011). Singular vectors under random perturbation. *Random Structures Algorithms* **39** 526–538. MR2846302 https://doi.org/10.1002/rsa.20367

Wedin, P. (1972). Perturbation bounds in connection with singular value decomposition. *Nordisk Tidskr. Informationsbehandling (BIT)* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678

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

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

Bernoulli **31**(3), 2025, 2276–2301
https://doi.org/10.3150/24-BEJ1805

Asymptotic normality of log likelihood ratio and fundamental limit of the weak detection for spiked Wigner matrices

HYE WON CHUNG[1,a], JIHO LEE[2,b] and JI OON LEE[3,c]

[1] *School of Electrical Engineering, KAIST, Daejeon, 34141, Korea,* [a] *hwchung@kaist.ac.kr*
[2] *Korea Science Academy of KAIST, Busan, 10547, Korea,* [b] *efidiaf@ksa.hs.kr*
[3] *Department of Mathematical Sciences, KAIST, Daejeon, 34141, Korea,* [c] *jioon.lee@kaist.edu*

We consider the problem of detecting the presence of a signal in a rank-one spiked Wigner model. For general non-Gaussian noise, assuming that the signal is drawn from the Rademacher prior, we prove that the log likelihood ratio (LR) of the spiked model against the null model converges to a Gaussian when the signal-to-noise ratio is below a certain threshold. The threshold is optimal in the sense that the reliable detection is possible by a transformed principal component analysis (PCA) above it. From the mean and the variance of the limiting Gaussian for the log-LR, we compute the limit of the sum of the Type-I error and the Type-II error of the likelihood ratio test. We also prove similar results for a rank-one spiked IID model where the noise is asymmetric but the signal is symmetric.

Keywords: Asymptotic normality; likelihood ratio; spiked Wigner matrix

References

[1] Aizenman, M., Lebowitz, J.L. and Ruelle, D. (1987). Some rigorous results on the Sherrington-Kirkpatrick spin glass model. *Comm. Math. Phys.* **112** 3–20. MR0904135
[2] 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
[3] Baik, J. and Lee, J.O. (2016). Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model. *J. Stat. Phys.* **165** 185–224. MR3554380 https://doi.org/10.1007/s10955-016-1610-0
[4] Barbier, J., Dia, M., Macris, N., Krzakala, F., Lesieur, T. and Zdeborová, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Adv. Neural Inf. Process. Syst.* **29** 424–432.
[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] Benaych-Georges, F. and Nadakuditi, R.R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019
[7] Capitaine, M., Donati-Martin, C. and Féral, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394
[8] Chen, Y., Cheng, C. and Fan, J. (2021). Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *Ann. Statist.* **49** 435–458. MR4206685 https://doi.org/10.1214/20-AOS1963
[9] Chung, H.W., Chung, H.W. and Lee, J.O. (2021). Detection of signal in the spiked rectangular models. In *Proceedings of the 38th International Conference on Machine Learning* **139** 5158–5167.

1350-7265

[10] Chung, H.W., Lee, J. and Lee, J.O. (2025). Supplement to "Asymptotic normality of log likelihood ratio and fundamental limit of the weak detection for spiked Wigner matrices." https://doi.org/10.3150/24-BEJ1805SUPP
[11] Chung, H.W. and Lee, J.O. (2019). Weak detection in the spiked Wigner model. In *Proceedings of the 36th International Conference on Machine Learning* **97** 1233–1241.
[12] Dobriban, E. (2017). Sharp detection in PCA under correlations: All eigenvalues matter. *Ann. Statist.* **45** 1810–1833. MR3670197 https://doi.org/10.1214/16-AOS1514
[13] Edwards, S.F. and Anderson, P.W. (1975). Theory of spin glasses. *J. Phys. F* **5** 965. https://doi.org/10.1088/0305-4608/5/5/017
[14] El Alaoui, A. and Jordan, M.I. (2018). Detection limits in the high-dimensional spiked rectangular model. In *Conference on Learning Theory* 410–438.
[15] El Alaoui, A., Krzakala, F. and Jordan, M. (2020). Fundamental limits of detection in the spiked Wigner model. *Ann. Statist.* **48** 863–885. MR4102679 https://doi.org/10.1214/19-AOS1826
[16] Féral, D. and Péché, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 https://doi.org/10.1007/s00220-007-0209-3
[17] 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
[18] Johnstone, I.M. and Onatski, A. (2020). Testing in high-dimensional spiked models. *Ann. Statist.* **48** 1231–1254. MR4124321 https://doi.org/10.1214/18-AOS1697
[19] Montanari, A., Reichman, D. and Zeitouni, O. (2017). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inf. Theory* **63** 1572–1579. MR3625981 https://doi.org/10.1109/TIT.2016.2637959
[20] Onatski, A., Moreira, M.J. and Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 https://doi.org/10.1214/13-AOS1100
[21] Onatski, A., Moreira, M.J. and Hallin, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485 https://doi.org/10.1214/13-AOS1181
[22] Parisi, G. (1980). A sequence of approximated solutions to the S–K model for spin glasses. *J. Phys. A* **13** L115. https://doi.org/10.1088/0305-4470/13/4/009
[23] Péché, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787 https://doi.org/10.1007/s00440-005-0466-z
[24] Perry, A., Wein, A.S., Bandeira, A.S. and Moitra, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. MR3845022 https://doi.org/10.1214/17-AOS1625
[25] Sherrington, D. and Kirkpatrick, S. (1975). Solvable model of a spin-glass. *Phys. Rev. Lett.* **35** 1792. https://doi.org/10.1103/PhysRevLett.35.1792
[26] Talagrand, M. (2006). The Parisi formula. *Ann. of Math. (2)* **163** 221–263. MR2195134 https://doi.org/10.4007/annals.2006.163.221

Bernoulli **31**(3), 2025, 2302–2322
https://doi.org/10.3150/24-BEJ1806

Computable bounds for solutions to Poisson's equation and perturbation of Markov kernels

LOÏC HERVÉ[a] and JAMES LEDOUX[b]

Univ Rennes, INSA Rennes, CNRS, IRMAR-UMR 6625, F-35000, France, [a]*Loic.Herve@insa-rennes.fr,* [b]*James.Ledoux@insa-rennes.fr*

We consider a Markov kernel on a measurable space, satisfying a minorization condition and a modulated drift condition. Then we show that there exists a solution to the so-called Poisson equation whose norm can be bounded from above using the modulated drift condition. This new bound is very simple and can be easily computed. This result is obtained using the submarkov residual kernel given by the minorization condition. Such a bound allows us to provide new control on the weighted total variation norms of the deviation between the invariant probability measure π_{θ_0} of a Markov kernel P_{θ_0} and the invariant probability measure π_θ of some perturbation P_θ of P_{θ_0}. From the standard connexion between Poisson's equation and the central limit theorem, a simple and computable bound on the asymptotic variance is also derived.

Keywords: Asymptotic variance; drift conditions; invariant probability measure; perturbed Markov kernels; Poisson's equation

References

Alquier, P., Friel, N., Everitt, R. and Boland, A. (2016). Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Stat. Comput.* **26** 29–47. MR3439357 https://doi.org/10.1007/s11222-014-9521-x

Altman, E., Avrachenkov, K.E. and Núñez-Queija, R. (2004). Perturbation analysis for denumerable Markov chains with application to queueing models. *Adv. in Appl. Probab.* **36** 839–853. MR2079917 https://doi.org/10.1239/aap/1093962237

Andrieu, C., Fort, G. and Vihola, M. (2015). Quantitative convergence rates for subgeometric Markov chains. *J. Appl. Probab.* **52** 391–404. MR3372082 https://doi.org/10.1239/jap/1437658605

Baxendale, P.H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987 https://doi.org/10.1214/105051604000000710

Delyon, B. (2017). Convergence rate of the powers of an operator. Applications to stochastic systems. *Bernoulli* **23** 2129–2180. MR3648028 https://doi.org/10.3150/15-BEJ778

Douc, R., Fort, G., Moulines, E. and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14** 1353–1377. MR2071426 https://doi.org/10.1214/105051604000000323

Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018). *Markov Chains*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

Dubins, L. and Freedman, D. (1964). Measurable sets of measures. *Pacific J. Math.* **14** 1211–1222. MR0174687

Durmus, A., Fort, G. and Moulines, É. (2016). Subgeometric rates of convergence in Wasserstein distance for Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1799–1822. MR3573296 https://doi.org/10.1214/15-AIHP699

Ferré, D., Hervé, L. and Ledoux, J. (2013). Regular perturbation of V-geometrically ergodic Markov chains. *J. Appl. Probab.* **50** 184–194. MR3076780 https://doi.org/10.1239/jap/1363784432

Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statist. Probab. Lett.* **49** 401–410. MR1796485 https://doi.org/10.1016/S0167-7152(00)00074-2

Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** 57–99. MR1947960 https://doi.org/10.1016/S0304-4149(02)00182-5

Glynn, P.W. and Meyn, S.P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

Hairer, M. and Mattingly, J.C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*. *Progress in Probability* **63** 109–117. Basel: Birkhäuser/Springer Basel AG. MR2857021 https://doi.org/10.1007/978-3-0348-0021-1_7

Hervé, L. and Ledoux, J. (2014). Approximating Markov chains and V-geometric ergodicity via weak perturbation theory. *Stochastic Process. Appl.* **124** 613–638. MR3131307 https://doi.org/10.1016/j.spa.2013.09.003

Hervé, L. and Ledoux, J. (2021). State-discretization of V-geometrically ergodic Markov chains and convergence to the stationary distribution. *Methodol. Comput. Appl. Probab.* **22** 905–925. MR4129119 https://doi.org/10.1007/s11009-019-09746-0

Hervé, L. and Ledoux, J. (2023a). Rate of convergence of Nummelin-type representation of the invariant distribution of a Markov chain via the residual kernel. *Electron. Commun. Probab.* **28** Paper No. 58, 13. MR4684071 https://doi.org/10.1007/s13253-022-00509-y

Hervé, L. and Ledoux, J. (2023b). Robustness of iterated function systems of Lipschitz maps. *J. Appl. Probab.* **60** 921–944. MR4624049 https://doi.org/10.1017/jpr.2022.107

Hervé, L. and Ledoux, J. (2024). Explicit bounds for spectral theory of geometrically ergodic Markov kernels and applications. *Bernoulli* **30** 581–609. MR4665590 https://doi.org/10.3150/23-bej1609

Jarner, S.F. and Roberts, G.O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063 https://doi.org/10.1214/aoap/1015961162

Jarner, S.F. and Tweedie, R.L. (2002). Convergence rates and moments of Markov chains associated with the mean of Dirichlet processes. *Stochastic Process. Appl.* **101** 257–271. MR1931269 https://doi.org/10.1016/S0304-4149(02)00139-4

Jarner, S.F. and Tweedie, R.L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* **9** 559–578. MR1996270 https://doi.org/10.3150/bj/1066223269

Johndrow, J.E., Mattingly, J.C., Mukherjee, S. and Dunson, D. (2015). Optimal approximating Markov chains for Bayesian inference. arXiv:1508.03387v3.

Jones, G.L. (2004). On the Markov chain central limit theorem. *Probab. Surv.* **1** 299–320. MR2068475 https://doi.org/10.1214/154957804100000051

Kartashov, N.V. (1986). Inequalities in stability and ergodicity theorems for Markov chains with a general phase space. Parts I and II. *Teor. Veroyatn. Primen.* **30** 478–485, 505–515.

Keller, G. (1982). Stochastic stability in some chaotic dynamical systems. *Monatsh. Math.* **94** 313–333. MR0685377 https://doi.org/10.1007/BF01667385

Keller, G. and Liverani, C. (1999). Stability of the spectrum for transfer operators. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (4)* **28** 141–152. MR1679080

Li, X. and Hou, Z. (2012). Subgeometric rates of convergence of the $GI/G/1$ queueing system. *Acta Math. Sci. Ser. B Engl. Ed.* **32** 1983–1996. MR2960764 https://doi.org/10.1016/S0252-9602(12)60154-4

Liu, Y. and Li, W. (2018). Error bounds for augmented truncation approximations of Markov chains via the perturbation method. *Adv. in Appl. Probab.* **50** 645–669. MR3832888 https://doi.org/10.1017/apr.2018.28

Liu, Y.Y. and Hou, Z.T. (2007). Explicit convergence rates of the embedded $M/G/1$ queue. *Acta Math. Sin. (Engl. Ser.)* **23** 1289–1296. MR2331144 https://doi.org/10.1007/s10114-005-0917-6

Masuyama, H. (2019). Specific bounds for a probabilistically interpretable solution of the Poisson equation for general state-space Markov chains with queueing applications. arXiv:1909.07567.

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

Medina-Aguayo, F.J., Lee, A. and Roberts, G.O. (2016). Stability of noisy Metropolis-Hastings. *Stat. Comput.* **26** 1187–1211. MR3538632 https://doi.org/10.1007/s11222-015-9604-3

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

Meyn, S.P. (2022). *Control Systems and Reinforcement Learning*. Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/9781009051873

Mitrophanov, A.Yu. (2005). Sensitivity and convergence of uniformly ergodic Markov chains. *J. Appl. Probab.* **42** 1003–1014. MR2203818 https://doi.org/10.1239/jap/1134587812

Mouhoubi, Z. (2021). Perturbation and stability bounds for ergodic general state Markov chains with respect to various norms. *Matematiche (Catania)* **76** 243–276. MR4284893 https://doi.org/10.4418/2021.76.1.13

Mouhoubi, Z. and Aïssani, D. (2010). New perturbation bounds for denumerable Markov chains. *Linear Algebra Appl.* **432** 1627–1649. MR2592907 https://doi.org/10.1016/j.laa.2009.11.020

Negrea, J. and Rosenthal, J.S. (2021). Approximations of geometrically ergodic reversible Markov chains. *Adv. in Appl. Probab.* **53** 981–1022. MR4342575 https://doi.org/10.1017/apr.2021.10

Nummelin, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. *Cambridge Tracts in Mathematics* **83**. Cambridge: Cambridge Univ. Press. MR0776608 https://doi.org/10.1017/CBO9780511526237

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

Rudolf, D. and Schweizer, N. (2018). Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli* **24** 2610–2639. MR3779696 https://doi.org/10.3150/17-BEJ938

Schweitzer, P.J. (1968). Perturbation theory and finite Markov chains. *J. Appl. Probab.* **5** 401–413. MR0234527 https://doi.org/10.2307/3212261

Seneta, E. (1993). Sensitivity of finite Markov chains under perturbation. *Statist. Probab. Lett.* **17** 163–168. MR1223539 https://doi.org/10.1016/0167-7152(93)90011-7

Shardlow, T. and Stuart, A.M. (2000). A perturbation theory for ergodic Markov chains and application to numerical approximations. *SIAM J. Numer. Anal.* **37** 1120–1137. MR1756418 https://doi.org/10.1137/S0036142998337235

Tuominen, P. and Tweedie, R.L. (1994). Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. in Appl. Probab.* **26** 775–798. MR1285459 https://doi.org/10.2307/1427820

Bernoulli **31**(3), 2025, 2323–2350
https://doi.org/10.3150/24-BEJ1807

Scaling of piecewise deterministic Monte Carlo for anisotropic targets

JORIS BIERKENS[1,a], KENGO KAMATANI[2,b] and GARETH O. ROBERTS[3,c]

[1] *Delft Institute of Applied Mathematics, Delft, Netherlands,* [a] *joris.bierkens@tudelft.nl*
[2] *Institute of Statistical Mathematics, Tokyo, Japan,* [b] *kamatani@ism.ac.jp*
[3] *Department of Statistics, University of Warwick, Coventry, United Kingdom,* [c] *Gareth.O.Roberts@warwick.ac.uk*

Piecewise deterministic Markov processes (PDMPs) are a type of continuous-time Markov process that combine deterministic flows with jumps. Recently, PDMPs have garnered attention within the Monte Carlo community as a potential alternative to traditional Markov chain Monte Carlo (MCMC) methods. The Zig-Zag sampler and the Bouncy Particle Sampler are commonly used examples of the PDMP methodology which have also yielded impressive theoretical properties, but little is known about their robustness to extreme dependence or anisotropy of the target density. It turns out that PDMPs may suffer from poor mixing due to anisotropy and this paper investigates this effect in detail in the stylised but important Gaussian case. To this end, we employ a multi-scale analysis framework in this paper. Our results show that when the Gaussian target distribution has two scales, of order 1 and ϵ, the computational cost of the Bouncy Particle Sampler is of order ϵ^{-1}, and the computational cost of the Zig-Zag sampler is ϵ^{-2}. In comparison, the cost of the traditional MCMC methods such as RWM is of order ϵ^{-2}, at least when the dimensionality of the small component is more than 1. Therefore, there is a robustness advantage to using PDMPs in this context.

Keywords: Markov process; Monte Carlo methods; multi-scale analysis

References

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

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

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

Beskos, A., Roberts, G., Thiery, A. and Pillai, N. (2018). Asymptotic analysis of the random walk Metropolis algorithm on ridged densities. *Ann. Appl. Probab.* **28** 2966–3001. MR3847978 https://doi.org/10.1214/18-AAP1380

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

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

Bierkens, J., Kamatani, K. and Roberts, G.O. (2022). High-dimensional scaling limits of piecewise deterministic sampling algorithms. *Ann. Appl. Probab.* **32** 3361–3407. MR4497848 https://doi.org/10.1214/21-aap1762

Bierkens, J., Kamatani, K. and Roberts, G.O. (2025). Supplement to "Scaling of piecewise deterministic Monte Carlo for anisotropic targets." https://doi.org/10.3150/24-BEJ1807SUPP

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

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. A Wiley-Interscience Publication. MR1700749 https://doi.org/10.1002/9780470316962

1350-7265

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

Chevallier, A., Fearnhead, P. and Sutton, M. (2023). Reversible jump PDMP samplers for variable selection. *J. Amer. Statist. Assoc.* **118** 2915–2927. MR4681630 https://doi.org/10.1080/01621459.2022.2099402

Davis, M.H.A. (1984). Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46** 353–388. With discussion. MR0790622

Deligiannidis, G., Paulin, D., Bouchard-Côté, A. and Doucet, A. (2021). Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. *Ann. Appl. Probab.* **31** 2612–2662. MR4350970 https://doi.org/10.1214/20-aap1659

Durmus, A., Guillin, A. and Monmarché, P. (2021). Piecewise deterministic Markov processes and their invariant measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1442–1475. MR4291455 https://doi.org/10.1214/20-aihp1125

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

Glynn, P.W. and Meyn, S.P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

Graham, M.M., Thiery, A.H. and Beskos, A. (2022). Manifold Markov chain Monte Carlo methods for Bayesian inference in diffusion models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1229–1256. MR4494159 https://doi.org/10.1111/rssb.12497

Holderrieth, P. (2021). Cores for piecewise-deterministic Markov processes used in Markov chain Monte Carlo. *Electron. Commun. Probab.* **26** Paper No. 57, 12. MR4346861 https://doi.org/10.1214/21-ecp430

Kulik, A. and Scheutzow, M. (2015). A coupling approach to Doob's theorem. *Atti Accad. Naz. Lincei, Rend. Lincei, Mat. Appl.* **26** 83–92. MR3345324 https://doi.org/10.4171/RLM/694

Pavliotis, G.A. and Stuart, A.M. (2008). *Multiscale Methods: Averaging and Homogenization*. *Texts in Applied Mathematics* **53**. New York: Springer. MR2382139

Peters, E.A.J.F. and de With, G. (2012). Rejection-free Monte Carlo sampling for general potentials. *Phys. Rev. E* **85** 026703. https://doi.org/10.1103/PhysRevE.85.026703

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

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

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

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

Roberts, G.O. and Sahu, S.K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59** 291–317. MR1440584 https://doi.org/10.1111/1467-9868.00070

Sherlock, C. and Thiery, A.H. (2022). A discrete bouncy particle sampler. *Biometrika* **109** 335–349. MR4430961 https://doi.org/10.1093/biomet/asab013

Sherlock, C., Thiery, A.H., Roberts, G.O. and Rosenthal, J.S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606 https://doi.org/10.1214/14-AOS1278

Zanella, G. and Roberts, G. (2021). Multilevel linear models, Gibbs samplers and multigrid decompositions (with discussion). *Bayesian Anal.* **16** 1308–1390. MR4381136 https://doi.org/10.1214/20-BA1242

Bernoulli **31**(3), 2025, 2351–2376
https://doi.org/10.3150/24-BEJ1808

Asymptotic limits of spiked eigenvalues and eigenvectors of signal-plus-noise matrices with weak signals and heteroskedastic noise

XIAOYU LIU[1,a], YIMING LIU[1,b], GUANGMING PAN[2,c], LINGYUE ZHANG[3,d] and ZHIXIANG ZHANG[4,e]

[1]*Department of Statistics, School of Economics, Jinan University, Guangzhou, China,* [a]*xyliu0075@jnu.edu.cn,* [b]*liuyiming@jnu.edu.cn*
[2]*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore,* [c]*gmpan@ntu.edu.sg*
[3]*School of Statistics, Dongbei University of Finance and Economics, Dalian, China,* [d]*lingyue_zhang@126.com*
[4]*Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau, China,* [e]*zhixzhang@um.edu.mo*

This paper is to study a signal-plus-noise model in high dimensional settings when the dimension and the sample size are comparable. Specifically, we assume that the noise has a general covariance matrix that allows for heteroskedasticity, and that the deterministic signal has the same magnitude as the noise and can have a rank that tends to infinity. We develop the asymptotic limits of the left and right spiked singular vectors of the signal-plus-noise data matrix and the limits of the spiked eigenvalues of the corresponding Gram matrix. As an application, we propose a new criterion to estimate the number of clusters in clustering problems.

Keywords: Deterministic equivalents; signal-plus-noise matrices; spectral clustering; spiked eigenvalues and eigenvectors

References

[1] Abbe, E., Fan, J., Wang, K. and Zhong, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

[2] Bai, Z., 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

[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 Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

[5] Bai, Z.D. and Silverstein, J.W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051 https://doi.org/10.1214/aop/1022855421

[6] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

[7] Bao, Z., Ding, X., Wang, J. and Wang, K. (2022). Statistical inference for principal components of spiked covariance matrices. *Ann. Statist.* **50** 1144–1169. MR4404931 https://doi.org/10.1214/21-aos2143

[8] Bao, Z., Ding, X. and Wang, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960

1350-7265

[9] Benaych-Georges, F., Guionnet, A. and Maida, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. MR2835249 https://doi.org/10.1214/EJP.v16-929
[10] 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
[11] Benaych-Georges, F. and Nadakuditi, R.R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019
[12] Berry, S. and Tamer, E. (2006). *Identification in Models of Oligopoly Entry* **42**. Cambridge: Cambridge Univ. Press.
[13] Bradley, P.S., Fayyad, U.M. and Mangasarian, O.L. (1999). Mathematical programming for data mining: Formulations and challenges. *INFORMS J. Comput.* **11** 217–238. MR1704656 https://doi.org/10.1287/ijoc.11.3.217
[14] Cai, T.T., Ma, J. and Zhang, L. (2019). CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality. *Ann. Statist.* **47** 1234–1267. MR3911111 https://doi.org/10.1214/18-AOS1711
[15] Cape, J., Tang, M. and Priebe, C.E. (2019). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070
[16] Capitaine, M., Donati-Martin, C. and Féral, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394
[17] Chen, X., Ponomareva, M. and Tamer, E. (2014). Likelihood inference in some finite mixture models. *J. Econometrics* **182** 87–99. MR3212763 https://doi.org/10.1016/j.jeconom.2014.04.010
[18] Chen, Y., Chi, Y., Fan, J., Ma, C. et al. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806.
[19] Couillet, R. and Benaych-Georges, F. (2016). Kernel spectral clustering of large dimensional data. *Electron. J. Stat.* **10** 1393–1454. MR3507369 https://doi.org/10.1214/16-EJS1144
[20] David, P.H. (2020). Degrees of freedom and model selection for k-means. *Comput. Statist. Data Anal.* **149** 1–14.
[21] 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
[22] Ding, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli* **26** 387–417. MR4036038 https://doi.org/10.3150/19-BEJ1129
[23] Ding, X. and Yang, F. (2021). Spiked separable covariance matrices and principal components. *Ann. Statist.* **49** 1113–1138. MR4255121 https://doi.org/10.1214/20-aos1995
[24] Donoho, D. and Gavish, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257
[25] Duda, R.O. and Hart, P.E. (1973). *Pattern Classification and Scene Analysis* **3**. New York: Wiley.
[26] 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
[27] 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
[28] Fan, J., Fan, Y., Han, X. and Lv, J. (2022). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *J. Amer. Statist. Assoc.* **117** 996–1009. MR4436328 https://doi.org/10.1080/01621459.2020.1840990
[29] Hachem, W., Loubaton, P. and Najim, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235 https://doi.org/10.1214/105051606000000925
[30] Hachem, W., Loubaton, P., Najim, J. and Vallet, P. (2013). On bilinear forms based on the resolvent of large random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 36–63. MR3060147 https://doi.org/10.1214/11-AIHP450
[31] Han, X., Tong, X. and Fan, Y. (2023). Eigen selection in spectral clustering: A theory-guided practice. *J. Amer. Statist. Assoc.* **118** 109–121. MR4571110 https://doi.org/10.1080/01621459.2021.1917418

[32] Huang, J. (2018). Mesoscopic perturbations of large random matrices. *Random Matrices Theory Appl.* **7** 1850004. MR3786885 https://doi.org/10.1142/S2010326318500041

[33] Ipsen, I.C.F. and Nadler, B. (2009). Refined perturbation bounds for eigenvalues of Hermitian and non-Hermitian matrices. *SIAM J. Matrix Anal. Appl.* **31** 40–53. MR2487048 https://doi.org/10.1137/070682745

[34] Jiang, D. and Bai, Z. (2021). Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices. *Bernoulli* **27** 274–294. MR4177370 https://doi.org/10.3150/20-BEJ1237

[35] Jin, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

[36] 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

[37] Keane, M.P. and Wolpin, K.I. (1997). The career decisions of young men. *J. Polit. Econ.* **105** 473–522.

[38] Knowles, A. and Yin, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. MR3103909 https://doi.org/10.1002/cpa.21450

[39] 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

[40] Liu, X., Liu, Y., Pan, G., Zhang, L. and Zhang, Z. (2025). Supplement to "Asymptotic limits of spiked eigenvalues and eigenvectors of signal-plus-noise matrices with weak signals and heteroskedastic noise." https://doi.org/10.3150/24-BEJ1808SUPP

[41] Liu, Y., Liang, Y.-C., Pan, G. and Zhang, Z. (2023). Random or nonrandom signal in high-dimensional regimes. *IEEE Trans. Inf. Theory* **69** 298–315. MR4544962 https://doi.org/10.1109/tit.2022.3204355

[42] Löffler, M., Zhang, A.Y. and Zhou, H.H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *Ann. Statist.* **49** 2506–2530. MR4338373 https://doi.org/10.1214/20-aos2044

[43] Loubaton, P. and Vallet, P. (2011). Almost sure localization of the eigenvalues in a Gaussian information plus noise model—application to the spiked models. *Electron. J. Probab.* **16** 1934–1959. MR2851051 https://doi.org/10.1214/EJP.v16-943

[44] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proc. Fifth Berkeley Sympos. Math. Statist. And Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics* 281–297. Berkeley, CA: Univ. California Press. MR0214227

[45] Maimon, O. and Rokach, L. (2005). *Data Mining and Knowledge Discovery Handbook* **2**. Berlin: Springer.

[46] Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inf. Theory* **54** 5113–5129. MR2589886 https://doi.org/10.1109/TIT.2008.929938

[47] Mestre, X. (2008). On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. *IEEE Trans. Signal Process.* **56** 5353–5368. MR2472837 https://doi.org/10.1109/TSP.2008.929662

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

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

[50] Rdusseeun, L. and Kaufman, P. (1987). Clustering by means of medoids. In *Proceedings of the Statistical Data Analysis Based on the L1 Norm Conference, Neuchatel, Switzerland* **31**.

[51] Redner, R.A. and Walker, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26** 195–239. MR0738930 https://doi.org/10.1137/1026034

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

[53] Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20** 53–65.

[54] Shabalin, A.A. and Nobel, A.B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *J. Multivariate Anal.* **118** 67–76. MR3054091 https://doi.org/10.1016/j.jmva.2013.03.005

[55] Silverstein, J.W. and Bai, Z.D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534 https://doi.org/10.1006/jmva.1995.1051
[56] Thorndike, R.L. (1953). Who belongs in the family? *Psychometrika* **18** 267–276.
[57] Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503 https://doi.org/10.1111/1467-9868.00293
[58] Vallet, P., Loubaton, P. and Mestre, X. (2012). Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case. *IEEE Trans. Inf. Theory* **58** 1043–1068. MR2918009 https://doi.org/10.1109/TIT.2011.2173718
[59] Yang, D., Ma, Z. and Buja, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *J. Mach. Learn. Res.* **17** 92. MR3543498
[60] 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
[61] Zhang, Z., Zheng, S., Pan, G. and Zhong, P.-S. (2022). Asymptotic independence of spiked eigenvalues and linear spectral statistics for large sample covariance matrices. *Ann. Statist.* **50** 2205–2230. MR4474488 https://doi.org/10.1214/22-aos2183
[62] Zhou, Z. and Amini, A.A. (2019). Analysis of spectral clustering algorithms for community detection: The general bipartite setting. *J. Mach. Learn. Res.* **20** 47. MR3948087

Bernoulli **31**(3), 2025, 2377–2402
https://doi.org/10.3150/24-BEJ1809

Covariance operator estimation: Sparsity, lengthscale, and ensemble Kalman filters

OMAR AL-GHATTAS[1,a], JIAHENG CHEN[2,c], DANIEL SANZ-ALONSO[1,b] and NATHAN WANIOREK[2,d]

[1]*Department of Statistics, University of Chicago, Chicago, USA*, [a]*ghattas@uchicago.edu*, [b]*sanzalonso@uchicago.edu*

[2]*Committee on Computational and Applied Mathematics, University of Chicago, Chicago, USA*, [c]*jiaheng@uchicago.edu*, [d]*waniorek@uchicago.edu*

This paper investigates covariance operator estimation via thresholding. For Gaussian random fields with approximately sparse covariance operators, we establish non-asymptotic bounds on the estimation error in terms of the sparsity level of the covariance and the expected supremum of the field. We prove that thresholded estimators enjoy an exponential improvement in sample complexity compared with the standard sample covariance estimator if the field has a small correlation lengthscale. As an application of the theory, we study thresholded estimation of covariance operators within ensemble Kalman filters.

Keywords: Covariance operator estimation; ensemble Kalman filters; small lengthscale regime; thresholding

References

[1] Abdalla, P. and Zhivotovskiy, N. (2024). Covariance estimation: Optimal dimension-free guarantees for adversarial corruption and heavy tails. *J. Eur. Math. Soc.* https://doi.org/10.4171/JEMS/1505

[2] Al-Ghattas, O., Bao, J. and Sanz-Alonso, D. (2024). Ensemble Kalman filters with resampling. *SIAM/ASA J. Uncertain. Quantificat.* **12** 411–441. https://doi.org/10.1137/23M1594935

[3] Al-Ghattas, O., Chen, J., Sanz-Alonso, D. and Waniorek, N. (2025). Supplement to "Covariance operator estimation: sparsity, lengthscale, and ensemble Kalman filters." https://doi.org/10.3150/24-BEJ1809SUPP

[4] Al-Ghattas, O. and Sanz-Alonso, D. (2024). Non-asymptotic analysis of ensemble Kalman updates: Effective dimension and localization. *Inf. Inference* **13** iaad043. https://doi.org/10.1093/imaiai/iaad043

[5] Bai, Z.D. and Yin, Y.Q. (2008). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. In *Advances in Statistics* 108–127. Singapore: World Scientific. https://doi.org/10.1214/aop/1176989118

[6] Bellec, P.C., Lecué, G. and Tsybakov, A.B. (2018). Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. https://doi.org/10.1214/17-AOS1670

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

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

[9] Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. https://doi.org/10.1214/08-AOS620

[10] Bigoni, D., Chen, Y., Garcia Trillos, N., Marzouk, Y. and Sanz-Alonso, D. (2020). Data-driven forward discretizations for Bayesian inversion. *Inverse Probl.* **36** 105008. https://doi.org/10.1088/1361-6420/abb2fa

[11] 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. https://doi.org/10.1137/12089586X

[12] Burgers, G., Jan van Leeuwen, P. and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126** 1719–1724. https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2
[13] Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. https://doi.org/10.1198/jasa.2011.tm10560
[14] Cai, T.T. and Yuan, M. (2012). Adaptive covariance matrix estimation through block thresholding. *Ann. Statist.* **40** 2014–2042. https://doi.org/10.1214/12-AOS999
[15] Cai, T.T. and Zhou, H.H. (2012). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statist. Sinica* 1319–1349. https://doi.org/10.5705/ss.2010.253
[16] Cai, T.T. and Zhou, H.H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. https://doi.org/10.1214/12-AOS998
[17] Calvello, E., Reich, S. and Stuart, A.M. (2022). Ensemble Kalman methods: A mean field perspective. arXiv preprint. arXiv:2209.11371.
[18] Carrassi, A., Bocquet, M., Bertino, L. and Geir, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev.: Clim. Change* **9**. https://doi.org/10.1002/wcc.535
[19] Chen, J. and Stein, M.L. (2023). Linear-cost covariance functions for Gaussian random fields. *J. Amer. Statist. Assoc.* **118** 147–164. https://doi.org/10.1080/01621459.2021.1919122
[20] Chen, R.Y., Gittens, A. and Tropp, J.A. (2012). The masked sample covariance estimator: An analysis using matrix concentration inequalities. *Inf. Inference* **1** 2–20. https://doi.org/10.1093/imaiai/ias001
[21] Chen, Y. and Anitescu, M. (2023). Scalable physics-based maximum likelihood estimation using hierarchical matrices. *SIAM/ASA J. Uncertain. Quantificat.* **11** 682–725. https://doi.org/10.1137/21M1458880
[22] Chen, Y., Sanz-Alonso, D. and Willett, R. (2022). Autodifferentiable ensemble Kalman filters. *SIAM J. Math. Data Sci.* **4** 801–833. https://doi.org/10.1137/21M143447
[23] de Hoop, M.V., Kovachki, N.B., Nelsen, N.H. and Stuart, A.M. (2023). Convergence rates for learning linear operators from noisy data. *SIAM/ASA J. Uncertain. Quantificat.* **11** 480–513. https://doi.org/10.1137/21M144294
[24] Diakonikolas, I. and Kane, D.M. (2023). *Algorithmic High-Dimensional Robust Statistics*. Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/9781108943161
[25] El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. https://doi.org/10.1214/07-AOS559
[26] Evensen, G. (2009). *Data Assimilation: The Ensemble Kalman Filter*. Berlin: Springer.
[27] Fang, Q., Guo, S. and Qiao, X. (2023). Adaptive functional thresholding for sparse covariance function estimation in high dimensions. *J. Amer. Statist. Assoc.* **546** 1473–1485. https://doi.org/10.1080/01621459.2023.2200522
[28] Farchi, A. and Bocquet, M. (2019). COn the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Front. Appl. Math. Stat.* **3**. https://doi.org/10.3389/fams.2019.00003
[29] Fernique, X.M., Conze, J., Gani, J. and Fernique, X. (1975). *Regularité des trajectoires des fonctions aléatoires gaussiennes*. Berlin: Springer. https://doi.org/10.1007/BFb0080190
[30] Furrer, R., Genton, M.G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15**. 502–523. https://doi.org/10.1198/106186006X132178
[31] Ghosal, S. and van der Vaart, A.W. (2017). *Fundamentals of Nonparametric Bayesian Inference* **44**. Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/9781139029834
[32] Goes, J., Lerman, G. and Nadler, B. (2020). Robust sparse covariance estimation by thresholding Tyler's M-estimator. *Ann. Statist.* **48** 86–110. https://doi.org/10.1214/18-AOS1793
[33] Gordon, Y. (1985). Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** 265–289. https://doi.org/10.1007/BF02759761
[34] Herty, M. and Visconti, G. (2019). Kinetic methods for inverse problems. *Kinet. Relat. Models* **12** 1109. https://doi.org/10.3934/krm.2019042
[35] Houtekamer, P.L. and Mitchell, H.L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**. 123–137. https://doi.org/10.1175/1520-0493(2001)129%3C0123:ASEKFF%3E2.0.CO; 2
[36] Houtekamer, P.L. and Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **144** 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1

[37] Hunter, J.K. and Nachtergaele, B. (2001). *Applied Analysis*. Singapore: World Scientific. https://doi.org/10.1142/4319
[38] James, G.M., Hastie, T.J. and Sugar, C.A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602. https://doi.org/10.1093/biomet/87.3.587
[39] James, G.M. and Sugar, C.A. (2003). Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* **98** 397–408. https://doi.org/10.1198/016214503000189
[40] Jin, J., Lu, Y., Blanchet, J. and Ying, L. (2022). Minimax optimal kernel operator learning via multilevel training. In *The Eleventh International Conference on Learning Representations*.
[41] Kalnay, E. (2003). *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511802270
[42] Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. https://doi.org/10.3150/15-BEJ730
[43] Kwiatkowski, E. and Mandel, J. (2015). Convergence of the square root ensemble Kalman filter in the large ensemble limit. *SIAM/ASA J. Uncertain. Quantificat.* **3** 1–17. https://doi.org/10.1137/140965363
[44] Mendelson, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680. https://doi.org/10.1016/j.spa.2016.04.019
[45] Mollenhauer, M., Mücke, N. and Sullivan, T. (2022). Learning linear operators: Infinite-dimensional regression as a well-behaved non-compact inverse problem. arXiv preprint. arXiv:2211.08875.
[46] Nobile, F. and Tesei, F. (2015). A multi level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients. *Stoch. Partial Differ. Equ. Anal. Comput.* **3** 398–444. https://doi.org/10.1007/s40072-015-0055-9
[47] Qiao, X., Qian, C., James, G.M. and Guo, S. (2020). Doubly functional graphical models in high dimensions. *Biometrika* **107** 415–431. https://doi.org/10.1093/biomet/asz072
[48] Sanz-Alonso, D., Stuart, A. and Taeb, A. (2023). *Inverse Problems and Data Assimilation* **7**. Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/9781009414319
[49] Sanz-Alonso, D. and Waniorek, N. (2024). Analysis of a computational framework for Bayesian inverse problems: Ensemble Kalman updates and MAP estimators under mesh refinement. *SIAM/ASA J. Uncertain. Quantificat.* **12** 30–68. https://doi.org/10.1137/23M1567035
[50] Sanz-Alonso, D. and Yang, R. (2022). Unlabeled data help in graph-based semi-supervised learning: A Bayesian nonparametrics perspective. *J. Mach. Learn. Res.* **23** 1–28.
[51] Stein, M.L. (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. Berlin: Springer. https://doi.org/10.1007/978-1-4612-1494-6
[52] Stuart, A.M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. https://doi.org/10.1017/S0962492910000061
[53] Talagrand, M. (2022). *Upper and Lower Bounds for Stochastic Processes: Decomposition Theorems* **60**. Berlin: Springer. https://doi.org/10.1007/978-3-030-82595-9
[54] Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M. and Whitaker, J.S. (2003). Ensemble square root filters. *Mon. Weather Rev.* **131** 1485–1490. https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2
[55] Tong, X.T. (2018). Performance analysis of local ensemble Kalman filter. *J. Nonlinear Sci.* **28** 1397–1442. https://doi.org/10.1007/s00332-018-9453-2
[56] Tong, X.T. and Morzfeld, M. (2023). Localized ensemble Kalman inversion. *Inverse Probl.* **39** 064002. https://doi.org/10.1088/1361-6420/accb08
[57] Van Handel, R. (2014). Probability in high dimension. Technical Report, Princeton University.
[58] Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. arXiv preprint. arXiv:1011.3027.
[59] Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science* **47**. Cambridge: Cambridge Univ. Press. https://doi.org/10.1017/9781108231596
[60] Wainwright, M.J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint* **48**. Cambridge: Cambridge Univ. Press.
[61] Wendland, H. (2004). *Scattered Data Approximation* **17**. Cambridge: Cambridge Univ. Press.
[62] Williams, C.K.I. and Rasmussen, C.E. (2006). *Gaussian Processes for Machine Learning* **2**. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001

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

[64] Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. https://doi.org/10.1198/016214504000001745

[65] Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. https://doi.org/10.1214/009053605000000660

Bernoulli **31**(3), 2025, 2403–2433
https://doi.org/10.3150/24-BEJ1810

Posterior consistency in multi-response regression models with non-informative priors for the error covariance matrix in growing dimensions

PARTHA SARKAR[a], KSHITIJ KHARE[b] and MALAY GHOSH[c]

Department of Statistics, University of Florida, Gainesville, United States, [a]*sarkarpartha@ufl.edu,* [b]*kdkhare@stat.ufl.edu,* [c]*ghoshm@ufl.edu*

The Inverse-Wishart (IW) distribution is a standard and popular choice of priors for covariance matrices and has attractive properties such as conditional conjugacy. However, the IW family of priors has crucial drawbacks, including the lack of effective choices for non-informative priors. Several classes of priors for covariance matrices that alleviate these drawbacks, while preserving computational tractability, have been proposed in the literature. These priors can be obtained through appropriate scale mixtures of IW priors. However, in the era of increasing dimensionality, the posterior consistency of models that incorporate such priors has not been investigated. We address this issue for the multi-response regression setting (q responses, n samples) under a wide variety of IW scale mixture priors for the error covariance matrix. Posterior consistency and contraction rates for both the regression coefficient matrix and the error covariance matrix are established in the "large q, large n" setting under mild assumptions on the true data-generating covariance matrix and relevant hyperparameters. In particular, the number of responses q_n is allowed to grow with n, but with $q_n = o(n)$. Also, some results related to the inconsistency of the posterior distribution and posterior mean for $q_n/n \to \gamma$, where $\gamma \in (0, \infty)$ are provided.

Keywords: High-dimensional covariance estimation; non-informative prior; scale-mixed inverse Wishart

References

Bai, R. and Ghosh, M. (2018). High-dimensional multivariate posterior consistency under global-local shrinkage priors. *J. Multivariate Anal.* **167** 157–170. MR3830639 https://doi.org/10.1016/j.jmva.2018.04.010

Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

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

Banerjee, S. and Ghosal, S. (2015). Bayesian structure learning in graphical models. *J. Multivariate Anal.* **136** 147–162. MR3321485 https://doi.org/10.1016/j.jmva.2015.01.015

Barnard, J., McCulloch, R. and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statist. Sinica* **10** 1281–1311. MR1804544

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

Bouriga, M. and Féron, O. (2013). Estimation of covariance matrices based on hierarchical inverse-Wishart priors. *J. Statist. Plann. Inference* **143** 795–808. MR3003891 https://doi.org/10.1016/j.jspi.2012.09.006

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

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

1350-7265

Daniels, M.J. and Kass, R.E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Amer. Statist. Assoc.* **94** 1254–1263. MR1731487 https://doi.org/10.2307/2669939

Eaton, M.L. (2007). Matrix factorizations and Jacobians. In *Multivariate Stat.* **53** 159–184. IMS.

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

Gao, C. and Zhou, H.H. (2016). Bernstein–von Mises theorems for functionals of the covariance matrix. *Electron. J. Stat.* **10** 1751–1806. MR3522660 https://doi.org/10.1214/15-EJS1048

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–533. MR2221284 https://doi.org/10.1214/06-BA117A

Gelman, A. Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. *Analytical Methods for Social Research*. Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511790942

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. Boca Raton, FL: CRC Press. MR3235677

Huang, A. and Wand, M.P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.* **8** 439–451. MR3066948 https://doi.org/10.1214/13-BA815

Jones, G.L. (2004). On the Markov chain central limit theorem. *Probab. Surv.* **1** 299–320. MR2068475 https://doi.org/10.1214/154957804100000051

Kass, R.E. and Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 535–542. MR2221285 https://doi.org/10.1214/06-BA117B

Leonard, T. and Hsu, J.S.J. (1992). Bayesian inference for a covariance matrix. *Ann. Statist.* **20** 1669–1696. MR1193308 https://doi.org/10.1214/aos/1176348885

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

Mulder, J. and Pericchi, L.R. (2018). The matrix-F prior for estimating and testing covariance matrices. *Bayesian Anal.* **13** 1189–1210. MR3855368 https://doi.org/10.1214/17-BA1092

Narisetty, N.N. and He, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. MR3210987 https://doi.org/10.1214/14-AOS1207

O'Malley, A.J. and Zaslavsky, A.M. (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *J. Amer. Statist. Assoc.* **103** 1405–1418. MR2655721 https://doi.org/10.1198/016214508000000724

Roberts, G.O. and Rosenthal, J.S. (2001). Markov chains and de-initializing processes. *Scand. J. Statist.* **28** 489–504. MR1858413 https://doi.org/10.1111/1467-9469.00250

Sarkar, P., Khare, K. and Ghosh, M. (2025). Supplement to "Posterior consistency in multi-response regression models with non-informative priors for the error covariance matrix in growing dimensions." https://doi.org/10.3150/24-BEJ1810SUPP

Song, Q. and Liang, F. (2023). Nearly optimal Bayesian shrinkage for high-dimensional regression. *Sci. China Math.* **66** 409–442. MR4535982 https://doi.org/10.1007/s11425-020-1912-6

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A. and Tuerlinckx, F. (2011). Visualizing distributions of covariance matrices. Technical Report 18–18, Columbia Univ. New York, USA..

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

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

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

Wang, S.-H., Bai, R. and Huang, H.-H. (2024). Two-Step Mixed-Type Multivariate Bayesian Sparse Variable Selection with Shrinkage Priors.

Xiang, R., Khare, K. and Ghosh, M. (2015). High dimensional posterior convergence rates for decomposable graphical models. *Electron. J. Stat.* **9** 2828–2854. MR3439186 https://doi.org/10.1214/15-EJS1084

Zhang, R., Yao, Y. and Ghosh, M. (2022). Contraction of a quasi-Bayesian model with shrinkage priors in precision matrix estimation. *J. Statist. Plann. Inference* **221** 154–171. MR4420542 https://doi.org/10.1016/j.jspi.2022.03.003

Bernoulli **31**(3), 2025, 2434–2458
https://doi.org/10.3150/24-BEJ1811

F statistics for high-dimensional inference of linear model

YUMOU QIU[1,a] and YUSHAN GU[2,b]

[1] *School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing, 100871, China,* [a] *qiuyumou@math.pku.edu.cn*
[2] *Department of Statistics, Iowa State University, Ames, IA 50011, USA,* [b] *yushang@iastate.edu*

We aim to estimate and conduct inference for the effects of multiple covariates of interest simultaneously after adjusting the impact of high-dimensional control variables under a linear model. An F-type statistic is proposed based on the residuals from the regularized fitting of the response variable and the target covariates on the control covariates. A testing procedure and confidence interval are constructed. The proposed procedures reduce the impact of potential over-fitting errors from the regularized regression on the inference of the target parameters. The essence is to eliminate the prediction errors in the direction of the actual regression error and achieve a more accurate size and coverage rate. Expansions of the proposed statistics are derived, showing the proposed method's error reduction property. Simulation studies verify the theoretical results and demonstrate the proposed method has better performance than the existing methods. A real data analysis for S&P 500 stock returns is conducted to show the utility of the proposed method in practice.

Keywords: Error reduction; F statistic; high dimensionality; inference for penalized regression; lasso

References

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

Belloni, A., Chernozhukov, V. and Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. *J. Bus. Econom. Statist.* **34** 606–619. MR3547999 https://doi.org/10.1080/07350015.2016.1166116

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

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

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

Cai, T.T., Liu, W. and Zhou, H.H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **44** 455–488. MR3476606 https://doi.org/10.1214/13-AOS1171

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

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

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

Chetverikov, D., Liao, Z. and Chernozhukov, V. (2021). On cross-validated Lasso in high dimensions. *Ann. Statist.* **49** 1300–1317. MR4298865 https://doi.org/10.1214/20-aos2000

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

1350-7265

Fei, Z., Zhu, J., Banerjee, M. and Li, Y. (2019). Drawing inferences for high-dimensional linear models: A selection-assisted partial regression and smoothing approach. *Biometrics* **75** 551–561. MR3999178 https://doi.org/10.1111/biom.13013

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

Lee, J.D., Sun, D.L., Sun, Y. and Taylor, J.E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523 https://doi.org/10.1111/j.1467-9868.2010.00740.x

Meinshausen, N., Meier, L. and Bühlmann, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584 https://doi.org/10.1198/jasa.2009.tm08647

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

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

Qiu, Y. and Gu, Y. (2025). Supplement to "F statistics for high-dimensional inference of linear model." https://doi.org/10.3150/24-BEJ1811SUPP

Qiu, Y. and Zhou, X.-H. (2020). Estimating c-level partial correlation graphs with application to brain imaging. *Biostatistics* **21** 641–658. MR4164049 https://doi.org/10.1093/biostatistics/kxy076

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

Robinson, P.M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762 https://doi.org/10.2307/1912705

Shah, R.D. and Bühlmann, P. (2018). Goodness-of-fit tests for high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 113–135. MR3744714 https://doi.org/10.1111/rssb.12234

Shah, R.D. and Samworth, R.J. (2013). Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 55–80. MR3008271 https://doi.org/10.1111/j.1467-9868.2011.01034.x

Shi, C., Song, R., Chen, Z. and Li, R. (2019). Linear hypothesis testing for high dimensional generalized linear models. *Ann. Statist.* **47** 2671–2703. MR3988769 https://doi.org/10.1214/18-AOS1761

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

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

Wang, J., He, X. and Xu, G. (2020). Debiased inference on treatment effect in a high-dimensional model. *J. Amer. Statist. Assoc.* **115** 442–454. MR4078474 https://doi.org/10.1080/01621459.2018.1558062

Wong, K.C., Li, Z. and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Ann. Statist.* **48** 1124–1142. MR4102690 https://doi.org/10.1214/19-AOS1840

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

Zhang, X. and Cheng, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

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

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

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

Zhu, Y. and Bradic, J. (2018b). Significance testing in non-sparse high-dimensional linear models. *Electron. J. Stat.* **12** 3312–3364. MR3861831 https://doi.org/10.1214/18-EJS1443

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

Bernoulli **31**(3), 2025, 2459–2483
https://doi.org/10.3150/24-BEJ1812

Inference for change-plane regression

CHAERYON KANG[1,a], HUNYONG CHO[2,b], RUI SONG[3,d], MOULINATH BANERJEE[4,e], ERIC B. LABER[5,f] and MICHAEL R. KOSOROK[2,c]

[1]*University of Pittsburgh, Pittsburgh, USA,* [a]*crkang@pitt.edu*
[2]*University of North Carolina at Chapel Hill, Chapel Hill, USA,* [b]*hunycho@live.unc.edu,* [c]*kosorok@unc.edu*
[3]*North Carolina State University, Raleigh, USA,* [d]*rsong@ncsu.edu*
[4]*University of Michigan, Ann Arbor, USA,* [e]*moulib@umich.edu*
[5]*Duke University, Durham, USA,* [f]*eric.laber@duke.edu*

A key challenge in analyzing the behavior of change-plane estimators is that the objective function has multiple minimizers. Two estimators are proposed to deal with this non-uniqueness. For each estimator, an n-rate of convergence is established, and the limiting distribution is derived. Based on these results, we provide a parametric bootstrap procedure for inference. The validity of our theoretical results and the finite sample performance of the bootstrap are demonstrated through simulation experiments. We demonstrate the proposed methods by applying them to latent subgroup identification in precision medicine using the ACTG175 AIDS study data.

Keywords: Change-plane; latent subgroup analysis; M-estimation; precision medicine

References

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

Chan, K.S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21** 520–533. MR1212191 https://doi.org/10.1214/aos/1176349040

Chan, N.H., Ng, W.L., Yau, C.Y. and Yu, H. (2021). Optimal change-point estimation in time series. *Ann. Statist.* **49** 2336–2355. https://doi.org/10.1214/20-AOS2039

Chernoff, H. and Rubin, H. (1956). The estimation of the location of a discontinuity in density. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 19–37. Berkeley-Los Angeles, CA: Univ. California Press. MR0084910

Deng, Y., Cai, J. and Zeng, D. (2022). Maximum likelihood estimation for Cox proportional hazards model with a change hyperplane. *Statist. Sinica* **32** 983–1006. MR4405106 https://doi.org/10.5705/ss.202020.0141

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

Fan, A., Song, R. and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. *J. Amer. Statist. Assoc.* **112** 769–778. MR3671769 https://doi.org/10.1080/01621459.2016.1166115

Ferger, D. (2004). A continuous mapping theorem for the argmax-functional in the non-unique case. *Stat. Neerl.* **58** 83–96. MR2042258 https://doi.org/10.1046/j.0039-0402.2003.00111.x

Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51** 231–251. MR1707773 https://doi.org/10.1023/A:1003802007064

Hammer, S.M., Squires, K.E., Hughes, M.D., Grimes, J.M., Demeter, L.M., Currier, J.S., Eron, J.J. Jr., Feinberg, J.E., Balfour, H.H. Jr., Deyton, L.R., Chodakewitz, J.A. and Fischl, M.A. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N. Engl. J. Med.* **337** 725–733. https://doi.org/10.1056/NEJM199709113371101

Hansen, B.E. (2000). Sample splitting and threshold estimation. *Econometrica* **68** 575–603. MR1769379 https://doi.org/10.1111/1468-0262.00124

Hernán, M.A. and Robins, J.M. (2020). *Causal Inference: What If.* Boca Raton, FL: CRC Press/CRC.

1350-7265

Huang, Y., Cho, J. and Fong, Y. (2021). Threshold-based subgroup testing in logistic regression models in two-phase sampling designs. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **70** 291–311. MR4226669 https://doi.org/10.1111/rssc.12459

Jin, P., Lu, W., Chen, Y. and Liu, M. (2023). Change-plane analysis for subgroup detection with a continuous treatment. *Biometrics* **79** 1920–1933. MR4643966 https://doi.org/10.1111/biom.13762

Kang, C. (2011). New statistical learning methods for chemical toxicity data analysis. Ph.D. thesis, Univ. North Carolina at Chapel Hill. https://doi.org/10.17615/5b2q-pr45

Kang, C., Cho, H., Song, R., Banerjee, M., Laber, E.B. and Kosorok, M.R. (2025). Supplement to "Inference for change-plane regression." https://doi.org/10.3150/24-BEJ1812SUPP

Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. New York: Springer. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

Kosorok, M.R. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.* **35** 957–989. MR2341694 https://doi.org/10.1214/009053606000001244

Lan, Y., Banerjee, M. and Michailidis, G. (2009). Change-point estimation under adaptive sampling. *Ann. Statist.* **37** 1752–1791. MR2533471 https://doi.org/10.1214/08-AOS602

Lee, S., Liao, Y., Seo, M.H. and Shin, Y. (2021). Factor-driven two-regime regression. *Ann. Statist.* **49** 1656–1678. MR4298876 https://doi.org/10.1214/20-aos2017

Li, J., Li, Y., Jin, B. and Kosorok, M.R. (2021). Multithreshold change plane model: Estimation theory and applications in subgroup identification. *Stat. Med.* **40** 3440–3459. MR4269063 https://doi.org/10.1002/sim.8976

Luo, X. (1996). The asymptotic distribution of MLE of treatment lag threshold. *J. Statist. Plann. Inference* **53** 33–61. MR1405736 https://doi.org/10.1016/0378-3758(95)00142-5

Mukherjee, D., Banerjee, M., Mukherjee, D. and Ritov, Y. (2023). Asymptotic normality of a change plane estimator in fixed dimension with near-optimal rate. *Electron. J. Stat.* **17** 2289–2316. MR4649982 https://doi.org/10.1214/23-ejs2144

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann. Statist.* **31** 442–463.

Ross, S.M. (1996). *Stochastic Processes*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. MR1373653

Rubin, D.B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100** 322–331. MR2166071 https://doi.org/10.1198/016214504000001880

Seijo, E. and Sen, B. (2011). A continuous mapping theorem for the smallest argmax functional. *Electron. J. Stat.* **5** 421–439. MR2802050 https://doi.org/10.1214/11-EJS613

Seo, M.H. and Linton, O. (2007). A smoothed least squares estimator for threshold regression models. *J. Econometrics* **141** 704–735. MR2413485 https://doi.org/10.1016/j.jeconom.2006.11.002

Song, R., Banerjee, M. and Kosorok, M.R. (2016). Asymptotics for change-point models under varying degrees of mis-specification. *Ann. Statist.* **44** 153–182. MR3449765 https://doi.org/10.1214/15-AOS1362

Verzelen, N., Fromont, M., Lerasle, M. and Reynaud-Bouret, P. (2023). Optimal change-point detection and localization. *Ann. Statist.* **51** 1586–1610. MR4658569 https://doi.org/10.1214/23-aos2297

Wang, J. and Li, J. (2023). Multi-threshold structural equation model. *J. Bus. Econom. Statist.* **41** 377–387. MR4568029 https://doi.org/10.1080/07350015.2021.2023553

Wei, S. and Kosorok, M.R. (2013). Latent supervised learning. *J. Amer. Statist. Assoc.* **108** 957–970. MR3174676 https://doi.org/10.1080/01621459.2013.789695

Wei, K., Qin, G. and Zhu, Z. (2023). Subgroup analysis for longitudinal data based on a partial linear varying coefficient model with a change plane. *Stat. Med.* **42** 3716–3731. MR4629071 https://doi.org/10.1002/sim.9827

Bernoulli **31**(3), 2025, 2484–2507
https://doi.org/10.3150/24-BEJ1814

Sandpiles on the Vicsek fractal explode with probability $\frac{1}{4}$

NICO HEIZMANN[1,a], ROBIN KAISER[2,b] and ECATERINA SAVA-HUSS[2,c]

[1] *Technische Universität Chemnitz, Chemnitz, Germany,* [a]*nico.heizmann@math.tu-chemnitz.de*
[2] *Universität Innsbruck, Innsbruck, Austria,* [b]*robin.kaiser@uibk.ac.at,* [c]*ecaterina.sava-huss@uibk.ac.at*

Vicsek fractal graphs are an important class of infinite graphs with self similar properties, polynomial growth and treelike features, on which several dynamical processes such as random walks or Abelian sandpiles can be rigorously analyzed and one can obtain explicit closed form expressions. While such processes on Vicsek fractals and on Euclidean lattices $\mathbb{Z}^2$ share some properties for instance in the recurrence behavior, many quantities related to sandpiles on Euclidean lattices are still poorly understood. The current work focuses on the stabilization and explosion of Abelian sandpiles on Vicsek fractal graphs, and we prove that a sandpile sampled from the infinite volume limit plus one additional particle stabilizes with probability $3/4$, that is, it does not stabilize almost surely and it explodes with the complementary probability $1/4$. We prove the main result by using two different approaches: one of probabilistic nature and one of algebraic flavor. The first approach is based on investigating the particles sent to the boundary of finite volumes and showing that their number stays above four with positive probability. In the second approach we relate the question of stabilization and explosion of sandpiles in infinite volume to the order of elements of the sandpile group on finite approximations of the infinite Vicsek graph. The method applies to more general state spaces and by employing it we also find all invariant factors of the sandpile groups on the finite approximations of the infinite Vicsek fractal.

Keywords: Abelian sandpile; absorbing states; critical group; infinite volume limit; invariant factors; Markov chains; stabilization; toppling; uniform spanning trees; Vicsek fractal

References

Athreya, S.R. and Járai, A.A. (2004). Infinite volume limit for the stationary distribution of Abelian sandpile models. *Comm. Math. Phys.* **249** 197–213. MR2077255 https://doi.org/10.1007/s00220-004-1080-0

Bak, P., Tang, C. and Wiesenfeld, K. (1988). Self-organized criticality. *Phys. Rev. A (3)* **38** 364–374. MR0949160 https://doi.org/10.1103/PhysRevA.38.364

Benjamini, I., Lyons, R., Peres, Y. and Schramm, O. (2001). Uniform spanning forests. *Ann. Probab.* **29** 1–65. MR1825141 https://doi.org/10.1214/aop/1008956321

Bhupatiraju, S., Hanson, J. and Járai, A.A. (2017). Inequalities for critical exponents in d-dimensional sandpiles. *Electron. J. Probab.* **22** Paper No. 85, 51. MR3718713 https://doi.org/10.1214/17-EJP111

Bou-Rabee, A. (2024). Integer superharmonic matrices on the F-lattice. *Adv. Math.* **436** Paper No. 109400, 98. MR4669323 https://doi.org/10.1016/j.aim.2023.109400

Caracciolo, S., Paoletti, G. and Sportiello, A. (2008). Explicit characterization of the identity configuration in an Abelian sandpile model. *J. Phys. A* **41** 495003, 17. MR2515893 https://doi.org/10.1088/1751-8113/41/49/495003

Daerden, F., Priezzhev, V.B. and Vanderzande, C. (2001). Waves in the sandpile model on fractal lattices. *Phys. A* **292** 43–54. MR1822430 https://doi.org/10.1016/S0378-4371(00)00553-7

Dhar, D. (1990). Self-organized critical state of sandpile automaton models. *Phys. Rev. Lett.* **64** 1613–1616. MR1044086 https://doi.org/10.1103/PhysRevLett.64.1613

Dhar, D. and Majumdar, S.N. (1990). Abelian sandpile model on the Bethe lattice. *J. Phys. A* **23** 4333–4350. MR1076905 https://doi.org/10.1088/0305-4470/23/19/018

Dhar, D., Ruelle, P., Sen, S. and Verma, D.-N. (1995). Algebraic aspects of Abelian sandpile models. *J. Phys. A* **28** 805–831. MR1326322

Heizmann, N., Kaiser, R. and Sava-Huss, E. (2023). Height probabilities for Abelian sandpiles and the looping constant on Sierpinski graphs. Available at arXiv:2308.03445.

Holroyd, A.E., Levine, L., Mészáros, K., Peres, Y., Propp, J. and Wilson, D.B. (2008). Chip-firing and rotor-routing on directed graphs. In *In and Out of Equilibrium. 2*. *Progress in Probability* **60** 331–364. Basel: Birkhäuser. MR2477390 https://doi.org/10.1007/978-3-7643-8786-0_17

Hutchcroft, T. (2020). Universality of high-dimensional spanning forests and sandpiles. *Probab. Theory Related Fields* **176** 533–597. MR4055195 https://doi.org/10.1007/s00440-019-00923-3

Járai, A.A. (2018). Sandpile models. *Probab. Surv.* **15** 243–306. MR3857602 https://doi.org/10.1214/14-PS228

Járai, A.A. and Sun, M. (2019). Toppling and height probabilities in sandpiles. *J. Stat. Mech. Theory Exp.* **11** 113204, 26. MR4059705 https://doi.org/10.1088/1742-5468/ab2ccb

Járai, A.A. and Werning, N. (2014). Minimal configurations and sandpile measures. *J. Theoret. Probab.* **27** 153–167. MR3174221 https://doi.org/10.1007/s10959-012-0446-z

Jerison, D.C., Levine, L. and Pike, J. (2019). Mixing time and eigenvalues of the Abelian sandpile Markov chain. *Trans. Amer. Math. Soc.* **372** 8307–8345. MR4029698 https://doi.org/10.1090/tran/7585

Kaiser, R. and Sava-Huss, E. (2024). Scaling limit of the sandpile identity element on the Sierpinski gasket. In *From Classical Analysis to Analysis on Fractals: A Tribute to Robert Strichartz* **2**. Birkhäuser. Available at arXiv:2308.12183.

Kaiser, R., Sava-Huss, E. and Wang, Y. (2024). Abelian sandpiles on Sierpiński gasket graphs. *Electron. J. Combin.* **31** Paper No. 1.6, 23. MR4695551 https://doi.org/10.37236/11520

Le Borgne, Y. and Rossin, D. (2002). On the identity of the sandpile group *Discrete Math.* **256** 775–790. LaCIM 2000 Conference on Combinatorics, Computer Science and Applications (Montreal, QC). MR1935788 https://doi.org/10.1016/S0012-365X(02)00347-3

Levine, L., Pegden, W. and Smart, C.K. (2016). Apollonian structure in the Abelian sandpile. *Geom. Funct. Anal.* **26** 306–336. MR3494492 https://doi.org/10.1007/s00039-016-0358-7

Levine, L., Pegden, W. and Smart, C.K. (2017). The Apollonian structure of integer superharmonic matrices. *Ann. of Math. (2)* **186** 1–67. MR3664999 https://doi.org/10.4007/annals.2017.186.1.1

Manna, S.S. (1990). Large-scale simulation of avalanche cluster distribution in sandpile model. *J. Stat. Phys.* **59** 509–521. https://doi.org/10.1007/BF01015580

Paoletti, G. (2014). *Deterministic Abelian Sandpile Models and Patterns*. *Springer Theses*. Cham: Springer. Thesis, University of Pisa, Pisa, 2012. MR3100415 https://doi.org/10.1007/978-3-319-01204-9

Pegden, W. and Smart, C.K. (2013). Convergence of the Abelian sandpile. *Duke Math. J.* **162** 627–642. MR3039676 https://doi.org/10.1215/00127094-2079677

Pegden, W. and Smart, C.K. (2020). Stability of patterns in the Abelian sandpile. *Ann. Henri Poincaré* **21** 1383–1399. MR4078284 https://doi.org/10.1007/s00023-020-00898-1

Priezzhev, V.B. (1994). Structure of two-dimensional sandpile. I. height probabilities. *J. Stat. Phys.* **74** 955–979. https://doi.org/10.1007/BF02188212

Saloff-Coste, L. (2004). Random walks on finite groups. In *Probability on Discrete Structures* (H. Kesten, ed.). *Encyclopaedia Math. Sci.* **110** 263–346. Berlin: Springer. MR2023654 https://doi.org/10.1007/978-3-662-09444-0_5

Strichartz, R.S. (2006). *Differential Equations on Fractals: A Tutorial*. Princeton, NJ: Princeton Univ. Press. MR2246975

Bernoulli **31**(3), 2025, 2508–2524
https://doi.org/10.3150/24-BEJ1815

Extrinsic derivative formula for distribution dependent SDEs

PANPAN REN[a]

Mathematics Department, City University of Hong Kong, Hong Kong, China, [a]*panparen@cityu.edu.hk*

To characterize the regularity of nonlinear Fokker-Planck equations with respect to weighted variational distances, we establish for the first time a Bismut type formula for the extrinsic derivative of distribution dependent SDEs (DDSDEs). As an application, the Lipschitz continuity in the weighted variational distance is derived for the associated nonlinear Fokker-Planck equation, which can be regarded as the counterpart of the classical contraction property in the linear setting. The main results are illustrated by non-degenerate DDSDEs with space-time singular drift, as well as degenerate DDSDEs with weakly monotone coefficients.

Keywords: Bismut formula; distribution dependent SDEs; extrinsic formula; stochastic Hamiltonian system

References

Bao, J., Ren, P. and Wang, F.-Y. (2021). Bismut formulas for Lions derivative of McKean-Vlasov SDEs with memory. *J. Differ. Equ.* **282** 285–329. https://doi.org/10.1016/j.jde.2021.02.019

Fang, S. and Zhang, T. (2005). A study of a class of stochastic differential equations with non-Lipschitzian coefficients. *Probab. Theory Related Fields* **132** 356–390. MR2197106 https://doi.org/10.1007/s00440-004-0398-z

Huang, X., Ren, P. and Wang, F.-Y. (2021). Distribution dependent stochastic differential equations. *Front. Math. China* **16** 257–301. MR4254653 https://doi.org/10.1007/s11464-021-0920-y

Huang, X., Song, Y. and Wang, F.-Y. (2022). Bismut formula for intrinsic/Lions derivatives of distribution dependent SDEs with singular coefficients. *Discrete Contin. Dyn. Syst.* **42** 4597–4614. MR4455243 https://doi.org/10.3934/dcds.2022065

Kondratiev, Y., Lytvynov, E. and Vershik, A. (2015). Laplace operators on the cone of Radon measures. *J. Funct. Anal.* **269** 2947–2976. MR3394626 https://doi.org/10.1016/j.jfa.2015.06.007

McKean, H.P. Jr. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** 1907–1911. MR0221595 https://doi.org/10.1073/pnas.56.6.1907

Ren, P. (2023). Singular McKean-Vlasov SDEs: Well-posedness, regularities and Wang's Harnack inequality. *Stochastic Process. Appl.* **156** 291–311. MR4518750 https://doi.org/10.1016/j.spa.2022.11.010

Ren, P. and Wang, F.-Y. (2019). Bismut formula for Lions derivative of distribution dependent SDEs and applications. *J. Differ. Equ.* **267** 4745–4777. MR3983053 https://doi.org/10.1016/j.jde.2019.05.016

Ren, P. and Wang, F.-Y. (2020). Spectral gap for measure-valued diffusion processes. *J. Math. Anal. Appl.* **483** 123624, 16. MR4024966 https://doi.org/10.1016/j.jmaa.2019.123624

Ren, P. and Wang, F.-Y. (2021). Derivative formulas in measure on Riemannian manifolds. *Bull. Lond. Math. Soc.* **53** 1786–1800. MR4379563 https://doi.org/10.1112/blms.12542

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

Wang, F.-Y. (2018). Distribution dependent SDEs for Landau type equations. *Stochastic Process. Appl.* **128** 595–621. MR3739509 https://doi.org/10.1016/j.spa.2017.05.006

Wang, F.-Y. (2023a). Derivative formula for singular McKean-Vlasov SDEs. *Commun. Pure Appl. Anal.* **22** 1866–1898. MR4599393 https://doi.org/10.3934/cpaa.2023050

Wang, F.-Y. (2023b). Distribution dependent reflecting stochastic differential equations. *Sci. China Math.* **66** 2411–2456. MR4658659 https://doi.org/10.1007/s11425-021-2028-y

1350-7265

Xia, P., Xie, L., Zhang, X. and Zhao, G. (2020). $L^q(L^p)$-theory of stochastic differential equations. *Stochastic Process. Appl.* **130** 5188–5211. MR4108486 https://doi.org/10.1016/j.spa.2020.03.004

Bernoulli **31**(3), 2025, 2525–2543
https://doi.org/10.3150/24-BEJ1816

The fewest-big-jumps principle and an application to random graphs

CÉLINE KERRIOU[a] and PETER MÖRTERS[b]

Department of Mathematics, University of Cologne, Cologne, Germany, [a]*ckerriou@math.uni-koeln.de,* [b]*p.moerters@uni-koeln.de*

We prove a large deviation principle for the sum of n independent heavy-tailed random variables, which are subject to a moving truncation at location n. Conditional on the sum being large at scale n, we show that a finite number of summands take values near the truncation boundary, while the remaining variables still obey the law of large numbers. This generalises the well-known single-big-jump principle for random variables without truncation to a situation where just the minimal necessary number of jumps occur. As an application, we consider a random graph with vertex set given by the lattice points of a torus with sidelength $2N + 1$. Every vertex is the centre of a ball with random radius sampled from a heavy-tailed distribution. Oriented edges are drawn from the central vertex to all other vertices in this ball. When this graph is conditioned on having an exceptionally large number of edges we use our main result to show that, as $N \to \infty$, the excess outdegrees condense in a fixed, finite number of randomly scattered vertices of macroscopic outdegree. By contrast, no condensation occurs for the indegrees of the vertices, which all remain microscopic in size.

Keywords: Condensation; geometric random graphs; heavy-tailed random variables; large deviations; truncated random variables

References

Asmussen, S. and Klüppelberg, C. (1996). Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic Process. Appl.* **64** 103–125. MR1419495 https://doi.org/10.1016/S0304-4149(96)00087-7

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

Caravenna, F. and Doney, R. (2019). Local large deviations and the strong renewal theorem. *Electron. J. Probab.* **24** Paper No. 72, 48. MR3978222 https://doi.org/10.1214/19-EJP319

Chatterjee, S. and Harel, M. (2020). Localization in random geometric graphs with too many edges. *Ann. Probab.* **48** 574–621. MR4089488 https://doi.org/10.1214/19-AOP1387

Chen, B., Blanchet, J., Rhee, C.-H. and Zwart, B. (2019). Efficient rare-event simulation for multiple jump events in regularly varying random walks and compound Poisson processes. *Math. Oper. Res.* **44** 919–942. MR3996652 https://doi.org/10.1287/moor.2018.0950

Deijfen, M., van der Hofstad, R. and Hooghiemstra, G. (2013). Scale-free percolation. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 817–838. MR3112435 https://doi.org/10.1214/12-AIHP480

Denisov, D., Dieker, A.B. and Shneer, V. (2008). Large deviations for random walks under subexponentiality: The big-jump domain. *Ann. Probab.* **36** 1946–1991. MR2440928 https://doi.org/10.1214/07-AOP382

Foss, S., Konstantopoulos, T. and Zachary, S. (2005). The principle of a single big jump: Discrete and continuous time modulated random walks with heavy-tailed increments. https://doi.org/10.48550/arXiv.math/0509605

Hall, P. (1985). On continuum percolation. *Ann. Probab.* **13** 1250–1266. MR0806222

Janson, S. (2012). Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.* **9** 103–252. MR2908619 https://doi.org/10.1214/11-PS188

Nagaev, S.V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR0542129

Nair, J., Wierman, A. and Zwart, B. (2022). *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press. MR4585805 https://doi.org/10.1017/9781009053730

Petrov, V.V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete [Results in Mathematics and Related Areas], Band 82*. New York: Springer. MR0388499

Rhee, C.-H., Blanchet, J. and Zwart, B. (2019). Sample path large deviations for Lévy processes and random walks with regularly varying increments. *Ann. Probab.* **47** 3551–3605. MR4038038 https://doi.org/10.1214/18-aop1319

Stegehuis, C. and Zwart, B. (2023). Scale-free graphs with many edges. *Electron. Commun. Probab.* **28** Paper No. 62, 11. MR4684075 https://doi.org/10.1214/23-ecp567

Vezzani, A., Barkai, E. and Burioni, R. (2019). Single-big-jump principle in physical modeling. *Phys. Rev. E* **100** 012108. https://doi.org/10.1103/PhysRevE.100.012108