BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Thirty Two Number Three August 2026 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 Two Number Three August 2026 ISSN: 1350-7265

CONTENTS

BERNOULLI Volume 32 Number 3 August 2026 Pages 1687–2544 ISI/BS

Volume 32 Number 3   August 2026   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 2026 Bernoulli consists of 4 issues published in February, May, August and November.

Bernoulli Society
for Mathematical Statistics
and Probability

Bernoulli **32**(3), 2026, 1687–1712
https://doi.org/10.3150/25-BEJ1927

SID: A novel class of nonparametric tests of independence for censored outcomes

JINHONG LI[1,a], JICAI LIU[2,b], JINHONG YOU[3,c] and RIQUAN ZHANG[4,d]

[1] *School of Statistics and Data Science, Laboratory for Statistical Monitoring and Intelligent Governance of Common Prosperit, Zhejiang Gongshang University, Hangzhou, China,* [a]*jinhongli0106@gmail.com*
[2] *School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai, China,* [b]*liujicai1234@126.com*
[3] *School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China,* [c]*johnyou07@163.com*
[4] *School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China,* [d]*zhangriquan@163.com*

We propose a new class of metrics, called the survival independence divergence (SID), to test dependence between a right-censored outcome and covariates. A key technique for deriving the SIDs is to use a counting process strategy, which equivalently transforms the intractable independence test due to the presence of censoring into a test problem for complete observations. The SIDs are equal to zero if and only if the right-censored response and covariates are independent, and they are capable of detecting various types of nonlinear dependence. We propose empirical estimates of the SIDs and establish their asymptotic properties. We further develop a wild bootstrap method to estimate the critical values and show the consistency of the bootstrap tests. The numerical studies demonstrate that our SID-based tests are highly competitive with existing methods in a wide range of settings.

Keywords: Characteristic function; counting process; nonparametric independence test; reproducing kernel Hilbert space; survival analysis

References

Belanche, L. and Villegas, M. (2013). Kernel functions for categorical variables with application to problems in the life sciences. In *International Conference of the Catalan Association for Artificial Intelligence* 171–180.

Chatterjee, S. (2021). A new coefficient of correlation. *J. Amer. Statist. Assoc.* **116** 2009–2022.

Chen, C., Li, K. and Wang, J. (1999). Dimension reduction for censored regression data. *Ann. Statist.* **27** 1–23.

Chwialkowski, K., Strathmann, H. and Gretton, A. (2016). A kernel test of goodness of fit. In *International Conference on Machine Learning* 2606–2615.

Cox, D.R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B, Methodol.* **34** 187–202.

Dugué, A.E., Pulido, M., Chabaud, S., Belin, L. and Gal, J. (2016). How to deal with interval-censored data practically while assessing the progression-free survival: A step-by-step guide using SAS and R software. *Clin. Cancer Res.* **22** 5629–5635.

Edelmann, D., Welchowski, T. and Benner, A. (2021). A consistent version of distance covariance for right-censored survival data and its application in hypothesis testing. *Biometrics* **78** 867–879.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* **64** 865–890.

Fernández, T., Gretton, A., Rindt, D. and Sejdinovic, D. (2023). A kernel log-rank test of independence for right-censored data. *J. Amer. Statist. Assoc.* **118** 925–936.

Fleming, T.R. and Harrington, D.P. (2011). *Counting Processes and Survival Analysis*. John Wiley & Sons.

Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B. and Smola, A. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* **20** 585–592.

Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B. and Smola, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27** 1536–1563.

Janson, S. (1984). The asymptotic distributions of incomplete U-statistics. *Z. Wahrsch. Verw. Gebiete* **66** 495–505.

Ke, C. and Yin, X. (2020). Expected conditional characteristic function-based measures for testing independence. *J. Amer. Statist. Assoc.* **115** 985–996.

Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data 1230*. Springer.

Laurie, J.A., Moertel, C.G., Fleming, T.R., Wieand, H.S., Leigh, J.E., Rubin, J., McCormack, G.W., Gerstner, J.B., Krook, J.E. and Malliard, J. (1989). Surgical adjuvant therapy of large-bowel carcinoma: An evaluation of levamisole and the combination of levamisole and fluorouracil. The North Central Cancer Treatment Group and the Mayo Clinic. *J. Clin. Oncol.* **7** 1447–1456.

Lee, J. (1990). *U-Statistics: Theory and Practice*. New York: Marcel Dekker, Inc.

Li, J., Liu, J., You, J. and Zhang, R. (2026). Supplement to "SID: a novel class of nonparametric tests of independence for censored outcomes." https://doi.org/10.3150/25-BEJ1927SUPP

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* **50** 163–170.

Rodrigues, A.S., Calsavara, V.F., Silva, F.I., Alves, F.A. and Vivas, A.P. (2018). Use of interval-censored survival data as an alternative to Kaplan-Meier survival curves: Studies of oral lesion occurrence in liver transplants and cancer recurrence. *J. Cancer Res.* **38** 1–10.

Schrab, A., Kim, I., Guedj, B. and Gretton, A. (2022). Efficient aggregated kernel tests using incomplete U-statistics. In *Advances in Neural Information Processing Systems* **35** 18793–18807.

Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. CRC Press.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45** 89–103.

Su, L. and White, H. (2007). A consistent characteristic function-based test for conditional independence. *J. Econometrics* **141** 807–834.

Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20** 595–601.

Wang, Y., He, S., Zhu, L. and Yuen, K. (2007). Asymptotics for a censored generalized linear model with unknown link function. *Probab. Theory Related Fields* **138** 235–267.

Wang, X., Pan, W., Hu, W., Tian, Y. and Zhang, H. (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.* **110** 1726–1734.

Wei, L. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Stat. Med.* **11** 1871–1879.

Weihs, L., Drton, M. and Meinshausen, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562.

Yang, B., Yin, X. and Zhang, N. (2019). Sufficient variable selection using independence measures for continuous response. *J. Multivariate Anal.* **173** 480–493.

Zhong, Q., Mueller, J. and Wang, J. (2022). Deep learning for the partially linear Cox model. *Ann. Statist.* **50** 1348–1375.

Bernoulli **32**(3), 2026, 1713–1736
https://doi.org/10.3150/25-BEJ1929

Identifiability of overcomplete independent component analysis

KEXIN WANG[a] and ANNA SEIGAL[b]

SEAS, Harvard University, Cambridge, United States, [a]*kexin_wang@g.harvard.edu,* [b]*aseigal@seas.harvard.edu*

Independent component analysis (ICA) studies mixtures of independent latent sources. An ICA model is identifiable if the mixing can be recovered uniquely. When the number of sources is at most the number of observations, Comon proved in 1994 that ICA is identifiable if and only if at most one source is Gaussian. However, in the overcomplete setting, where the number of sources exceeds the number of observations, an if and only if characterization for identifiability has been missing. In this paper, we give such a characterization. The proof studies linear spaces of rank one symmetric matrices. For generic mixing, we present an identifiability condition in terms of the number of sources and the number of observations. We use our identifiability results to design a coupled matrix and tensor decomposition algorithm to recover the mixing matrix from data and apply it to synthetic data and two real datasets.

Keywords: Independent component analysis; identifiability; real algebraic geometry

References

[1] Albera, L., Comon, P., Chevalier, P. and Ferréol, A. (2004). Blind identification of underdetermined mixtures based on the hexacovariance. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* **2** ii–29. IEEE.

[2] Alexander, J. and Hirschowitz, A. (1995). Polynomial interpolation in several variables. *J. Algebraic Geom.* **4**.

[3] Anandkumar, A., Ge, R. and Janzamin, M. (2015). Learning overcomplete latent variable models through tensor methods. In *Proceedings of the 28th Conference on Learning Theory*. *Proceedings of Machine Learning Research* **40** 36–112. Paris, France: PMLR.

[4] Bartlett, M.S., Movellan, J.R. and Sejnowski, T.J. (2002). Face recognition by independent component analysis. *IEEE Trans. Neural Netw.* **13** 1450–1464.

[5] Bashelor, A., Ksir, A. and Traves, W. (2008). Enumerative algebraic geometry of conics. *Amer. Math. Monthly* **115** 701–728.

[6] Cardoso, J.-F. and Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals. *IEE Proc., F, Radar Signal Process.* **140** 362–370.

[7] Chiantini, L. and Ciliberto, C. (2002). Weakly defective varieties. *Trans. Amer. Math. Soc.* **354** 151–178.

[8] Chiantini, L., Ottaviani, G. and Vannieuwenhoven, N. (2017). On generic identifiability of symmetric tensors of subgeneric rank. *Trans. Amer. Math. Soc.* **369** 4021–4042.

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

[10] Comon, P. and Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press.

[11] Darmois, G. (1953). Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire. *Rev. Inst. Int. Stat.* 2–8.

[12] Davies, M. and Mitianoudis, N. (2004). Simple mixture model for sparse overcomplete ICA. *IEE Proc., Vis. Image Signal Process.* **151** 35–43.

[13] De Lathauwer, L., Castaing, J. and Cardoso, J.-F. (2007). Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Trans. Signal Process.* **55** 2965–2973.

[14] De Lathauwer, L., De Moor, B. and Vandewalle, J. (2001). Independent component analysis and (simultaneous) third-order tensor diagonalization. *IEEE Trans. Signal Process.* **49** 2262–2271.

[15] Duff, T., Hill, C., Jensen, A., Lee, K., Leykin, A. and Sommars, J. (2019). Solving polynomial systems via homotopy continuation and monodromy. *IMA J. Numer. Anal.* **39** 1421–1446.

[16] Eriksson, J. and Koivunen, V. (2004). Identifiability, separability, and uniqueness of linear ICA models. *IEEE Signal Process. Lett.* **11** 601–604.

[17] Fevola, C., Mandelshtam, Y. and Sturmfels, B. (2021). Pencils of quadrics: Old and new. *Matematiche* **76** 319–335.

[18] Grayson, D.R. and Stillman, M.E. (2002). Macaulay2, a software system for research in algebraic geometry.

[19] Harris, J. (1992). *Algebraic Geometry: A First Course*. *Graduate Texts in Mathematics*. Springer.

[20] Hartshorne, R. (2013). *Algebr. Geom.* **52**. Springer Science & Business Media.

[21] Heath, T.L. (1921). *A History of Greek Mathematics, Vol. 1: From Thales to Euclid*. Clarendon Press.

[22] Hyvarinen, A., Cristescu, R. and Oja, E. (1999). A fast algorithm for estimating overcomplete ICA bases for image windows. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)* **2** 894–899. IEEE.

[23] Jung, T.-P., Makeig, S., McKeown, M.J., Bell, A.J., Lee, T.-W. and Sejnowski, T.J. (2001). Imaging brain dynamics using independent component analysis. *Proc. IEEE* **89** 1107–1122.

[24] Jutten, C. and Karhunen, J. (2004). Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures. *Int. J. Neural Syst.* **14** 267–292.

[25] Kagan, A.M., Linnik, I.U.V. and Rao, C.R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley. A Wiley-Interscience Publication.

[26] Kileel, J. and Pereira, J.M. (2025). Subspace power method for symmetric tensor decomposition. *Numer. Algorithms* 1–38.

[27] Krizhevsky, A., Nair, V. and Hinton, G. (2014). The CIFAR- 10 dataset. University of Toronto.

[28] Kruskal, J.B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **18** 95–138.

[29] Lewicki, M.S. and Sejnowski, T.J. (2000). Learning overcomplete representations. *Neural Comput.* **12** 337–365.

[30] Mayr, E.W. and Meyer, A.R. (1982). The complexity of the word problems for commutative semigroups and polynomial ideals. *Adv. Math.* **46** 305–329.

[31] McCullagh, P. (2018). *Tensor Methods in Statistics*. Courier Dover Publications.

[32] Moreau, E. (2001). A generalization of joint-diagonalization criteria for source separation. *IEEE Trans. Signal Process.* **49** 530–541.

[33] Odehnal, B., Stachel, H. and Glaeser, G. (2020). *The Universe of Quadrics*. Springer Nature.

[34] Olshausen, B.A. and Field, D.J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res.* **37** 3311–3325.

[35] Podosinnikova, A., Perry, A., Wein, A.S., Bach, F., d'Aspremont, A. and Sontag, D. (2019). Overcomplete independent component analysis via SDP. In *The 22nd International Conference on Artificial Intelligence and Statistics* 2583–2592. PMLR.

[36] Powell, M.J. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* **7** 155–162.

[37] Ranestad, K., Seigal, A. and Wang, K. (2024). A Real Generalized Trisecant Trichotomy. arXiv preprint, arXiv:2409.01356.

[38] Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A. and Nolan, G.P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308** 523–529.

[39] Shimizu, S., Hoyer, P.O., Hyvärinen, A., Kerminen, A. and Jordan, M. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **7** 2003–2030.

[40] Skitovič, V. (1962). Linear combinations of independent random variables and the normal distribution law. *Sel. Transl. Math. Stat. Probab.* **2** 211–28.

[41] Skitovitch, V.P. (1953). On a property of the normal distribution. *Dokl. Akad. Nauk SSSR* **89** 217–219.

[42] Sokol, A., Maathuis, M.H. and Falkeborg, B. (2014). Quantifying identifiability in independent component analysis. *Electron. J. Stat.* **8** 1438–1459.

[43] Theis, F.J., Lang, E.W. and Puntonet, C.G. (2004). A geometric algorithm for overcomplete linear ICA. *Neurocomputing* **56** 381–398.

[44] Uddin, Z., Ahmad, A., Iqbal, M. and Naeem, M. (2015). Applications of independent component analysis in wireless communication systems. *Wirel. Pers. Commun.* **83** 2711–2737.
[45] Wang, K. and Seigal, A. (2026). Supplement to "Identifiability of overcomplete independent component analysis." https://doi.org/10.3150/25-BEJ1929SUPP

Bernoulli **32**(3), 2026, 1737–1762
https://doi.org/10.3150/25-BEJ1930

Merging rate of opinions via optimal transport on random measures

MARTA CATALANO[1,a] and HUGO LAVENANT[2,b]

[1]*Luiss University, Rome Italy,* [a]*mcatalano@luiss.it*
[2]*Bocconi University, Milan, Italy,* [b]*hugo.lavenant@unibocconi.it*

Random measures provide flexible parameters for Bayesian nonparametric models. Given two different priors for a random measure, we develop a natural framework to investigate the rate at which the corresponding posteriors merge, as the sample size increases. We define a new distance between the laws of random measures that is built as a Wasserstein distance on the ground space of unbalanced measures, endowed with the bounded Lipschitz metric. We develop tight analytical bounds for its specification to completely random measures, including the special case of Poisson and gamma random measures. The bounds are interpreted in terms of an adapted extended Wasserstein distance between the Lévy measures and are used to investigate the merging between the posteriors of normalized gamma and generalized gamma priors. After a careful study on the identifiability of the law of the random measure, interesting asymptotic and finite-sample insights are derived without putting *any* assumption on the true data generating process.

Keywords: Bayesian nonparametrics; completely random measures; Cox process; impact of the prior; Lévy measure; merging of opinions; optimal transport; Poisson process; Wasserstein distance

References

Backhoff, J., Bartl, D., Beiglböck, M. and Wiesel, J. (2022). Estimating processes in adapted Wasserstein distance. *Ann. Appl. Probab.* **32** 529–550.

Barbour, A.D. and Brown, T.C. (1992). Stein's method and point process approximation. *Stoch. Process. Appl.* **43** 9–31.

Barrios, E., Lijoi, A., Nieto-Barajas, L.E. and Prünster, I. (2013). Modeling with normalized random measure mixture models. *Statist. Sci.* **28** 313–334.

Bartl, D., Beiglböck, M. and Pammer, G. (2024). The Wasserstein space of stochastic processes. *J. Eur. Math. Soc.*

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *Ann. Math. Stat.* **33** 882–886.

Bogachev, V.I. (2007). *Measure Theory. Vol. I, II.* Berlin: Springer-Verlag.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Probab.* **31** 929–953.

Caron, F. and Fox, E. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **5** 1295–1366.

Catalano, M. and Lavenant, H. (2024). Hierarchical integral probability metrics: A distance on random probability measures with low sample complexity. In *Proceedings of the 41st International Conference on Machine Learning* (R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett and F. Berkenkamp, eds.). *Proceedings of Machine Learning Research* **235** 5841–5861. PMLR.

Catalano, M. and Lavenant, H. (2026). Supplement to "Merging rate of opinions via optimal transport on random measures." https://doi.org/10.3150/25-BEJ1930SUPP

Catalano, M., Lijoi, A. and Prünster, I. (2020). Approximation of Bayesian models for time-to-event data. *Electron. J. Stat.* **14** 3366–3395.

Catalano, M., Lijoi, A. and Prünster, I. (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *Ann. Statist.* **49** 2916–2947.

Catalano, M., Lavenant, H., Lijoi, A. and Prünster, I. (2024). A Wasserstein index of dependence for random measures. *J. Amer. Statist. Assoc.* **119** 2396–2406.

Cox, D.R. (1955). Some statistical methods connected with series of events. *J. Roy. Statist. Soc. Ser. B, Methodol.* **17** 129–157.

De Blasi, P., Peccati, G. and Prünster, I. (2009). Asymptotics for posterior hazards. *Ann. Statist.* **37** 1906–1945.

Decreusefond, L., Schulte, M. and Thäle, C. (2016). Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Probab.* **44** 2147–2197.

Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–26.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201.

Dukler, Y., Li, W., Lin, A. and Montúfar, G. (2019). Wasserstein of Wasserstein loss for learning generative models. In *International Conference on Machine Learning* 1716–1725. PMLR.

Dykstra, R.L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367.

Erbar, M., Huesmann, M., Jalowy, J. and Müller, B. (2025). Optimal transport of stationary point processes: Metric structure, gradient flow and convexity of the specific entropy. *J. Funct. Anal.* **289** 110974.

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

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

Figalli, A. and Gigli, N. (2010). A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *J. Math. Pures Appl.* **94** 107–130.

Fortet, R. and Mourier, E. (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. Éc. Norm. Supér.* **70** 267–285.

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.

Giordano, R., Liu, R., Jordan, M.I. and Broderick, T. (2022). Evaluating sensitivity to the stick-breaking prior in Bayesian nonparametrics (with discussion). *Bayesian Anal.* 1–53.

Griffin, J.E. and Leisen, F. (2017). Compound random measures and their use in Bayesian non-parametrics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 525–545.

Guillen, N., Mou, C. and Święch, A. (2019). Coupling Lévy measures and comparison principles for viscosity solutions. *Trans. Amer. Math. Soc.* **372** 7327–70.

Gustafson, P. (1996). Local sensitivity of posterior expectations. *Ann. Statist.* **24** 174–195.

Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294.

Ho, N., Nguyen, X., Yurochkin, M., Bui, H.H., Huynh, V. and Phung, D. (2017). Multilevel clustering via Wasserstein means. In *Proceedings of the 34th International Conference on Machine Learning* **70** 1501–1509.

Huesmann, M. (2016). Optimal transport between random measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 196–232.

James, L.F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33** 1771–1799.

James, L.F., Lijoi, A. and Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scand. J. Stat.* **33** 105–120.

James, L.F., Lijoi, A. and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36** 76–97.

Kallenberg, O. (2017). *Random Measures, Theory and Applications*. *Probability Theory and Stochastic Modelling*. Cham: Springer International Publishing.

Kingman, J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78.

Korwar, R.M. and Hollander, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.* **1** 705–711.

Lau, J.W. and Cripps, E. (2022). Thinned completely random measures with applications in competing risks models. *Bernoulli* **28** 638–662.

Lee, J., Miscouridou, X. and Caron, F. (2023). A unified construction for series representations and finite approximations of completely random measures. *Bernoulli* **29** 2142–2166.

Ley, C., Reinert, G. and Swan, Y. (2017). Distances between nested densities and a measure of the impact of the prior in Bayesian statistics. *Ann. Appl. Probab.* **27** 216–241.

Lijoi, A., Mena, R.H. and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100** 1278–1291.

Lijoi, A., Mena, R.H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 715–740.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* 80–136. Cambridge: Cambridge University Press.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12** 351–357.

Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400.

Nguyen, X. (2016). Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure. *Bernoulli* **22** 1535–1571.

Nieto-Barajas, L.E. and Prünster, I. (2009). A sensitivity analysis for Bayesian nonparametric density estimators. *Statist. Sinica* **19** 685–705.

Peyré, G. and Cuturi, M. (2019). Computational optimal transport – with applications to data science. *Found. Trends Mach. Learn.* **11** 355–607.

Pflug, G.C. and Pichler, A. (2012). A distance for multistage stochastic optimization models. *SIAM J. Optim.* **22** 1–23.

Pitman, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Mathematics* **1875**. Berlin: Springer-Verlag. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 2002.

Pitman, J. and Yor, M. (1992). Arcsine laws and interval partitions derived from a stable subordinator. *Proc. Lond. Math. Soc.* **s3-65** 326–356.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900.

Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist.* **31** 560–585.

Saha, A. and Kurtek, S. (2019). Geometric sensitivity measures for Bayesian nonparametric density estimation models. *Sankhya, Ser. A* **81**.

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians* **55** 94. New York: Birkhäuser.

Schuhmacher, D. and Xia, A. (2008). A new metric between distributions of point processes. *Adv. Appl. Probab.* **40** 651–672.

Villani, C. (2009). *Optimal Transport: Old and New 338*. Heidelberg: Springer.

Yurochkin, M., Claici, S., Chien, E., Mirzazadeh, F. and Solomon, J.M. (2019). Hierarchical optimal transport for document representation. *Adv. Neural Inf. Process. Syst.* **32**.

Bernoulli **32**(3), 2026, 1763–1787
https://doi.org/10.3150/25-BEJ1931

Dimension-free uniform concentration bound for logistic regression

SHOGO NAKAKITA[a]

Komaba Institute for Science, University of Tokyo, Tokyo, Japan, [a]*nakakita@g.ecc.u-tokyo.ac.jp*

We provide a novel dimension-free uniform concentration bound for the empirical risk function of constrained logistic regression. Our bound yields a milder sufficient condition for a uniform law of large numbers than conditions derived by the Rademacher complexity argument and McDiarmid's inequality. The derivation is based on the PAC-Bayes approach with second-order expansion and Rademacher-complexity-based bounds for the residual term of the expansion.

Keywords: Effective rank; logistic regression; uniform concentration bound; uniform law of large numbers

References

Adamczak, R. (2015). A note on the Hanson-Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.* **20** 1–13. https://doi.org/10.1214/ECP.v20-3829

Alquier, P. (2024). User-friendly introduction to PAC-Bayes bounds. *Found. Trends Mach. Learn.* **17** 174–303. https://doi.org/10.1561/2200000100

Alquier, P., Ridgway, J. and Chopin, N. (2016). On the properties of variational approximations of Gibbs posteriors. *J. Mach. Learn. Res.* **17** 1–41.

Appert, G. and Catoni, O. (2021). New bounds for k-means and information k-means. Preprint. Available at arXiv:2101.05728 [math.ST].

Arsov, N., Pavlovski, M. and Kocarev, L. (2019). Stability of decision trees and logistic regression. Preprint. Available at arXiv:1903.00816 [cs.LG].

Aubin, B., Krzakala, F., Lu, Y. and Zdeborová, L. (2020). Generalization error in high-dimensional perceptrons: Approaching Bayes error with convex optimization. *Adv. Neural Inf. Process. Syst.* **33** 12199–12210.

Bach, F. (2024). *Learning Theory from First Principles*. MIT Press.

Bakry, D., Gentil, I. and Ledoux, M. (2014). *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der mathematischen Wissenschaften*. Springer International Publishing. https://doi.org/10.1007/978-3-319-00227-9

Bartlett, P.L. and Mendelson, S. (2002). Rademacher and Gaussian complexities: Risk bounds and structural results. *J. Mach. Learn. Res.* **3** 463–482.

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. https://doi.org/10.1073/pnas.1907378117

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

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

Catoni, O. (2004). *Statistical Learning Theory and Stochastic Optimization: Ecole d'Eté de Probabilités de Saint-Flour, XXXI-2001*. Springer Science & Business Media. https://doi.org/10.1007/b99352

Catoni, O. (2007). *PAC-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning*. Institute of Mathematical Statistics. https://doi.org/10.1214/074921707000000391

Catoni, O. and Giulini, I. (2017). Dimension-free PAC-Bayesian bounds for matrices, vectors, and linear least squares regression. Preprint. Available at arXiv:1712.02747 [math.ST].

1350-7265

Cheng, C. and Montanari, A. (2024). Dimension free ridge regression. *Ann. Statist.* **52** 2879–2912. https://doi.org/10.1214/24-AOS2449

Cover, T.M. (1964). Geometrical and Statistical Properties of Linear Threshold Devices. PhD thesis, Stanford University.

Emami, M., Sahraee-Ardakan, M., Pandit, P., Rangan, S. and Fletcher, A. (2020). Generalization error of generalized linear models in high dimensions. In *International Conference on Machine Learning 2892–2901*. PMLR.

Giulini, I. (2018). Robust dimension-free Gram operator estimates. *Bernoulli* **24** 3864–3923. https://doi.org/10.3150/17-BEJ981

Hsu, D. and Mazumdar, A. (2024). On the sample complexity of parameter estimation in logistic regression with normal design. In *The Thirty Seventh Annual Conference on Learning Theory* 2418–2437. PMLR.

James, G.M. and Radchenko, P. (2009). A generalized Dantzig selector with shrinkage tuning. *Biometrika* **96** 323–337. https://doi.org/10.1093/biomet/asp013

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

Kuchelmeister, F. and van de Geer, S. (2024). Finite sample rates for logistic regression with small noise or few samples. *Sankhya A*. Advance online publication. https://doi.org/10.1007/s13171-024-00358-6

Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. American Mathematical Society.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-20212-4

Levy, T. and Abramovich, F. (2023). Generalization error bounds for multiclass sparse linear classifiers. *J. Mach. Learn. Res.* **24** 1–35.

Liang, H. and Du, P. (2012). Maximum likelihood estimation in logistic regression models with a diverging number of covariates. *Electron. J. Stat.* **6** 1838–1846. https://doi.org/10.1214/12-EJS731

Meir, R. and Zhang, T. (2003). Generalization error bounds for Bayesian mixture algorithms. *J. Mach. Learn. Res.* **4** 839–860.

Salehi, F., Abbasi, E. and Hassibi, B. (2019). The impact of regularization on high-dimensional logistic regression. *Adv. Neural Inf. Process. Syst.* **32**.

Sawaya, K., Uematsu, Y. and Imaizumi, M. (2023). Moment-Based Adjustments of Statistical Inference in High-Dimensional Generalized Linear Models. Preprint. Available at arXiv:2305.17731 [math.ST].

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

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

Tsigler, A. and Bartlett, P.L. (2023). Benign overfitting in ridge regression. *J. Mach. Learn. Res.* **24** 1–76.

Tsuzuku, Y., Sato, I. and Sugiyama, M. (2020). Normalized flat minima: Exploring scale invariant definition of flat minima for neural networks using PAC-Bayesian analysis. In *International Conference on Machine Learning* 9636–9647. PMLR.

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

van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511802256

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press. https://doi.org/10.1017/9781108231596

Wainwright, M.J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press. https://doi.org/10.1017/9781108627771

Wang, H., Keskar, N.S., Xiong, C. and Socher, R. (2018). Identifying generalization properties in neural networks. Preprint. Available at arXiv:1809.07402 [cs.LG].

Wu, J., Zou, D., Chen, Z., Braverman, V., Gu, Q. and Kakade, S.M. (2023). Finite-sample analysis of learning high-dimensional single ReLU neuron. In *International Conference on Machine Learning* 37919–37951. PMLR.

Zhao, Q., Sur, P. and Candes, E.J. (2022). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli* **28** 1835–1861. https://doi.org/10.3150/21-BEJ1401

Zhivotovskiy, N. (2024). Dimension-free bounds for sums of independent matrices and simple tensors via the variational principle. *Electron. J. Probab.* **29** 1–28. https://doi.org/10.1214/23-EJP1021

Bernoulli **32**(3), 2026, 1788–1811
https://doi.org/10.3150/25-BEJ1932

Optimal designs for regression on Lie groups

SOMNATH CHAKRABORTY[1,a], HOLGER DETTE[2,b] and MARTIN KROLL[3,c]

[1]*Department of Mathematics and Statistics, University of Helsinki, FI-00560 Helsinki, Finland,* [a]*somnath.chakraborty@helsinki.fi*
[2]*Fakultät für Mathematik, Ruhr-Universität Bochum, D-44801 Bochum, Germany,* [b]*holger.dette@rub.de*
[3]*Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, D-95440 Bayreuth, Germany,* [c]*martin.kroll@uni-bayreuth.de*

We consider a linear regression model with complex-valued response and predictors from a compact and connected Lie group. The regression model is formulated in terms of eigenfunctions of the Laplace-Beltrami operator on the Lie group. We show that the normalized Haar measure is an *approximate* optimal design with respect to all Kiefer's Φ_p-criteria. Inspired by the concept of t-designs in the field of algebraic combinatorics, we then consider so-called λ-designs in order to construct *exact* Φ_p-optimal designs for fixed sample sizes in the considered regression problem. In particular, we explicitly construct Φ_p-optimal designs for regression models with predictors in the Lie groups $\mathrm{SU}(2)$ and $\mathrm{SO}(3)$, the groups of 2×2 unitary matrices and 3×3 orthogonal matrices with determinant equal to 1, respectively. We also discuss the advantages of the derived theoretical results in a concrete biological application.

Keywords: Approximate design; Haar measure; Kiefer's optimality criteria; Laplace-Beltrami operator; Lie groups; linear regression; optimal design; spherical t-design; Wigner's D-matrices; λ-design

References

Abrial, P., Moudden, Y., Starck, J.L., Fadili, J., Delabrouille, J. and Nguyen, M.K. (2008). CMB data analysis and sparsity. *Stat. Methodol.* **5** 289–298. https://doi.org/10.1016/j.stamet.2007.11.005

Alfeld, P., Neamtu, M. and Schumaker, L.L. (1996). Fitting scattered data on sphere-like surfaces using spherical splines. *J. Comput. Appl. Math.* **73** 5–43. MR1424867 https://doi.org/10.1016/0377-0427(96)00034-9

Avery, J. and Wen, Z. (1982). Angular integrations in m-dimensional spaces and hyperspherical harmonics. *Int. J. Quantum Chem.* **22** 717–738. https://doi.org/10.1002/qua.560220406

Bajnok, B. (1991). Construction of spherical 4- and 5-designs. *Graphs Combin.* **7** 219–233. MR1130484 https://doi.org/10.1007/BF01787629

Bannai, E. and Bannai, E. (2009). A survey on spherical designs and algebraic combinatorics on spheres. *European J. Combin.* **30** 1392–1425. Association Schemes: Ideas and Perspectives. https://doi.org/10.1016/j.ejc.2008.11.007

Bannai, E. and Damerell, R.M. (1979). Tight spherical designs. I. *J. Math. Soc. Japan* **31** 199–207. MR519045 https://doi.org/10.2969/jmsj/03110199

Bannai, E. and Damerell, R.M. (1980). Tight spherical designs. II. *J. Lond. Math. Soc. (2)* **21** 13–30. MR576179 https://doi.org/10.1112/jlms/s2-21.1.13

Bannai, E., Nakata, Y., Okuda, T. and Zhao, D. (2022). Explicit construction of exact unitary designs. *Adv. Math.* **405**. Paper No. 108457, 25. https://doi.org/10.1016/j.aim.2022.108457

Berger, M., Gauduchon, P. and Mazet, E. (1971). *Le Spectre d'une Variété Riemannienne*. Berlin Heidelberg: Springer. https://doi.org/10.1007/bfb0064643

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. https://doi.org/10.1214/aos/1046294456

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds — II. *Ann. Statist.* **33**. https://doi.org/10.1214/009053605000000093

Bondarenko, A., Radchenko, D. and Viazovska, M. (2013). Optimal asymptotic bounds for spherical designs. *Ann. of Math. (2)* **178** 443–452. https://doi.org/10.4007/annals.2013.178.2.2

Chakraborty, S., Dette, H. and Kroll, M. (2026). Supplement to "Optimal Designs for Regression on Lie Groups." https://doi.org/10.3150/25-BEJ1932SUPP

Chang, T. (1986). Spherical regression. *Ann. Statist.* **14** 907–924. MR856797 https://doi.org/10.1214/aos/1176350041

Chirikjian, G.S. and Kyatkin, A.B. (2021). *Engineering Applications of Noncommutative Harmonic Analysis*. Taylor & Francis Group. https://doi.org/10.1201/9780429289385

Delsarte, P., Goethals, J.M. and Seidel, J.J. (1977). Spherical codes and designs. *Geom. Dedicata* **6** 363–388. MR485471 https://doi.org/10.1007/bf03187604

Dette, H., Melas, V.B. and Pepelyshev, A. (2005). Optimal designs for three-dimensional shape analysis with spherical harmonic descriptors. *Ann. Statist.* **33** 2758–2788. https://doi.org/10.1214/009053605000000552

Dette, H., Konstantinou, M., Schorning, K. and Gösmann, J. (2019). Optimal designs for regression with spherical data. *Electron. J. Stat.* **13** 361–390. https://doi.org/10.1214/18-ejs1524

Di Marzio, M., Panzera, A. and Taylor, C.C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.* **79** 2066–2075. MR2571770 https://doi.org/10.1016/j.spl.2009.06.014

Downs, T.D. (2003). Spherical regression. *Biometrika* **90** 655–668. https://doi.org/10.1093/biomet/90.3.655

Filová, L., Harman, R. and Klein, T. (2011). Approximate E-optimal designs for the model of spring balance weighing with a constant bias. *J. Statist. Plann. Inference* **141** 2480–2488. https://doi.org/10.1016/j.jspi.2011.02.011

Gariboldi, B. and Gigante, G. (2021). Optimal asymptotic bounds for designs on manifolds. *Anal. PDE* **14** 1701–1724. https://doi.org/10.2140/apde.2021.14.1701

Giné M, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3**. https://doi.org/10.1214/aos/1176343283

Gräf, M. and Potts, D. (2009). Sampling sets and quadrature formulae on the rotation group. *Numer. Funct. Anal. Optim.* **30** 665–688. https://doi.org/10.1080/01630560903163508

Guedes, I.A., Pereira, F.S.S. and Dardenne, L.E. (2018). Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Front. Pharmacol.* **9**. https://doi.org/10.3389/fphar.2018.01089

Haines, L.M. (2024). Optimal designs for spherical harmonic regression. *Stat. Pap.* **66**. https://doi.org/10.1007/s00362-024-01630-5

Harman, R. (2004). Minimal efficiency of designs under the class of orthogonally invariant information criteria. *Metrika* **60** 137–153. https://doi.org/10.1007/s001840300301

Hastings, M.B. and Harrow, A.W. (2009). Classical and quantum tensor product expanders. *Quantum Inf. Comput.* **9** 336–360.

Hendriks, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.* **18** 832–849. https://doi.org/10.1214/aos/1176347628

Hielscher, R. (2013). Kernel density estimation on the rotation group and its application to crystallographic texture analysis. *J. Multivariate Anal.* **119** 119–143. MR3061419 https://doi.org/10.1016/j.jmva.2013.03.014

Jeon, J.M., Park, B.U. and Van Keilegom, I. (2022). Nonparametric regression on Lie groups with measurement errors. *Ann. Statist.* **50** 2973–3008. https://doi.org/10.1214/22-aos2218

Kerkyacharian, G., Nickl, R. and Picard, D. (2011). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Related Fields* **153** 363–404. https://doi.org/10.1007/s00440-011-0348-5

Khalid, Z., Durrani, S., Kennedy, R.A., Wiaux, Y. and McEwen, J.D. (2016). Gauss-Legendre sampling on the rotation group. *IEEE Signal Process. Lett.* **23** 207–211. https://doi.org/10.1109/LSP.2015.2503295

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879. https://doi.org/10.1214/aos/1176342810

Kitchen, D.B., Decornez, H., Furr, J.R. and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.* **3** 935–949. https://doi.org/10.1038/nrd1549

Korevaar, J. and Meyers, J.L.H. (1994). Chebyshev-type quadrature on multidimensional domains. *J. Approx. Theory* **79** 144–164. MR1294326 https://doi.org/10.1006/jath.1994.1119

Kostelec, P.J. and Rockmore, D.N. (2008). FFTs on the rotation group. *J. Fourier Anal. Appl.* **14** 145–179. https://doi.org/10.1007/s00041-008-9013-5

Kovacs, J.A., Chacón, P., Cong, Y., Metwally, E. and Wriggers, W. (2003). Fast rotational matching of rigid bodies by fast Fourier transform acceleration of five degrees of freedom. *Acta Crystallogr., Sect. D, Biol. Crystallogr.* **59** 1371–1376. https://doi.org/10.1107/s0907444903011247

Kowalski, E. (2014). *An Introduction to the Representation Theory of Groups*. *Graduate Studies in Mathematics* **155**. Providence, RI: American Mathematical Society. MR3236265 https://doi.org/10.1090/gsm/155

Lacour, C. and Pham Ngoc, T.M. (2014). Goodness-of-fit test for noisy directional data. *Bernoulli* **20** 2131–2168. MR3263101 https://doi.org/10.3150/13-BEJ553

Lee, J.M. and Ruymgaart, F.H. (1996). Nonparametric curve estimation on Stiefel manifolds. *J. Nonparametr. Stat.* **6** 57–68. https://doi.org/10.1080/10485259608832663

Li, H. and Castillo, E.D. (2024). Optimal design of experiments on Riemannian manifolds. *J. Amer. Statist. Assoc.* **119** 875–886. https://doi.org/10.1080/01621459.2022.2146587

Lin, S.-B. (2019). Nonparametric regression using needlet kernels for spherical data. *J. Complexity* **50** 66–83. https://doi.org/10.1016/j.jco.2018.09.003

Lin, L., Mu, N., Cheung, P. and Dunson, D. (2019). Extrinsic Gaussian processes for regression and classification on manifolds. *Bayesian Anal.* **14**. https://doi.org/10.1214/18-ba1135

Marzio, M.D., Panzera, A. and Taylor, C.C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.* **109** 748–763. https://doi.org/10.1080/01621459.2013.866567

Mason, J.K. and Patala, S. (2019). Basis functions on the grain boundary space: Theory.

Mason, J.K. and Schuh, C.A. (2008). Hyperspherical harmonics for the representation of crystallographic texture. *Acta Mater.* **56** 6141–6155. https://doi.org/10.1016/j.actamat.2008.08.031

Miller, K.S. (1973). Complex linear least squares. *SIAM Rev.* **15** 706–726. https://doi.org/10.1137/1015094

Mimura, Y. (1990). A construction of spherical 2-design. *Graphs Combin.* **6** 369–372. MR1092586 https://doi.org/10.1007/BF01787704

Monnier, J.-B. (2011). Nonparametric regression on the hyper-sphere with uniform design: Needlets-based regression on the hyper-sphere. *TEST* **20** 412–446. MR2834054 https://doi.org/10.1007/s11749-011-0233-7

Narcowich, F.J., Petrushev, P. and Ward, J.D. (2006). Localized tight frames on spheres. *SIAM J. Math. Anal.* **38** 574–594. https://doi.org/10.1137/040614359

Padhorny, D., Kazennov, A., Zerbe, B.S., Porter, K.A., Xia, B., Mottarella, S.E., Kholodov, Y., Ritchie, D.W., Vajda, S. and Kozakov, D. (2016). Protein–protein docking by fast generalized Fourier transforms on 5D rotational manifolds. *Proc. Natl. Acad. Sci. USA* **113**. https://doi.org/10.1073/pnas.1603929113

Patala, S., Mason, J.K. and Schuh, C.A. (2012). Improved representations of misorientation information for grain boundary science and engineering. *Prog. Mater. Sci.* **57** 1383–1425. https://doi.org/10.1016/j.pmatsci.2012.04.002

Pukelsheim, F. (2006). *Optimal Design of Experiments*. *Classics in Applied Mathematics* **50**. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). Reprint of the 1993 original. https://doi.org/10.1137/1.9780898719109

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79** 763–770. https://doi.org/10.1093/biomet/79.4.763

Schaeben, H., Hielscher, R., Fundenberger, J.-J., Potts, D. and Prestin, J. (2007). Orientation density function-controlled pole probability density function measurements: Automated adaptive control of texture goniometers. *J. Appl. Crystallogr.* **40** 570–579. https://doi.org/10.1107/s0021889807019711

Sen, P. (2018). Efficient quantum tensor product expanders and unitary t-designs via the zigzag product. https://doi.org/10.48550/arXiv.1808.10521

Shen, J., Xu, J. and Zhang, P. (2017). Approximations on SO(3) by Wigner D-matrix and applications. *J. Sci. Comput.* **74** 1706–1724. https://doi.org/10.1007/s10915-017-0515-7

Szegö, G. (1975). *Orthogonal Polynomials*. Providence, RI: American Mathematical Society.

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* **2** 5–16. https://doi.org/10.1137/0902002

Womersley, R. (2017a). Efficient Spherical t-Designs on the Sphere S^3 with Good Goemetric Properties. https://web.maths.unsw.edu.au/~rsw/Sphere/EffSphDes/S3SD.html.

Womersley, R. (2017b). Efficient Spherical Designs on the sphere S^2 with good geometric properties. https://web.maths.unsw.edu.au/~rsw/Sphere/EffSphDes/sf.html.

Womersley, R.S. (2018). Efficient spherical designs with good geometric properties. In *Contemporary Computational Mathematics - a Celebration of the 80th Birthday of Ian Sloan* 1243–1285. Springer International Publishing. https://doi.org/10.1007/978-3-319-72456-0_57

Bernoulli **32**(3), 2026, 1812–1826
https://doi.org/10.3150/25-BEJ1933

Strong convergence for tensor GUE random matrices

BENOÎT COLLINS[1,a] and WANGJUN YUAN[2,b]

[1] *Department of Mathematics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan,*
[a] *collins@math.kyoto-u.ac.jp*
[2] *Department of Mathematics, Southern University of Science and Technology, Shenzhen, China,*
[b] *ywangjun@connect.hku.hk*

(*Ann. of Math. (2)* **162** (2005) 711–775) proved that iid GUEs converge strongly to free semicircular elements as the dimension grows to infinity. Motivated by considerations from quantum physics – in particular, understanding nearest neighbor interactions in quantum spin systems – we consider iid GUE acting on multipartite state spaces, with components on more than half of the sites and identity on the remaining sites. In particular, any two GUEs have some sites in common. We show that under proper assumptions on the dimension of the sites, strong asymptotic freeness still holds. Our proof relies on an interpolation technology recently introduced by (*Invent. Math.* **234** (2023) 419–487).

Keywords: Gaussian unitary ensemble; strong asymptotic freeness; tensor

References

Bandeira, A.S., Boedihardjo, M.T. and van Handel, R. (2023). Matrix concentration inequalities and free probability. *Invent. Math.* **234**. 419–487. MR4635836 https://doi.org/10.1007/s00222-023-01204-6

Belinschi, S. and Capitaine, M. (2022). Strong convergence of tensor products of independent G.U.E. matrices. arXiv:2205.07695v1.

Bordenave, C. and Collins, B. (2019). Eigenvalues of random lifts and polynomials of random permutation matrices. *Ann. of Math. (2)* **190** 811–875. MR4024563 https://doi.org/10.4007/annals.2019.190.3.3

Bordenave, C. and Collins, B. (2023). Norm of matrix-valued polynomials in random unitaries and permutations. arXiv:2304.05714.

Bordenave, C. and Collins, B. (2024). Strong asymptotic freeness for independent uniform variables on compact groups associated to nontrivial representations. *Invent. Math.* **237** 221–273. MR4756991 https://doi.org/10.1007/s00222-024-01259-z

Caspers, M. and Fima, P. (2017). Graph products of operator algebras. *J. Noncommut. Geom.* **11** 367–411. MR3626564 https://doi.org/10.4171/JNCG/11-1-9

Charlesworth, I. and Collins, B. (2021). Matrix models for ε-free independence. *Arch. Math. (Basel)* **116** 585–600. MR4248552 https://doi.org/10.1007/s00013-020-01569-7

Chen, C.-F., Garza-Vargas, J. and van Handel, R. (2024). A new approach to strong convergence II. The classical ensembles. arXiv:2412.00593.

Chen, C.-F., Garza-Vargas, J., Tropp, J.A. and van Handel, R. (2024). A new approach to strong convergence. *Ann. Math.* To appear. arXiv:2405.16026.

Collins, B. and Male, C. (2014). The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér. (4)* **47** 147–163. MR3205602 https://doi.org/10.24033/asens.2211

Haagerup, U. and Thorbjørnsen, S. (2005). A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)* **162** 711–775. MR2183281 https://doi.org/10.4007/annals.2005.162.711

Hayes, B. (2022). A random matrix approach to the Peterson-Thom conjecture. *Indiana Univ. Math. J.* **71** 1243–1297. MR4448584

1350-7265

Magee, M. and Thomas, J. (2023). Strongly convergent unitary representations of right-angled Artin groups. arXiv:2308.00863.

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

Młotkowski, W. (2004). Λ-Free probability. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **7** 27–41. MR2044286 https://doi.org/10.1142/S0219025704001517

Morampudi, S.C. and Laumann, C.R. (2019). Many-body systems with random spatially local interactions. *Phys. Rev. B* **100** 245152. https://doi.org/10.1103/PhysRevB.100.245152

Nica, A. and Speicher, R. (2006). *Lectures on the Combinatorics of Free Probability*. *London Mathematical Society Lecture Note Series* **335**. Cambridge: Cambridge University Press. MR2266879 https://doi.org/10.1017/CBO9780511735127

Parraud, F. (2024). The spectrum of a tensor of random and deterministic matrices. arXiv:2410.04481.

Speicher, R. and Wysoczański, J. (2016). Mixtures of classical and free independence. *Arch. Math. (Basel)* **107** 445–453. MR3552222 https://doi.org/10.1007/s00013-016-0955-6

Voiculescu, D. (1991). Limit laws for random matrices and free products. *Invent. Math.* **104** 201–220. MR1094052 https://doi.org/10.1007/BF01245072

Bernoulli **32**(3), 2026, 1827–1852
https://doi.org/10.3150/25-BEJ1934

Optimal importance sampling for overdamped Langevin dynamics

MARTIN CHAK[1,a], TONY LELIÈVRE[2,3,b], GABRIEL STOLTZ[2,3,c] and URBAIN VAES[3,2,d]

[1] *LJLL, Sorbonne Université, Paris, France,* [a]*martin.chak@unibocconi.it*
[2] *CERMICS, École des Ponts, Champs-sur-Marne, France,* [b]*tony.lelievre@enpc.fr,* [c]*gabriel.stoltz@enpc.fr*
[3] *MATHERIALS project-team, Inria Paris, Paris, France,* [d]*urbain.vaes@inria.fr*

Calculating averages with respect to multimodal probability distributions is often necessary in applications. Markov chain Monte Carlo (MCMC) methods to this end, which are based on time averages along a realization of a Markov process ergodic with respect to the target probability distribution, are usually plagued by a large variance due to the metastability of the process. In this work, we mathematically analyze an importance sampling approach for MCMC methods that rely on the overdamped Langevin dynamics. Specifically, we study an estimator based on an ergodic average along a realization of an overdamped Langevin process for a modified potential. The estimator we consider incorporates a reweighting term in order to rectify the bias that would otherwise be introduced by this modification of the potential. We obtain an explicit expression in dimension 1 for the biasing potential that minimizes the asymptotic variance of the estimator for a given observable, and propose a general numerical approach for approximating the optimal potential in the multi-dimensional setting. We also investigate an alternative approach where, instead of the asymptotic variance for a single given observable, a weighted average of the asymptotic variances corresponding to a class of observables is minimized. Finally, we demonstrate the capabilities of the proposed method by means of numerical experiments.

Keywords: Importance sampling; overdamped Langevin dynamics; Poisson equation; variance reduction

References

Abdulle, A., Pavliotis, G.A. and Vaes, U. (2017). Spectral methods for multiscale stochastic differential equations. *SIAM/ASA J. Uncertain. Quantificat.* **5** 720–761.

Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D. and Stuart, A.M. (2017). Importance sampling: Intrinsic dimension and computational cost. *Statist. Sci.* **32** 405–431.

Allen, M.P. and Tildesley, D.J. (2017). *Computer Simulation of Liquids*, 2nd ed. Oxford University Press, Inc.

Ané, C., Blachère, S., Chafaï, D., Fougères, P., Gentil, I., Malrieu, F., Roberto, C. and Scheffer, G. (2000). *Sur les Inégalités de Sobolev Logarithmiques*. *Panoramas et Synthèses* **10**. Paris: Société Mathématique de France.

Armijo, L. (1966). Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific J. Math.* **16** 1–3.

Asmussen, S.r. and Glynn, P.W. (2007). *Stochastic Simulation: Algorithms and Analysis*. *Stochastic Modelling and Applied Probability* **57**. New York: Springer.

Bakry, D., Barthe, F., Cattiaux, P. and Guillin, A. (2008). A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.* **13** 60–66.

Barzilai, J. and Borwein, J.M. (1988). Two-point step size gradient methods. *IMA J. Numer. Anal.* **8** 141–148.

Cérou, F., Héas, P. and Rousset, M. (2022). Entropy minimizing distributions are worst-case optimal importance proposals. arXiv preprint, arXiv:2212.04292.

Chak, M., Kantas, N., Lelièvre, T. and Pavliotis, G.A. (2023). Optimal friction matrix for underdamped Langevin sampling. *ESAIM: Math. Model. Numer. Anal.* **57** 3335–3371.

Chak, M., Lelièvre, T., Stoltz, G. and Vaes, U. (2026). Supplement to "Optimal importance sampling for overdamped Langevin dynamics." https://doi.org/10.3150/25-BEJ1934SUPP

Chen, Y., Huang, D.Z., Huang, J., Reich, S. and Stuart, A.M. (2024). Efficient, multimodal, and derivative-free Bayesian inference with Fisher-Rao gradient flows. arXiv preprint. arXiv:2406.17263.

Dashti, M. and Stuart, A.M. (2017). The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification* 311–428. Cham: Springer International Publishing.

Dovera, L. and Della Rossa, E. (2011). Multimodal ensemble Kalman filtering using Gaussian mixture models. *Comput. Geosci.* **15** 307–323.

Durmus, A.O. and Eberle, A. (2023). Asymptotic bias of inexact Markov Chain Monte Carlo methods in high dimension. arXiv preprint, arXiv:2108.00682.

Evans, D.J. and Morriss, G.P. (2008). *Statistical Mechanics of Nonequilibrium Liquids*. Cambridge University Press.

García-Portugués, E., Sørensen, M., Mardia, K.V. and Hamelryck, T. (2019). Langevin diffusions on the torus: Estimation and applications. *Stat. Comput.* **29** 1–22.

Glynn, P.W. and Meyn, S.P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

Hesterberg, T.C. (1988). Advances in Importance Sampling. ProQuest LLC, Ann Arbor, MI Thesis (Ph.D.), Stanford University.

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

Kliemann, W. (1987). Recurrence and invariant measures for degenerate diffusions. *Ann. Probab.* **15** 690–707.

Law, K., Stuart, A. and Zygalakis, K. (2015). *Data Assimilation*. *Texts in Applied Mathematics* **62**. Cham: Springer.

Leimkuhler, B. and Matthews, C. (2015). *Molecular Dynamics*. *Interdisciplinary Applied Mathematics* **39**. Cham: Springer.

Leimkuhler, B., Matthews, C. and Stoltz, G. (2016). The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.* **36** 13–79.

Lelièvre, T., Rousset, M. and Stoltz, G. (2010). *Free Energy Computations*. London: Imperial College Press.

Lelièvre, T. and Stoltz, G. (2016). Partial differential equations and stochastic methods in molecular dynamics. *Acta Numer.* **25** 681–880.

Lorenzi, L. (2017). *Analytical Methods for Kolmogorov Equations*. *Monographs and Research Notes in Mathematics*. Boca Raton, FL: CRC Press.

Maruyama, G. and Tanaka, H. (1959). Ergodic property of N-dimensional recurrent Markov processes. *Mem. Fac. Sci., Kyushu Univ., Ser. A* **13** 157–172.

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

Mogensen, P.K. and Riseth, A.N. (2018). Optim: A mathematical optimization package for Julia. *J. Open Sour. Softw.* **3** 615.

Neal, R. (1993). Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems* (S. Hanson, J. Cowan and C. Giles, eds.) **5**. Morgan-Kaufmann.

Owen, A.B. (2013). Monte Carlo Theory, Methods and Examples. https://artowen.su.domains/mc/.

Pavliotis, G.A. (2014). *Stochastic Processes and Applications*. *Texts in Applied Mathematics* **60**. New York: Springer.

Reich, S. and Weissmann, S. (2021). Fokker–Planck particle systems for Bayesian inference: Computational approaches. *SIAM/ASA J. Uncertain. Quantificat.* **9** 446–482.

Rey-Bellet, L. (2006). Ergodic properties of Markov processes. In *Open Quantum Systems II: The Markovian Approach* 1–39. Berlin: Springer.

Richardson, S. and Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B, Methodol.* **59** 731–792.

Robert, C.P. (2007). *The Bayesian Choice*, 2nd ed. *Springer Texts in Statistics*. New York: Springer.

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. New York: Springer-Verlag.

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.

Roberts, G.O. and Tweedie, R.L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**.

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

Roussel, J. and Stoltz, G. (2018). Spectral methods for Langevin dynamics and associated error estimates. *ESAIM: Math. Model. Numer. Anal.* **52** 1051–1083.

Schütte, C. and Sarich, M. (2013). *Metastability and Markov State Models in Molecular Dynamics*. *Courant Lecture Notes in Mathematics* **24**. New York: Courant Institute of Mathematical Sciences. Providence, RI: American Mathematical Society.

Shen, J., Tang, T. and Wang, L.L. (2011). *Spectral Methods*. *Springer Series in Computational Mathematics* **41**. Heidelberg: Springer.

Stuart, A.M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559.

Tuckerman, M. (2010). *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press.

Bernoulli **32**(3), 2026, 1853–1879
https://doi.org/10.3150/25-BEJ1935

Detecting a late changepoint in the preferential attachment model

GIANMARCO BET[1,a], KAY BOGERD[2,b], RUI M. CASTRO[2,c] and REMCO VAN DER HOFSTAD[2,d]

[1]*Department of Mathematics and Computer Science, University of Florence, Florence, Italy,* [a]*gianmarco.bet@unifi.it*
[2]*Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands,* [b]*kaybogerd@hotmail.com,* [c]*rmcastro@tue.nl,* [d]*r.w.v.d.hofstad@tue.nl*

Motivated by the problem of detecting a change in the evolution of a network, we consider the preferential attachment random graph model with a *time-dependent* attachment function. Our goal is to detect whether the attachment mechanism changed over time, based on a single snapshot of the network and without directly observable information about the dynamics. We cast this question as a hypothesis testing problem, where the null hypothesis is a preferential attachment model with a constant affine attachment parameter δ_0, and the alternative hypothesis is a preferential attachment model where the affine attachment parameter changes from δ_0 to δ_1 at an unknown changepoint time τ_n. For our analysis we focus on the regime where δ_0 and δ_1 are fixed, and the changepoint occurs close to the observation time of the network (i.e., $\tau_n = n - cn^\gamma$ with $c > 0$ and $\gamma \in (0,1)$). This corresponds to a rather relevant scenario where we aim to detect the changepoint shortly after it has happened. We present two tests based on the number of vertices with minimal degree, and show that these are asymptotically powerful when $\frac{1}{2} < \gamma < 1$. We conjecture that any test based on the final network snapshot will be powerless when $\gamma < \frac{1}{2}$. The first test we propose requires knowledge of δ_0. The second test is significantly more involved, and does not require the knowledge of δ_0 while still achieving the same asymptotic performance guarantees. Furthermore, we prove that the test statistics for both tests are asymptotically normal, allowing for accurate calibration of the tests. This is demonstrated by numerical experiments, that also illustrate the finite sample test properties.

Keywords: Changepoint detection; preferential attachment model

References

[1] Adamic, L.A., Huberman, B.A., Barabási, A.L., Albert, R., Jeong, H. and Bianconi, G. (2000). Power-law distribution of the world wide web. *Science* **287** 2115. https://doi.org/10.1126/science.287.5461.2115a
[2] Albert, R. and Barabási, A.L. (2000). Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.* **85** 5234–5237. https://doi.org/10.1103/PhysRevLett.85.5234
[3] Baldassarri, S. and Bet, G. (2022). Asymptotic normality of degree counts in a general preferential attachment model. *Markov Process. Related Fields* **28** 577–603.
[4] Banerjee, S., Bhamidi, S. and Carmichael, I. (2023). Fluctuation bounds for continuous time branching processes and evolution of growing trees with a change point. *Ann. Appl. Probab.* **33** 2919–2980. https://doi.org/10.1214/22-AAP1881
[5] Barabási, A.L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. https://doi.org/10.1126/science.286.5439.509
[6] Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Phys. A* **311** 590–614. https://doi.org/10.1016/S0378-4371(02)00736-7
[7] Bet, G., Bogerd, K., Castro, R.M. and van der Hofstad, R. (2026). Supplement to "Detecting a late changepoint in the preferential attachment model." https://doi.org/10.3150/25-BEJ1935SUPP
[8] Bhamidi, S., Jin, J. and Nobel, A. (2018). Change point detection in network models: Preferential attachment and long range dependence. *Ann. Appl. Probab.* **28** 35–78. https://doi.org/10.1214/17-AAP1297

1350-7265

[9] Bhattacharjee, M., Banerjee, M. and Michailidis, G. (2020). Change point estimation in a dynamic stochastic block model. *J. Mach. Learn. Res.* **21** 4330–4388.

[10] Bollobás, B. and Riordan, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24** 5–34. https://doi.org/10.1007/s00493-004-0002-2

[11] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. https://doi.org/10.1002/rsa.1009

[12] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J. (2000). Graph structure in the web. *Comput. Netw.* **33** 309–320. https://doi.org/10.1016/S1389-1286(00)00083-9

[13] Bubeck, S., Devroye, L. and Lugosi, G. (2017). Finding Adam in random growing trees. *Random Structures Algorithms* **50** 158–172. https://doi.org/10.1002/rsa.20649

[14] Bubeck, S., Mossel, E. and Rácz, M.Z. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Trans. Netw. Sci. Eng.* **2** 30–39. https://doi.org/10.1109/TNSE.2015.2397592

[15] Bubeck, S., Eldan, R., Mossel, E. and Rácz, M.Z. (2017). From trees to seeds: On the inference of the seed from large trees in the uniform attachment model. *Bernoulli* **23** 2887–2916. https://doi.org/10.3150/16-BEJ831

[16] Cirkovic, D., Wang, T. and Zhang, X. (2022). Likelihood-based changepoint detection in preferential attachment networks. Preprint. Available at arXiv:2206.01076.

[17] Curien, N., Duquesne, T., Kortchemski, I. and Manolescu, I. (2015). Scaling limits and influence of the seed graph in preferential attachment trees. *J. Éc. Polytech. Math.* **2** 1–34. https://doi.org/10.5802/jep.15

[18] Deijfen, M., van den Esker, H., van der Hofstad, R. and Hooghiemstra, G. (2007). A preferential attachment model with random initial degrees. *Ark. Mat.* **47** 41–72. https://doi.org/10.1007/s11512-007-0067-4

[19] Dommers, S., van der Hofstad, R. and Hooghiemstra, G. (2010). Diameters in preferential attachment graphs. *J. Stat. Phys.* **139** 72–107. https://doi.org/10.1007/s10955-010-9921-z

[20] Douwes, A.S. (2024). Detection of change points and anomalies in Preferential Attachment Models. Master thesis.

[21] Du, H., Gong, S. and Xu, J. (2025). A proof of the changepoint detection threshold conjecture in preferential attachment models. Preprint. Available at arXiv:2502.00514.

[22] Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). On power-law relationships of the internet topology. *Comput. Commun. Rev.* **29** 251–261. https://doi.org/10.1145/316194.316229

[23] Farkas, I., Jeong, H., Vicsek, T., Barabási, A.L. and Oltvai, Z.N. (2003). The topology of the transcription regulatory network in the yeast, Saccharomyces cerevisiae. *Phys. A* **318** 601–612. https://doi.org/10.1016/S0378-4371(02)01731-4

[24] Gao, F. (2011). Modeling and interference of the internet movie database. Master thesis, Eindhoven University of Technology.

[25] Gao, F. and van der Vaart, A. (2017). On the asymptotic normality of estimating the affine preferential attachment network models with random initial degrees. *Stoch. Process. Appl.* **127** 3754–3775. https://doi.org/10.1016/J.SPA.2017.03.008

[26] Jeong, H., Tombor, B., Albert, R., Oltval, Z.N. and Barabási, A.L. (2000). The large-scale organization of metabolic networks. *Nature* **407** 651–654. https://doi.org/10.1038/35036627

[27] Kaddouri, I., Naulet, Z. and Gassiat, É. (2024). On the impossibility of detecting a late change-point in the preferential attachment random graph model. Preprint. Available at arXiv:2407.18685.

[28] Marchand, D.C. and Manolescu, I. (2020). Influence of the seed in affine preferential attachment trees. *Bernoulli* **26** 1665–1705. https://doi.org/10.3150/19-BEJ1152

[29] Middendorf, M., Ziv, E. and Wiggins, C.H. (2005). Inferring network mechanisms: The Drosophila melanogaster protein interaction network. *Proc. Natl. Acad. Sci. USA* **102** 3192–3197. https://doi.org/10.1073/pnas.0409515102

[30] Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98** 404–409. https://doi.org/10.1073/pnas.021544898

[31] Pensky, M. and Zhang, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **13** 678–709. https://doi.org/10.1214/19-EJS1533

[32] Perc, M. (2014). The Matthew effect in empirical data. *J. R. Soc. Interface* **11**. https://doi.org/10.1098/rsif.2014.0378

[33] Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* **4** 131–134. https://doi.org/10.1007/s100510050359
[34] Resnick, S.I. and Samorodnitsky, G. (2016). Asymptotic normality of degree counts in a preferential attachment model. *Adv. Appl. Probab.* **48** 283–299. https://doi.org/10.1017/apr.2016.56
[35] van der Hofstad, R. (2017). *Random Graphs and Complex Networks* **1**. Cambridge University Press. https://doi.org/10.1017/9781316779422
[36] van der Hofstad, R. (2024). *Random Graphs and Complex Networks* **2**. Cambridge University Press. https://doi.org/10.1017/9781316795552
[37] Wang, D., Yu, Y. and Rinaldo, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *Ann. Statist.* **49** 203–232. https://doi.org/10.1214/20-AOS1953
[38] Wang, M., Yu, G. and Yu, D. (2008). Measuring the preferential attachment mechanism in citation networks. *Phys. A* **387** 4692–4698. https://doi.org/10.1016/j.physa.2008.03.017
[39] Wang, H., Tang, M., Park, Y. and Priebe, C.E. (2014). Locality statistics for anomaly detection in time series of graphs. *IEEE Trans. Signal Process.* **62** 703–717. https://doi.org/10.1109/TSP.2013.2294594
[40] Watts, D.J. (1999). *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press.
[41] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393** 440–442. https://doi.org/10.1038/30918
[42] Zhao, Z., Chen, L. and Lin, L. (2019). Change-point detection in dynamic networks via graphon estimation. Preprint. Available at arXiv:1908.01823.

Bernoulli **32**(3), 2026, 1880–1906
https://doi.org/10.3150/25-BEJ1936

Nonparametric estimation of ordinary differential equations: Snake and stubble

CHRISTOF SCHÖTZ[1,2,a]

[1] *Munich Climate Center and Earth System Modelling Group, Department of Aerospace and Geodesy, TUM School of Engineering and Design, Technical University of Munich, Munich, Germany,* [a] *christof.schoetz@tum.de*
[2] *Potsdam Institute for Climate Impact Research, Potsdam, Germany*

We study nonparametric estimation in dynamical systems described by ordinary differential equations (ODEs). Specifically, we focus on estimating the unknown function $f\colon \mathbb{R}^d \to \mathbb{R}^d$ that governs the system dynamics through the ODE $\dot{u}(t) = f(u(t))$, where observations $Y_{j,i} = u_j(t_{j,i}) + \varepsilon_{j,i}$ of solutions u_j of the ODE are made at times $t_{j,i}$ with independent noise $\varepsilon_{j,i}$. We introduce two novel models—the Stubble model and the Snake model—to mitigate the issue of observation location dependence on f, an inherent difficulty in nonparametric estimation of ODE systems. In the Stubble model, we observe many short solutions with initial conditions that adequately cover the domain of interest. Here, we study an estimator based on multivariate local polynomial regression and univariate polynomial interpolation. In the Snake model, we observe few long trajectories that traverse the domain of interest. Here, we study an estimator that combines univariate local polynomial estimation with multivariate polynomial interpolation. For both models, we establish error bounds of order $n^{-\frac{\beta}{2(\beta+1)+d}}$ for β-smooth functions f in an infinite-dimensional function class of Hölder-type and establish minimax optimality for the Stubble model in general and for the Snake model under some conditions via comparison to lower bounds from parallel work.

Keywords: Minimax optimal; nonparametric regression; ordinary differential equations; rate of convergence

References

[1] Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developments and their applications. *Stat. Surv.* **1** 16–55. MR2520413 https://doi.org/10.1214/07-SS014
[2] Arnold, V.I. (2006). *Ordinary Differential Equations*. *Universitext*. Berlin: Springer-Verlag. Translated from the Russian by Roger Cooke, Second printing of the 1992 edition. MR2242407
[3] Brunel, N.J.-B. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electron. J. Stat.* **2** 1242–1267. MR2471285 https://doi.org/10.1214/07-EJS132
[4] Butcher, J.C. (2016). *Numerical Methods for Ordinary Differential Equations*, 3rd ed. Chichester: John Wiley & Sons, Ltd. With a foreword by J. M. Sanz-Serna. MR3559553 https://doi.org/10.1002/9781119121534
[5] Chen, R.T.Q., Rubanova, Y., Bettencourt, J. and Duvenaud, D. (2018). Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. *NIPS'18* 6572–6583. Red Hook, NY, USA: Curran Associates Inc.
[6] Dattner, I. (2021). Differential equations in data analysis. *Wiley Interdiscip. Rev.: Comput. Stat.* **13** Paper No. e1534, 20. MR4336079 https://doi.org/10.1002/wics.1534
[7] Dattner, I. and Klaassen, C.A.J. (2015). Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *Electron. J. Stat.* **9** 1939–1973. MR3391125 https://doi.org/10.1214/15-EJS1053
[8] Donovan, T. (2010). *Replay: The History of Video Games*. Yellow Ant.
[9] Gottwald, G.A. and Reich, S. (2021). Supervised learning from noisy observations: Combining machine-learning techniques with data assimilation. *Phys. D* **423** Paper No. 132911, 15. MR4249157 https://doi.org/10.1016/j.physd.2021.132911

1350-7265

[10] Gugushvili, S. and Klaassen, C.A.J. (2012). $\sqrt{n}$-Consistent parameter estimation for systems of ordinary differential equations: Bypassing numerical integration via smoothing. *Bernoulli* **18** 1061–1098. MR2948913 https://doi.org/10.3150/11-BEJ362

[11] Hartman, P. (2002). *Ordinary Differential Equations*. *Classics in Applied Mathematics* **38**. Philadelphia: Society for Industrial and Applied Mathematics (SIAM). Corrected reprint of the second (1982) edition [Birkhäuser, Boston, MA; MR0658490 (83e:34002)], With a foreword by Peter Bates. MR1929104 https://doi.org/10.1137/1.9780898719222

[12] Heinonen, M., Yildiz, C., Mannerström, H., Intosalmi, J. and Lähdesmäki, H. (2018). Learning unknown ODE models with Gaussian processes. In *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 1959–1968. PMLR.

[13] Lahouel, K., Wells, M., Rielly, V., Lew, E., Lovitz, D. and Jedynak, B.M. (2024). Learning nonparametric ordinary differential equations from noisy data. *J. Comput. Phys.* **507** Paper No. 112971, 22. MR4729758 https://doi.org/10.1016/j.jcp.2024.112971

[14] Marzouk, Y., Ren, Z., Wang, S. and Zech, J. (2024). Distribution learning via neural differential equations: A nonparametric statistical perspective. *J. Mach. Learn. Res.* **25** Paper No. [232], 61. https://doi.org/4796493

[15] McGoff, K., Mukherjee, S. and Pillai, N. (2015). Statistical inference for dynamical systems: A review. *Stat. Surv.* **9** 209–252. MR3422438 https://doi.org/10.1214/15-SS111

[16] Olver, P.J. (2006). On multivariate interpolation. *Stud. Appl. Math.* **116** 201–240. MR2200737 https://doi.org/10.1111/j.1467-9590.2006.00335.x

[17] Qi, X. and Zhao, H. (2010). Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. *Ann. Statist.* **38** 435–481. MR2589327 https://doi.org/10.1214/09-AOS724

[18] Ramsay, J. and Hooker, G. (2017). *Dynamic Data Analysis: Modeling Data with Differential Equations*. *Springer Series in Statistics*. New York: Springer. MR3645102 https://doi.org/10.1007/978-1-4939-7190-9

[19] Rasmussen, C.E. and Williams, C.K.I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001

[20] Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.* **48** 1875–1897. MR4134774 https://doi.org/10.1214/19-AOS1875

[21] Schötz, C. (2026). Supplement to "Nonparametric estimation of ordinary differential equations: Snake and stubble." https://doi.org/10.3150/25-BEJ1936SUPP

[22] Schötz, C. and Siebel, M. (2024). Lower bounds for nonparametric estimation of ordinary differential equations. https://doi.org/10.48550/arXiv.2407.14993

[23] Strogatz, S. (2024). *Nonlinear Dynamics and Chaos*, 3rd ed. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429398490

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

[25] Xiao, L. (2019). Asymptotic theory of penalized splines. *Electron. J. Stat.* **13** 747–794. MR3925516 https://doi.org/10.1214/19-ejs1541

Bernoulli **32**(3), 2026, 1907–1921
https://doi.org/10.3150/25-BEJ1937

Sample path properties of the fractional Wiener–Weierstrass bridge

ALEXANDER SCHIED[1,a] and ZHENYUAN ZHANG[2,b]

[1]*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada,* [a]*aschied@uwaterloo.ca*
[2]*Department of Mathematics, Stanford University, Stanford, USA,* [b]*zzy@stanford.edu*

Fractional Wiener–Weierstrass bridges are a class of Gaussian processes that arise from replacing the trigonometric function in the construction of classical Weierstrass functions by a fractional Brownian bridge. We investigate the sample path properties of such processes, including local and uniform moduli of continuity, Φ-variation, Hausdorff dimension, nowhere differentiability, and location of the maximum. Our analysis relies heavily on upper and lower bounds of fractional integrals, where we establish a novel improvement of the classical Hardy–Littlewood inequality for fractional integrals of a special class of step functions.

Keywords: Fractional Wiener–Weierstrass bridge; Hardy–Littlewood inequality for fractional integrals; Hausdorff dimension; moduli of continuity; Φ-variation

References

Adler, R.J. (1977). Hausdorff dimension and Gaussian fields. *Ann. Probab.* **5** 145–151.
Adler, R.J. (1981). *The Geometry of Random Fields*. SIAM.
Basse-O'Connor, A. and Weber, M. (2016). On the Φ-variation of stochastic processes with exponential moments. *Trans. Lond. Math. Soc.* **3** 1–27.
Bayer, C., Friz, P.K., Fukasawa, M., Gatheral, J., Jacquier, A. and Rosenbaum, M., eds. (2024). *Rough Volatility*. Philadelphia, PA: Society for Industrial and Applied Mathematics. MR4813430
Berman, S.M. and Getoor, R. (1973). Local nondeterminism and local times of Gaussian processes. *Indiana Univ. Math. J.* **23** 69–94.
Cont, R. and Perkowski, N. (2019). Pathwise integration and change of variable formulas for continuous paths with arbitrary regularity. *Trans. Amer. Math. Soc. Ser. B* **6** 161–186. MR3937343 https://doi.org/10.1090/btran/34
de Lima, A. and Smania, D. (2019). Central limit theorem for generalized Weierstrass functions. *Stoch. Dyn.* **19** 1950002.
Dudley, R.M. and Norvaiša, R. (2011). *Concrete Functional Calculus*. Springer.
Friz, P.K. and Hairer, M. (2020). *A Course on Rough Paths*. Springer.
Friz, P.K. and Victoir, N.B. (2010). *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications 120*. Cambridge University Press.
Galkin, O.E. and Galkina, S.Y. (2015). On properties of functions in exponential Takagi class. *Ufa Math. J.* **7** 28–37. MR3430688 https://doi.org/10.13108/2015-7-3-28
Galkina, S.Y. (1992). Fourier-Haar coefficients of functions of bounded variation. *Mat. Zametki* **51** 42–54, 160. MR1165279 https://doi.org/10.1007/BF01229431
Gamkrelidze, N.G. (1984). Modulus of continuity of the Weierstrass function. *Math. Notes USSR* **36** 510–512.
Gasbarra, D., Sottinen, T. and Valkeila, E. (2007). Gaussian bridges. In *Stochastic Analysis and Applications*. *Abel Symp.* **2** 361–382. Berlin: Springer. MR2397795 https://doi.org/10.1007/978-3-540-70847-6_15
Han, X. and Schied, A. (2022). Step roots of Littlewood polynomials and the extrema of functions in the Takagi class. *Math. Proc. Cambridge Philos. Soc.* **173** 591–618.
Han, X. and Schied, A. (2025). The roughness exponent and its model-free estimation. *Ann. Appl. Probab.* **35** 1049–1082.

Imkeller, P., Menoukeu Pamen, O., dos Reis, G. c. and Réveillac, A. (2022). Rough Weierstrass functions and dynamical systems: The smoothness of the SBR measure. *Pure Appl. Funct. Anal.* **7** 1273–1305. MR4493560

Izumi, M., Izumi, S.-i. and Kahane, J.-P. (1965). Théorèmes élémentaires sur les séries de Fourier lacunaires. *J. Anal. Math.* **14** 235–246.

Kahane, J.-P. (1959). Sur l'exemple, donné par M. de Rham, d'une fonction continue sans dérivée. *Enseign. Math.* **5** 53–57.

Kawada, T. and Kôno, N. (1973). On the variation of Gaussian processes. In *Proceedings of the Second Japan-Ussr Symposium on Probability Theory* 176–192. Springer.

Keller, G. (2017). A simpler proof for the dimension of the graph of the classical Weierstrass function. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 169–181.

Kubilius, K., Mishura, Y. and Ralchenko, K. (2017). *Parameter Estimation in Fractional Diffusion Models 8*. Springer.

Marcus, M.B. and Rosen, J. (2006). *Markov Processes, Gaussian Processes, and Local Times*. *Cambridge Studies in Advanced Mathematics* **100**. Cambridge: Cambridge University Press. MR2250510 https://doi.org/10.1017/CBO9780511617997

Meerschaert, M., Wang, W. and Xiao, Y. (2013). Fernique-type inequalities and moduli of continuity for anisotropic Gaussian random fields. *Trans. Amer. Math. Soc.* **365** 1081–1107.

Mishura, Y. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. *Lecture Notes in Mathematics* **1929**. Berlin: Springer-Verlag. MR2378138 https://doi.org/10.1007/978-3-540-75873-0

Mishura, Y. and Schied, A. (2019). On (signed) Takagi–Landsberg functions: pth variation, maximum, and modulus of continuity. *J. Math. Anal. Appl.* **473** 258–272. MR3912820 https://doi.org/10.1016/j.jmaa.2018.12.047

Norvaiša, R. (2008). Φ-Variation of a bifractional Brownian motion. *Lith. Math. J.* **48** 418–426.

Nourdin, I. (2012). *Selected Aspects of Fractional Brownian Motion 4*. Springer.

Pitt, L.D. (1978). Local times for Gaussian vector fields. *Indiana Univ. Math. J.* **27** 309–330.

Qi, H. and Yan, L. (2018). A law of iterated logarithm for the subfractional Brownian motion and an application. *J. Inequal. Appl.* **2018** 1–18.

Ren, H. and Shen, W. (2021). A dichotomy for the Weierstrass-type functions. *Invent. Math.* **226** 1057–1100.

Schied, A. and Zhang, Z. (2020). On the p-th variation of a class of fractal functions. *Proc. Amer. Math. Soc.* **148** 5399–5412.

Schied, A. and Zhang, Z. (2024). Weierstrass bridges. *Trans. Amer. Math. Soc.* **377** 2947–2989.

Schied, A. and Zhang, Z. (2026). Supplement to "Sample path properties of the fractional Wiener–Weierstrass bridge." https://doi.org/10.3150/25-BEJ1937SUPP

Shen, W. (2018). Hausdorff dimension of the graphs of the classical Weierstrass functions. *Math. Z.* **289** 223–266.

Tudor, C.A. and Xiao, Y. (2007). Sample path properties of bifractional Brownian motion. *Bernoulli* **13** 1023–1052.

Wu, D. and Xiao, Y. (2007). Geometric properties of fractional Brownian sheets. *J. Fourier Anal. Appl.* **13** 1–37.

Xiao, Y. (2006). Properties of local-nondeterminism of Gaussian and stable random fields and their applications. *Ann. Fac. Sci. Toulouse Math.* **15** 157–193.

Xiao, Y. (2007). Strong local nondeterminism of Gaussian random fields and its applications. In *Asymptotic Theory in Probability and Statistics with Applications* 136–176. Higher Education Press.

Bernoulli **32**(3), 2026, 1922–1946
https://doi.org/10.3150/25-BEJ1938

Asymptotic bias reduction of maximum likelihood estimates via penalized likelihoods with differential geometry

MASAYO Y. HIROSE[1,a] and SHUHEI MANO[2,b]

[1] *Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan,* [a]*masayo@imi.kyushu-u.ac.jp*
[2] *Department of Fundamental Statistical Mathematics, The Institute of Statistical Mathematics, Tachikawa, Japan,* [b]*smano@ism.ac.jp*

A method for asymptotic bias reduction of maximum likelihood estimates of generic estimands is developed. The estimator is realized as a plug-in estimator, where the parameter maximizes the penalized likelihood with a penalty function that satisfies a quasi-linear partial differential equation of the first order. The integration of the partial differential equation with the aid of differential geometry is discussed. Applications to generalized linear models, linear mixed-effects models, and a location-scale family are presented.

Keywords: Bias reduction; information geometry; Jeffreys prior; partial differential equation; plug-in estimator; shrinkage

References

Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*. Providence, RI: Amer. Math. Soc.
Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.
Copas, J.B. (1988). Binary regression models for contaminated data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **50** 225–65.
Courant, R. and Hilbert, D. (1962). *Methods of Mathematical Physics. Vol II*. New York: Wiley-Interscience.
Cox, D., Little, J. and O'Shea, D. (2007). *Ideals, Varieties, Algorithms*, 3rd ed. New York: Springer.
Cox, D.R. and Snell, E.J. (1968). A general definition of residuals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **30** 248–275.
Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika* **59** 335–347.
Eguchi, S. and Yanagimoto, T. (2008). Asymptotical improvement of maximum likelihood estimators on Kullback-Leibler loss. *J. Statist. Plann. Inference* **138** 3502–3511.
Elgmati, E., Fiaccone, R.L., Henderson, R. and Matthews, J.N.S. (2015). Penalized logistic regression and dynamic prediction for discrete-time recurrent event data. *Lifetime Data Anal.* **21** 542–560.
Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80** 27–38.
Ghosh, M. and Liu, R. (2011). Moment matching priors. *Sankhyā* **73-A** 185–201.
Hartigan, J.A. (1964). Invariant prior distributions. *Ann. Statist.* **35** 836–845.
Helgason, S. (1984). *Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions*. Orlando: Academic Press.
Hirose, Y.M. and Lahiri, P. (2018). Estimating variance of random effects to solve multiple problems simultaneously. *Ann. Statist.* **46** 1721–1741.
Hirose, Y.M. and Lahiri, P. (2021). Multi-goal prior selection: A way to reconcile Bayesian and classical approaches for random effects models. *J. Amer. Statist. Assoc.* **116** 1487–1497.
Hirose, M.Y., Mano, S. (2026). Supplement to "Asymptotic bias reduction of maximum likelihood estimates via penalized likelihoods with differential geometry." https://doi.org/10.3150/25-BEJ1938SUPP

1350-7265

Kobayashi, S. and Nomizu, K. (1963). *Foundations of Differential Geometry, Vol. I.* New York: Wiley and Sons.
Komaki, F. (1996). On asymptotic properties of predictive distributions. *Biometrika* **83** 299–313.
Kosmidis, I. (2014). Bias in parametric estimation: Reduction and useful side-effects. *Wiley Interdiscip. Rev.: Comput. Stat.* **6** 185–196.
Kosmidis, I. and Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika* **96** 793–804.
Lauritzen, S.L. (1987). Statistical manifolds. In *Differential Geometry in Statistical Inference*. *IMS Lecture Notes-Monograph Series* **10** 163–216. Hayward, CA: Inst. Math. Statist.
Lee, J.M. (2002). *Introduction to Smooth Manifolds*. New York: Springer.
Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. New York: Springer.
McCullagh, P. (1987). *Tensor Methods in Statistics*. New York: Chapman and Hall.
McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. New York: Chapman and Hall.
Morita, S. (2001). *Geometry of Differential Forms*. Providence, RI: Amer. Math. Soc.
Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32.
Nomizu, K. and Sasaki, T. (1994). *Affine Differential Geometry*. Cambridge: Cambridge Univ. Press.
Ratcliffe, J.G. (2019). *Foundations of Hyperbolic Manifolds*, 3rd ed. Cham: Springer.
Riesz, M. (1949). L'integrale de Riemann-Liouville et le problème de Cauchy. *Acta Math.* **81** 1–223.
Takeuchi, J. and Amari, S. (2005). α-Parallel prior and its properties. *IEEE Trans. Inf. Theory* **51** 1011–1023.

Bernoulli **32**(3), 2026, 1947–1974
https://doi.org/10.3150/25-BEJ1939

Regularized e-processes: Anytime valid inference with knowledge-based efficiency gains

RYAN MARTIN[a]

Department of Statistics, North Carolina State University, Raleigh, USA, [a]*rgmarti3@ncsu.edu*

Classical statistical methods have theoretical justification when the sample size is predetermined. In applications, however, sample sizes are often data-dependent rather than predetermined. The aforementioned methods aren't reliable in this latter case, hence the recent interest in e-processes and methods that are anytime valid, i.e., reliable for any dynamic data-collection plan. But if the investigator has relevant-yet-incomplete prior information about the quantity of interest, then there's an opportunity for efficiency gain. This paper proposes a *regularized e-process* framework featuring a knowledge-based, imprecise-probabilistic regularization with improved efficiency. A generalized version of Ville's inequality is established, ensuring that inference based on the regularized e-process is anytime valid in a novel, knowledge-dependent sense. Regularized e-processes also facilitate possibility-theoretic uncertainty quantification with strong frequentist-like calibration properties and other Bayesian-like properties: satisfies the likelihood principle, avoids sure-loss, and offers formal decision-making with reliability guarantees.

Keywords: Credal set; decision-making; e-value; possibility theory; uncertainty quantification

References

Aliprantis, C.D. and Border, K.C. (2006). *Infinite Dimensional Analysis*, 3rd ed. Berlin: Springer. A hitchhiker's guide.
Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* **533** 452–454.
Balch, M.S., Martin, R. and Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proc. R. Soc. A* **475** 2018.0565.
Basu, D. (1975). Statistical information and likelihood. *Sankhyā Ser. A* **37** 1–71. Discussion and correspondance between Barnard and Basu.
Berger, J. and Wolpert, R. (1984). *The Likelihood Principle*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **6**. Hayward, CA: Institute of Mathematical Statistics.
Birnbaum, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57** 269–326.
Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J. and Wu, H. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nat. Hum. Behav.* **2** 637–644.
Cella, L. and Martin, R. (2023). Possibility-theoretic statistical inference offers performance and probativeness assurances. *Internat. J. Approx. Reason.* **163** 109060.
Couso, I., Montes, S. and Gil, P. (2001). The necessity of the strong α-cuts of a fuzzy set. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems* **9** 249–262.
Denœux, T. (2019). Decision-making with belief functions: A review. *Internat. J. Approx. Reason.* **109** 87–110.
Dey, N., Martin, R. and Williams, J.P. (2024). Generalized universal inference on risk minimizers. arXiv:2402.00202.
Dubois, D. (2006). Possibility theory and statistical reasoning. *Comput. Statist. Data Anal.* **51** 47–69.
Dubois, D. and Prade, H. (1988). *Possibility Theory*. New York: Plenum Press.
Dubois, D., Foulloy, L., Mauris, G. and Prade, H. (2004). Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliab. Comput.* **10** 273–297.

Durrett, R. (2010). *Probability: Theory and Examples*, 4th ed. *Cambridge Series in Statistical and Probabilistic Mathematics* **31**. Cambridge: Cambridge University Press.

Good, I.J. (1950). *Probability and the Weighing of Evidence*. London: Hafner Publishing Co. New York: Charles Griffin & Co., Ltd.

Grünwald, P.D. (2023). The e-posterior. *Philos. Trans. - Royal Soc. A* **381** Paper No. 20220146, 21.

Grünwald, P., de Heide, R. and Koolen, W. (2024). Safe testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1091–1128.

Hose, D. (2022). Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering. PhD thesis, University of Stuttgart. https://dominikhose.github.io/dissertation/diss_dhose.pdf.

Kass, R.E. (1992). Bayes factors in practice. *Amer. Statist.* **42** 551–560.

Kass, R.E. and Greenhouse, J.B. (1989). Comment on 'investigating therapies of potentially great benefit: ECMO' by Ware (1989). *Statist. Sci.* **4** 310–317.

Levi, I. (1980). *The Enterprise of Knowledge*. Boston: The MIT Press.

Martin, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *Internat. J. Approx. Reason.* **113** 39–73.

Martin, R. (2021). An imprecise-probabilistic characterization of frequentist statistical inference. arXiv:2112.10904.

Martin, R. (2022). Valid and efficient imprecise-probabilistic inference with partial priors, II. General framework. arXiv:2211.14567.

Martin, R. (2023). Valid and efficient imprecise-probabilistic inference with partial priors, III. Marginalization. arXiv:2309.13454.

Martin, R. (2024a). A possibility-theoretic solution to Basu's Bayesian–frequentist via media. *Sankhya A* **86** 43–70.

Martin, R. (2024b). Which statistical hypotheses are afflicted by false confidence? In *BELIEF 2024* (Y. Bi, A.-L. Jousselme and T. Denoeux, eds.). *Lecture Notes in Artificial Intelligence* **14909** 140–149. Switzerland: Springer Nature.

Martin, R. (2025). Possibilistic inferential models: a review. arXiv:2507.09007.

Martin, R. (2026). Supplement to "Regularized e-processes: anytime valid inference with knowledge-based efficiency gains." https://doi.org/10.3150/25-BEJ1939SUPP

Martin, R. and Liu, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.* **108** 301–313.

Martin, R. and Liu, C. (2015). *Inferential Models*. *Monographs on Statistics and Applied Probability* **147**. Boca Raton, FL: CRC Press.

Mayo, D.G. (2018). *Statistical Inference as Severe Testing*. Cambridge: Cambridge University Press.

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature* **506** 150–152.

Ramdas, A. and Manole, T. (2023). Randomized and exchangeable improvements of Markov's, Chebyshev's and Chernoff's inequalities. arXiv:2304.02611.

Ramdas, A., Ruf, J., Larsson, M. and Koolen, W. (2022). Admissible anytime-valid sequential inference must rely on nonnegative martingales. arXiv:2009.03167.

Ramdas, A., Grünwald, P., Vovk, V. and Shafer, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statist. Sci.* **38** 576–601.

Reid, N. and Cox, D.R. (2015). On some principles of statistical inference. *Int. Stat. Rev.* **83** 293–308.

Sellke, T., Bayarri, M.J. and Berger, J.O. (2001). Calibration of p values for testing precise null hypotheses. *Amer. Statist.* **55** 62–71.

Shackle, G.L.S. (1961). *Decision Order and Time in Human Affairs*. Cambridge: Cambridge University Press.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, N.J.: Princeton University Press.

Shafer, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *J. R. Stat. Soc., Ser. A* **184** 407–478.

Shafer, G. and Vovk, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. Hoboken, NJ: Wiley.

Shafer, G., Shen, A., Vereshchagin, N. and Vovk, V. (2011). Test martingales, Bayes factors and p-values. *Statist. Sci.* **26** 84–101.

Troffaes, M.C.M. and de Cooman, G. (2014). *Lower Previsions*. *Wiley Series in Probability and Statistics*. Chichester: John Wiley & Sons, Ltd.

Turner, R.J. and Grünwald, P.D. (2023). Exact anytime-valid confidence intervals for contingency tables and beyond. *Statist. Probab. Lett.* **198** Paper No. 109835, 8.

Varadarajan, V.S. (1963). Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109** 191–220.

Vovk, V.G. (1993). A logic of probability, with application to the foundations of statistics. *J. Roy. Statist. Soc. Ser. B, Methodol.* **55** 317–351. With discussion and a reply by the author.

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. *Monographs on Statistics and Applied Probability* **42**. London: Chapman & Hall Ltd.

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *J. Roy. Statist. Soc. Ser. B, Methodol.* **58** 3–57. With discussion and a reply by the author.

Ware, J.H. (1989). Investigating therapies of potentially great benefit: ECMO. *Statist. Sci.* **4** 298–340. With comments and a rejoinder by the author.

Wasserman, L., Ramdas, A. and Balakrishnan, S. (2020). Universal inference. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890.

Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* **1** 3–28.

Bernoulli **32**(3), 2026, 1975–2000
https://doi.org/10.3150/25-BEJ1940

Misspecified Bernstein–von Mises theorem for hierarchical models

GEERTEN KOERS[1,a], BOTOND SZABÓ[2,c] and AAD VAN DER VAART[1,b]

[1] *DIAM, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, Netherlands,* [a]*g.j.koers@tudelft.nl,* [b]*a.w.vandervaart@tudelft.nl*

[2] *Department of Data Sciences and BIDSA, Bocconi University, Via Roentgen 1 20136, Milan, Italy,* [c]*botond.szabo@unibocconi.it*

We derive a Bernstein–von Mises theorem in the context of misspecified, non-i.i.d., hierarchical models parametrised by a finite-dimensional parameter of interest. We apply our results to hierarchical models containing non-linear operators, including the squared integral operator, and PDE-constrained inverse problems. More specifically, we consider the elliptic, time-independent Schrödinger equation with parametric boundary condition and general parabolic PDEs with parametric potential and boundary constraints. Our theoretical results are complemented with a numerical analysis of synthetic data sets, considering both the square integral operator and the Schrödinger equation.

Keywords: Bayesian estimation; Bernstein–von Mises; hierarchical model; misspecification; parametric model; posterior distribution

References

Alsing, J., Heavens, A., Jaffe, A., Kiessling, A., Wandelt, B. and Hoffmann, T. (2016). Hierarchical cosmic shear power spectrum inference. *Mon. Not. R. Astron. Soc.* **455** 4452–4466. https://doi.org/10.1093/mnras/stv2501

Bochkina, N. (2019). Bernstein–von Mises theorem and misspecified models: A review. In *Foundations of Modern Statistics* 355–380.

Bowman, F., Caffo, B., Bassett, S. and Kilts, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* **39** 146–156. https://doi.org/10.1016/j.neuroimage.2007.08.012

Castillo, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99. https://doi.org/10.1007/s00440-010-0316-5

Castillo, I. and Nickl, R. (2014). On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. https://doi.org/10.1214/14-AOS1246

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

Cornish, N. and Littenberg, T. (2015). Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical Quantum Gravity* **32** 135012. https://doi.org/10.1088/0264-9381/32/13/135012

Doob, J. (1949). Application of the theory of martingales. In *Le Calcul des probabilités et ses applications* **13** 23–27.

Durante, D., Pozza, F. and Szabo, B. (2024). Skewed Bernstein–von Mises theorem and skew-modal approximations. *Ann. Statist.* **52** 2714–2737.

Eriksen, H., O'Dwyer, I., Jewell, J., Wandelt, B., Larson, D., Górski, K., Levin, S., Banday, A. and Lilje, P. (2004). Power spectrum estimation from high-resolution maps by Gibbs sampling. *Astrophys. J., Suppl. Ser.* **155** 227. https://doi.org/10.1086/425219

Ezquiaga, J. and Zumalacárregui, M. (2018). Dark energy in light of multi-messenger gravitational-wave astronomy. *Front. Astron. Space Sci.* **5** 44. https://doi.org/10.3389/fspas.2018.00044

Feehan, P., Gong, R. and Song, J. (2015). Feynman-Kac Formulas for Solutions to Degenerate Elliptic and Parabolic Boundary-Value and Obstacle Problems with Dirichlet Boundary Conditions. ArXiv E-prints.

Freidlin, M. (1985). *Functional Integration and Partial Differential Equations*. *Annals of Mathematics Studies* **109**. Princeton, NJ: Princeton University Press. MR833742 https://doi.org/10.1515/9781400881598

Haralambie, L. (2011). On the Bernstein-von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. https://doi.org/10.1214/11-EJS611

Kekkonen, H. (2022). Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem. *Inverse Probl.* **38** 035002. https://doi.org/10.1088/1361-6420/ac4839

Kleijn, B. and van der Vaart, A. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electron. J. Stat.* **6** 354–381. https://doi.org/10.1214/12-EJS675

Koers, G., Szabó, B. and van der Vaart, A. (2023). Linear methods for non-linear inverse problems. ArXiv preprint.

Koers, G., Szabó, B. and van der Vaart, A. (2026). Supplement to "Misspecified Bernstein–von Mises theorem for hierarchical models." https://doi.org/10.3150/25-BEJ1940SUPP

Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*, 1st ed. *Springer Series in Statistics*. New York: Springer. https://doi.org/10.1007/978-1-4612-4946-7

Lehmann, E. (1983). *Theory of Point Estimation*, 1st ed. *Probability and Mathematical Statistics*. New York: John Wiley & Sons.

Magnus, J. and Neudecker, H. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 1st ed. *Probability and Statistics. Hoboken*. John Wiley & Sons. https://doi.org/10.1002/9781119541219

Marinucci, D. (2004). Testing for non-Gaussianity on cosmic microwave background radiation: A review. *Statist. Sci.* **19** 294–307. https://doi.org/10.1214/088342304000000783

Monard, F., Nickl, R. and Paternain, G. (2021). Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. *Ann. Statist.* **49** 3255–3298. https://doi.org/10.1214/21-aos2082

Monthus, C. and Garel, T. (2011). A critical Dyson hierarchical model for the Anderson localization transition. *J. Stat. Mech. Theory Exp.* **2011** P05005. https://doi.org/10.1088/1742-5468/2011/05/P05005

Nickl, R. (2018). Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc.* **22** 2697–2750. https://doi.org/10.4171/JEMS/975

Nickl, R. (2023). *Bayesian Non-linear Statistical Inverse Problems*. *Zurich Lectures in Advanced Mathematics*. Berlin: EMS Press. MR4604099 https://doi.org/10.4171/zlam/30

Nickl, R., van de Geer, S. and Wang, S. (2020). Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA J. Uncertain. Quantificat.* **8** 374–413. https://doi.org/10.1137/18M1236137

Sellentin, E., Heymans, C. and Harnois-Déraps, J. (2018). The skewed weak lensing likelihood: Why biases arise, despite data and theory being sound. *Mon. Not. R. Astron. Soc.* **477** 4879–4895. https://doi.org/10.1093/mnras/sty988

van der Vaart, A. (1998). *Asymptotic Statistics*, 1st ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511802256

Wikle, C. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **84** 1382–1394. https://doi.org/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2

Bernoulli **32**(3), 2026, 2001–2021
https://doi.org/10.3150/25-BEJ1941

Universality of estimators for high-dimensional linear models with block dependency

TOSHIKI TSUDA[1,a] and MASAAKI IMAIZUMI[2,b]

[1]*Yale University, New Haven, CT 06520, USA,* [a]*toshiki.tsuda@yale.edu*
[2]*The University of Tokyo / Kyoto University / RIKEN AIP, Tokyo, Japan,* [b]*imaizumi@g.ecc.u-tokyo.ac.jp*

We study the universality property of estimators for high-dimensional linear models, which implies that the distribution of estimators is independent of whether the covariates follow a Gaussian distribution. Recent developments in high-dimensional statistics typically require covariates to strictly follow a Gaussian distribution to precisely characterize the properties of estimators. To relax this Gaussianity requirement, the existing literature has examined conditions under which estimators achieve universality. In particular, independence among the elements of the high-dimensional covariates has played a critical role. In this study, we focus on high-dimensional linear models with covariates exhibiting block dependence, where covariate elements can only be dependent within each block, and show that estimators for such models retain universality. Specifically, we prove that the distribution of estimators with Gaussian covariates can be approximated by the distribution of estimators with non-Gaussian covariates having the same moments under block dependence. To establish this result, we develop a generalized Lindeberg principle suitable for handling block dependencies and derive new error bounds for correlated covariate elements. We further demonstrate the universality result across several different estimators.

Keywords: Dependency; high-dimension; linear models; robust estimators; universality

References

Akhtiamov, D., Ghane, R., Varma, N.K., Hassibi, B. and Bosch, D. (2025). A novel Gaussian min-max theorem and its applications. *IEEE Trans. Inf. Theory*.

Barbier, J., Krzakala, F., Macris, N., Miolane, L. and Zdeborová, L. (2019). Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. Natl. Acad. Sci.* **116** 5451–5460.

Bayati, M. and Montanari, A. (2011a). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785.

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

Bean, D., Bickel, P.J., El Karoui, N. and Yu, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci.* **110** 14563–14568.

Bellec, P.C. (2025). Observable adjustments in single-index models for regularized M-estimators with bounded p/n. *Ann. Statist.* **53** 531–560.

Celentano, M., Montanari, A. and Wei, Y. (2023). The lasso with general Gaussian designs with applications to hypothesis testing. *Ann. Statist.* **51** 2194–2220.

Chang, J., Chen, X. and Wu, M. (2024). Central limit theorems for high dimensional dependent data. *Bernoulli* **30** 712–742.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597.

Chernozhukov, V., Chetverikov, D., Kato, K. and Koike, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586.

Dobriban, E. and Wager, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *Ann. Statist.* **46** 247–279.

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

Donoho, D. and Montanari, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969.

Dudeja, R., Lu, Y.M. and Sen, S. (2023). Universality of approximate message passing with semirandom matrices. *Ann. Probab.* **51** 1616–1683.

Dudeja, R., Sen, S. and Lu, Y.M. (2024). Spectral universality in regularized linear regression with nearly deterministic sensing matrices. *IEEE Trans. Inf. Theory*.

El Karoui, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175.

El Karoui, N., Bean, D., Bickel, P.J., Lim, C. and Yu, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci.* **110** 14557–14562.

Fan, Z. (2022). Approximate message passing algorithms for rotationally invariant matrices. *Ann. Statist.* **50** 197–224.

Feng, O.Y., Venkataramanan, R., Rush, C., Samworth, R.J. et al. (2022). A unifying tutorial on approximate message passing. *Found. Trends Mach. Learn.* **15** 335–536.

Goldt, S., Loureiro, B., Reeves, G., Krzakala, F., Mézard, M. and Zdeborová, L. (2022). The Gaussian equivalence of generative models for learning with shallow neural networks. In *Math. Sci. Mach. Learn* 426–471.

Han, Q. (2025). Entrywise dynamics and universality of general first order methods. *Ann. Statist.* **53** 1783–1807.

Han, Q. and Shen, Y. (2023). Universality of regularized regression estimators in high dimensions. *Ann. Statist.* **51** 1799–1823.

Hastie, T., Montanari, A., Rosset, S. and Tibshirani, R.J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986.

Hu, H. and Lu, Y.M. (2022). Universality laws for high-dimensional learning with random features. *IEEE Trans. Inf. Theory* **69** 1932–1964.

Javanmard, A. and Montanari, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144.

Koehler, F., Zhou, L., Sutherland, D.J. and Srebro, N. (2021). Uniform convergence of interpolators: Gaussian width, norm bounds and benign overfitting. *Adv. Neural Inf. Process. Syst.* **34** 20657–20668.

Korada, S.B. and Montanari, A. (2011). Applications of the Lindeberg principle in communications and statistical learning. *IEEE Trans. Inf. Theory* **57** 2440–2450.

Lahiry, S. and Sur, P. (2024). Universality in block dependent linear models with applications to nonlinear regression. *IEEE Trans. Inf. Theory* **70** 8975–9000.

Liang, T. and Sur, P. (2022). A precise high-dimensional asymptotic theory for boosting and minimum-l1-norm interpolated classifiers. *Ann. Statist.* **50** 1669–1695.

Lopes, M.E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near 1/n rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513.

Loureiro, B., Gerbelot, C., Cui, H., Goldt, S., Krzakala, F., Mezard, M. and Zdeborová, L. (2021). Learning curves of generic features maps for realistic datasets with a teacher-student model. *Adv. Neural Inf. Process. Syst.* **34** 18137–18151.

Marchenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.* **114** 507–536.

Mezard, M. and Montanari, A. (2009). *Information, Physics, and Computation*. Oxford University Press.

Mézard, M., Parisi, G. and Virasoro, M.A. (1987). *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications 9*. World Scientific Publishing Company.

Miolane, L. and Montanari, A. (2021). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. *Ann. Statist.* **49** 2313–2335.

Mondelli, M. and Venkataramanan, R. (2021). Approximate message passing with spectral initialization for generalized linear models. *Artif. Intell. Stat.* 397–405.

Montanari, A. and Nguyen, P.-M. (2017). Universality of the elastic net error. In *IEEE Int. Symp. Inf. Theory* 2338–2342.

Montanari, A. and Saeed, B.N. (2022). Universality of empirical risk minimization. In *Conf. Learn. Theory* 4310–4312.

Montanari, A., Ruan, F., Sohn, Y. and Yan, J. (2025). The generalization error of max-margin linear classifiers: Benign overfitting and high dimensional asymptotics in the overparametrized regime. *Ann. Statist.* **53** 822–853.

Rangan, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In *IEEE Int. Symp. Inf. Theory* 2168–2172.

Salehi, F., Abbasi, E. and Hassibi, B. (2019). The impact of regularization on high-dimensional logistic regression. *Adv. Neural Inf. Process. Syst.* **32** 12005–12015.

Sawaya, K., Uematsu, Y. and Imaizumi, M. (2023). Moment-based adjustments of statistical inference in high-dimensional generalized linear models. arXiv preprint. Available at arXiv:2305.17731.

Sawaya, K., Uematsu, Y. and Imaizumi, M. (2024). High-dimensional single-index models: link estimation and marginal inference. arXiv preprint. Available at arXiv:2404.17812.

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

Sur, P. and Candès, E.J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci.* **116** 14516–14525.

Sur, P., Chen, Y. and Candès, E.J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probab. Theory Related Fields* **175** 487–558.

Thrampoulidis, C., Abbasi, E. and Hassibi, B. (2018). Precise error analysis of regularized M-estimators in high dimensions. *IEEE Trans. Inf. Theory* **64** 5592–5628.

Thrampoulidis, C., Oymak, S. and Hassibi, B. (2015). Regularized linear regression: A precise analysis of the estimation error. In *Conf. Learn. Theory* 1683–1709.

Tsuda, T. and Imaizumi, M. (2024). Benign overfitting of non-sparse high-dimensional linear regression with correlated noise. *Electron. J. Stat.* **18** 4119–4197.

Tsuda, T. and Imaizumi, M. (2026). Supplement to "Universality of estimators for high-dimensional linear models with block dependency." https://doi.org/10.3150/25-BEJ1941SUPP

Wang, T., Zhong, X. and Fan, Z. (2024). Universality of approximate message passing algorithms and tensor networks. *Ann. Appl. Probab.* **34** 3943–3994.

Yadlowsky, S., Yun, T., McLean, C.Y. and D'Amour, A. (2021). SLOE: A faster method for statistical inference in high-dimensional logistic regression. *Adv. Neural Inf. Process. Syst.* **34** 29517–29528.

Zhao, Q., Sur, P. and Candes, E.J. (2022). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli* **28** 1835–1861.

Bernoulli **32**(3), 2026, 2022–2045
https://doi.org/10.3150/25-BEJ1943

Sharp anti-concentration inequalities for extremum statistics via copulas

MATIAS D. CATTANEO[1,a], RICARDO P. MASINI[2,b] and WILLIAM G. UNDERWOOD[3,c]

[1] *Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA,* [a] *cattaneo@princeton.edu*
[2] *Department of Statistics, University of California Davis, Davis, CA 95616, USA,* [b] *rmasini@ucdavis.edu*
[3] *Statistical Laboratory, University of Cambridge, Cambridge, UK,* [c] *wgu21@cam.ac.uk*

We derive sharp upper and lower bounds for the pointwise concentration function of the maximum statistic of d identically distributed real-valued random variables. Our first main result places no restrictions either on the common marginal law of the samples or on the copula describing their joint distribution. We show that, in general, strictly sublinear dependence of the concentration function on the dimension d is not possible. We then introduce a new class of copulas, namely those with a convex diagonal section, and demonstrate that restricting to this class yields a sharper upper bound on the concentration function. This allows us to establish several new dimension-independent and poly-logarithmic-in-d anti-concentration inequalities for a variety of marginal distributions under mild dependence assumptions. Our theory improves upon the best known results in certain special cases. Applications to high-dimensional statistical inference are presented, including a specific example pertaining to Gaussian mixture approximations for factor models, for which our main results lead to superior distributional guarantees.

Keywords: Anti-concentration; concentration; copulas; extreme value theory; high-dimensional probability; order statistics

References

Aizenman, M., Germinet, F., Klein, A. and Warzel, S. (2009). On Bernoulli decompositions for random variables, concentration bounds, and spectral localization. *Probab. Theory Related Fields* **143** 219–238.

Arakelian, V. and Karlis, D. (2014). Clustering dependencies via mixtures of copulas. *Comm. Statist. Simulation Comput.* **43** 1644–1661.

Bakshi, A., Diakonikolas, I., Hopkins, S.B., Kane, D., Karmalkar, S. and Kothari, P.K. (2020). Outlier-robust clustering of Gaussians and other non-spherical mixtures. In *61st Annual Symposium on Foundations of Computer Science* 149–159. IEEE.

Belloni, A., Fang, E.X. and Shen, S. (2024). Anti-concentration inequalities for the difference of maxima of Gaussian random vectors. Available at arXiv:2408.13348.

Birnbaum, Z.W. (1942). An inequality for Mills' ratio. *Ann. Math. Statist.* **13** 245–246.

Bobkov, S.G. and Chistyakov, G.P. (2015). On concentration functions of random variables. *J. Theoret. Probab.* **28** 976–988.

Cattaneo, M.D., Feng, Y. and Underwood, W.G. (2024). Uniform inference for kernel density estimators with dyadic data. *J. Amer. Statist. Assoc.* **119** 2695–2708.

Cattaneo, M.D., Masini, R.P. and Underwood, W.G. (2025). Yurinskii's coupling for martingales. *Ann. Statist.* **53** 2179–2203.

Cattaneo, M.D. and Yu, R.R. (2025). Strong approximations for empirical processes indexed by Lipschitz functions. *Ann. Statist.* **81** 667–737.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818.
Chernozhukov, V., Chetverikov, D. and Kato, K. (2014b). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597.
Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70.
Chernozhukov, V., Chetverikov, D. and Kato, K. (2017a). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352.
Chernozhukov, V., Chetverikov, D. and Kato, K. (2017b). Detailed proof of Nazarov's inequality. Available at arXiv:1711.10696.
Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. *Probability and Mathematical Statistics*. New York: Academic Press.
Cuculescu, I. and Theodorescu, R. (2001). Copulas: Diagonals, tracks. *Rev. Roumaine Math. Pures Appl.* **46** 731–742.
Deng, H. and Zhang, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671.
Döbler, C. and Peccati, G. (2018). The Gamma Stein equation and noncentral de Jong theorems. *Bernoulli* **24** 3384–3421.
Durante, F. and Sempi, C. (2016). *Principles of Copula Theory*. New York: Chapman and Hall/CRC Press.
Embrechts, P., Klüppelberg, C. and Mikosch, T. (2013). *Modelling Extremal Events: For Insurance and Finance*. *Stochastic Modelling and Applied Probability* **33**. Berlin Heidelberg: Springer.
Fernández-Sánchez, J. and Úbeda-Flores, M. (2018). Constructions of copulas with given diagonal (and opposite diagonal) sections and some generalizations. *Depend. Model.* **6** 139–155.
Fox, J., Kwan, M. and Sauermann, L. (2021). Combinatorial anti-concentration inequalities, with applications. In *Math. Proc. Cambridge Philos. Soc* **171** 227–248.
Frank, M.J., Nelsen, R.B. and Schweizer, B. (1987). Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probab. Theory Related Fields* **74** 199–211.
Gaunt, R.E., Pickett, A.M. and Reinert, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *Ann. Appl. Probab.* **27** 720–756.
Giessing, A. (2023). Anti-concentration of suprema of Gaussian processes and Gaussian order statistics. Available at arXiv:2310.12119.
Götze, F., Naumov, A., Spokoiny, V.G. and Ulyanov, V.V. (2019). Large ball probabilities, Gaussian comparison and anti-concentration. *Bernoulli* **25** 2538–2563.
Jaworski, P. (2009). On copulas and their diagonals. *Inf. Sci.* **179** 2863–2871.
Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297.
Kozbur, D. (2021). Dimension-free anticoncentration bounds for Gaussian order statistics with discussion of applications to multiple testing. Available at arXiv:2107.10766.
Krishnapur, M. (2016). *Anti-Concentration Inequalities*. *Lecture Notes, Advanced Training in Mathematics Workshop in Applied Probability*. Bombay: Indian Institute of Technology.
Kuchibhotla, A.K., Mukherjee, S. and Banerjee, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192–217.
Lévy, P. (1954). *Théorie de l'Addition Des Variables Aléatoires*, 2nd ed. *Monographies des Probabilités*. Paris: Gauthier-Villars.
Lindvall, T. (1992). *Lectures on the Coupling Method*. New York: Wiley.
Litvak, A.E., Lytova, A., Tikhomirov, K., Tomczak-Jaegermann, N. and Youssef, P. (2017). Adjacency matrices of random digraphs: Singularity and anti-concentration. *J. Math. Anal. Appl.* **445** 1447–1491.
Livshyts, G.V. (2014). Maximal surface area of a convex set in $\mathbb{R}^n$ with respect to log concave rotation invariant measures. In *Geometric Aspects of Functional Analysis: Israel Seminar, 2011–2013* 355–383. Springer.
Livshyts, G.V. (2021). Some remarks about the maximal perimeter of convex sets with respect to probability measures. *Commun. Contemp. Math.* **23**.
Lopes, M.E., Lin, Z. and Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Ann. Statist.* **48** 1214–1229.

Meka, R., Nguyen, O. and Vu, V.H. (2015). Anti-concentration for polynomials of independent random variables. *Theory Comput.* **12** 1–17.

Miller, K.S. and Samko, S.G. (2001). Completely monotonic functions. *Integral Transforms Spec. Funct.* **12** 389–402.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis: Israel Seminar, 2001–2002* 169–187. Springer.

Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. New York: Springer.

Nie, Z. (2021). Matrix anti-concentration inequalities with applications. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* 568–581.

Paouris, G. (2012). Small ball probability estimates for log-concave measures. *Trans. Amer. Math. Soc.* **364** 287–308.

Paouris, G. and Valettas, P. (2018). A Gaussian small deviation inequality for convex functions. *Ann. Probab.* **46** 1441–1454.

Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. *Cambridge Series in Statistical and Probabilistic Mathematics*. New York: Cambridge University Press.

Rudelson, M. and Vershynin, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Not.* **2015** 9594–9617.

Saumard, A. and Wellner, J.A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114.

Vershynin, R. and Rudelson, M. (2007). Anti-concentration inequalities. In *Phenomena in High Dimensions, Third Annual Conference*. Samos, Greece.

Bernoulli **32**(3), 2026, 2046–2070
https://doi.org/10.3150/25-BEJ1944

On some geometric identities involving the sample covariance matrix and its adjugate

ALEXANDER DÜRRE[1,a] and DAVY PAINDAVEINE[2,b]

[1]*Mathematisch Instituut, Universiteit Leiden, Leiden, The Netherlands,* [a]*a.m.durre@math.leidenuniv.nl*
[2]*ECARES and Department of Mathematics, Université libre de Bruxelles, Brussels, Belgium,*
[b]*Davy.Paindaveine@ulb.be*

Several identities, involving the Lebesgue measure of data-based simplices or parallelotopes, have been obtained for functionals involving the sample covariance matrix S or its population counterpart Σ. This is the case in particular for Wilks' generalized variance $\det(S)$, which allowed one to obtain an explicit expression for $\mathrm{E}[\det(S)]$ whenever observations are randomly sampled from a distribution with finite second-order moments. To date, however, all such results are limited to scalar functionals. In this paper, we obtain geometric identities for the adjugate $\mathrm{adj}(S)$ of S and for other functionals involving $\mathrm{adj}(S)$ and the sample mean vector $\bar{X}$. This allows us in particular to define uniformly minimum risk unbiased (UMRU) estimators of the corresponding population quantities. Just as the results from (*Ann. Statist.* **36** (2008) 2261–2283) find applications when conditional independence is of interest, our results are relevant in an elliptical framework (resp., in a general framework) where conditional independence is replaced with partial uncorrelatedness (resp., with an original concept of partial median-uncorrelatedness).

Keywords: Adjugate matrices; graphical models; partial uncorrelatedness; random polytopes; random projections; sample covariance matrices; UMRU estimators

References

Bodnar, T., Mazur, S. and Podgórski, K. (2016). Singular inverse Wishart distribution and its application to portfolio theory. *J. Multivariate Anal.* **143** 314–326. https://doi.org/10.1016/j.jmva.2015.09.021

Bodnar, T. and Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *J. Multivariate Anal.* **99** 2389–2045. MR2463397 https://doi.org/10.1016/j.jmva.2008.02.024

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11** 368–385. https://doi.org/10.1016/0047-259X(81)90082-8

Drton, M., Massam, H. and Olkin, I. (2008). Moments of minors of Wishart matrices. *Ann. Statist.* **36** 2261–2283. MR2458187 https://doi.org/10.1214/07-AOS522

Dürre, A. and Paindaveine, D. (2022a). Affine-equivariant inference for multivariate location under L_p loss functions. *Ann. Statist.* **50** 2616–2640. https://doi.org/10.1214/22-AOS2199

Dürre, A. and Paindaveine, D. (2022b). Supplement to "Affine-equivariant inference for multivariate location under L_p loss functions". *Ann. Statist. Online supplement.* https://doi.org/10.1214/22-AOS2199SUPP

Dürre, A. and Paindaveine, D. (2026). Supplement to "On some geometric identities involving the sample covariance matrix and its adjugate." https://doi.org/10.3150/25-BEJ1944SUPP

Horn, R.A. and Johnson, C.R. (2013). *Matrix Analysis*, 2nd ed. New York: Cambridge Univ. Press.

Janssen, P., Serfling, R. and Veraverbeke, N. (1984). Asymptotic normality for a general class of statistical functions and applications to measures of spread. *Ann. Statist.* **12** 1369–1379. https://doi.org/10.1214/aos/1176346797

Magnus, J.R. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Chichester: Wiley.

Muirhead, R.J. (2005). *Aspects of Multivariate Statistical Theory*. Hoboken, New Jersey: Wiley.

Paindaveine, D. (2022). On the measure of anchored Gaussian simplices, with applications to multivariate medians. *Bernoulli* **28** 965–996. https://doi.org/10.3150/21-BEJ1373

Pfanzagl, J. (1994). *Parametric Statistical Theory*. Berlin: Walter de Gruyter.

1350-7265

Pronzato, L. (1998). On a property of the expected value of a determinant. *Statist. Probab. Lett.* **39** 161–165. https://doi.org/10.1016/S0167-7152(98)00057-1

Pronzato, L., Wynn, H.P. and Zhigljavsky, A.A. (2017). Extended generalised variances, with applications. *Bernoulli* **23** 2617–2642. https://doi.org/10.3150/16-BEJ821

Pronzato, L., Wynn, H.P. and Zhigljavsky, A.A. (2018). Simplicial variances, potentials and Mahalanobis distances. *J. Multivariate Anal.* **168** 276–289. https://doi.org/10.1016/j.jmva.2018.08.002

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Serfling, R.J. (1984). Generalized L-, M-, and R-statistics. *Ann. Statist.* **12** 76–86. MR0733500 https://doi.org/10.1214/aos/1176346393

Srivastava, M.S. (2003). Singular Wishart and multivariate beta distributions. *Ann. Statist.* **31** 1537–1560. https://doi.org/10.1214/aos/1065705118

Uhler, C. (2018). Gaussian graphical models. In *Handbook of Graphical Models* 217–238. CRC Press.

van der Vaart, H.R. (1965). A note on Wilks' internal scatter. *Ann. Math. Statist.* **36** 1308–1312. https://doi.org/10.1214/aoms/1177700006

Vogel, D. and Fried, R. (2011). Elliptical graphical modelling. *Biometrika* **98** 935–951. https://doi.org/10.1093/biomet/asr037

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.

Wilks, S.S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24** 471–494. https://doi.org/10.2307/2331979

Bernoulli **32**(3), 2026, 2071–2097
https://doi.org/10.3150/25-BEJ1945

Multivariate Hilbertian additive regression with general estimated variables

JEONG MIN JEON[1,a] and GERMAIN VAN BEVER[2,b]

[1]*Seoul National University, Seoul, South Korea,* [a]*jeongmin.jeon.stat@gmail.com*
[2]*Université libre de Bruxelles and Université de Namur, Brussels, Belgium,* [b]*germain.van.bever@ulb.be*

We investigate a multivariate Hilbertian additive model in which the response variable is Hilbert-space-valued and predictors are multi-dimensional Euclidean. We allow for the scenario where both variables are unobservable but they are estimable. This scenario includes the case of principal or singular component scores, the case of density-valued responses and the case of semiparametric regression. For such cases, we provide estimation errors for the variables, which are of importance in their own right. Additionally, we derive the full non-asymptotic and asymptotic properties of our regression estimator under such estimation errors. This allows us to handle various novel regression problems. We demonstrate the strong performance of our regression estimator via simulation studies and a real data application.

Keywords: Additive model; dimension reduction; non-Euclidean data; smooth backfitting

References

[1] Abdulaziz, A. (2023). A review of compositional data analysis and recent advances. *Comm. Statist. Theory Methods* **52** 5535–5567.
[2] Ali, M.H., Wainwright, B., Petersen, A., Jonnadula, G.B., Desai, M., Rao, H.L., Srinivas, M.B., Jammalamadaka, S.R., Senthil, S. and Pyne, S. (2021). Circular functional analysis of OCT data for precise identification of structural phenotypes in the eye. *Sci. Rep.* **11** 23336.
[3] Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555.
[4] Bunkure, J.K. (2019). Lebegue-Bochner spaces and evolution triples. *Int. J. Math. Appl.* **7** 41–52.
[5] Cai, T.T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179.
[6] Cheng, M.-Y. and Wu, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434.
[7] Cohn, D.L. (2013). *Measure Theory*. Basel: Birkhäuser.
[8] Dai, X. and Müller, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.* **46** 3334–3361.
[9] Hall, P. and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91.
[10] Han, K. and Park, B.U. (2018). Smooth backfitting for error-in-variables additive models. *Ann. Statist.* **46** 2216–2250.
[11] Han, K., Müller, H.-G. and Park, B.U. (2018). Smooth backfitting for additive modeling with small errors-in-variables, with an application to additive functional regression for multiple predictor functions. *Bernoulli* **24** 1233–1265.
[12] Han, K., Müller, H.-G. and Park, B.U. (2020). Additive functional regression for densities as responses. *J. Amer. Statist. Assoc.* **115** 997–1010.
[13] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.
[14] Jeon, J.M. and Park, B.U. (2020). Additive regression with Hilbertian responses. *Ann. Statist.* **48** 2671–2697.
[15] Jeon, J.M. and Van Bever, G. (2026). Supplement to "Multivariate Hilbertian additive regression with general estimated variables." https://doi.org/10.3150/25-BEJ1945SUPP

[16] Jeon, J.M., Park, B.U. and Van Keilegom, I. (2021). Additive regression for non-Euclidean responses and predictors. *Ann. Statist.* **49** 2611–2641.
[17] Jeon, J.M., Lee, Y.K., Mammen, E. and Park, B.U. (2022). Locally polynomial Hilbertian additive regression. *Bernoulli* **28** 2034–2066.
[18] Jeon, J.M., Park, B.U. and Van Keilegom, I. (2022). Nonparametric regression on Lie groups with measurement errors. *Ann. Statist.* **50** 2973–3008.
[19] Lee, Y.K., Mammen, E. and Park, B.U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.* **38** 2857–2883.
[20] Lee, Y.K., Mammen, E. and Park, B.U. (2012). Flexible generalized varying coefficient regression models. *Ann. Statist.* **40** 1906–1933.
[21] Lin, Z. and Yao, F. (2019). Intrinsic Riemannian functional data analysis. *Ann. Statist.* **47** 3533–3577.
[22] Linton, O. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–101.
[23] Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Ann. Statist.* **36** 686–718.
[24] Maier, E.-M., Stöcker, A., Fitzenberger, B. and Greven, S. (2025). Additive density-on-scalar regression in Bayes Hilbert spaces with an application to gender economics. *Ann. Appl. Stat.* **19** 680–700.
[25] Mammen, E., Linton, O. and Nielsen, J.P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443–1490.
[26] Nielsen, J.P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 43–61.
[27] Park, B.U., Chen, C.-J., Tao, W. and Müller, H.-G. (2018). Singular additive models for function to function regression. *Statist. Sinica* **28** 2497–2520.
[28] Petersen, A. and Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218.
[29] Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719.
[30] Preston, C. (2008). Some notes on standard Borel and related spaces. Available at arXiv:0809.3066.
[31] Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.
[32] Yang, W., Müller, H.-G. and Stadtmüller, U. (2011). Functional singular component analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 303–324.
[33] Yu, K., Park, B.U. and Mammen, E. (2008). Smooth backfitting in generalized additive models. *Ann. Statist.* **36** 228–260.
[34] Yu, K., Mammen, E. and Park, B.U. (2011). Semi-parametric regression: Efficiency gains from modeling the nonparametric part. *Bernoulli* **17** 736–748.

Bernoulli **32**(3), 2026, 2098–2122
https://doi.org/10.3150/25-BEJ1946

On the passage times of self-similar Gaussian processes on curved boundaries

DAVAR KHOSHNEVISAN[1,a] and CHEUK YIN LEE[2,b]

[1] *Department of Mathematics, The University of Utah, Salt Lake City, Utah 84112-0090, USA,* [a] *davar@math.utah.edu*
[2] *School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Longgang, Shenzhen, Guangdong, 518172, P.R. China,* [b] *leecheukyin@cuhk.edu.cn*

Let $T_{c,\beta}$ denote the smallest $t \geq 1$ that a continuous, self-similar Gaussian process with self-similarity index $\alpha > 0$ moves at least $\pm ct^{\beta}$ units. We prove that: (i) If $\beta > \alpha$, then $T_{c,\beta} = \infty$ with positive probability; (ii) If $\beta < \alpha$ and X is strongly locally nondeterministic in the sense of Pitt (1978), then $T_{c,\beta}$ has moments of all order; and (iii) If $\beta = \alpha$ and X is strongly locally nondeterministic in the sense of Pitt (1978), then there exists a continuous, strictly decreasing function $\lambda : (0, \infty) \to (0, \infty)$ such that $\mathrm{E}(T^{\mu}_{c,\beta})$ is finite when $0 < \mu < \lambda(c)$ and infinite when $\mu > \lambda(c)$. Together these results extend a celebrated theorem of Breiman (1967) and Shepp (1967) for passage times of a Brownian motion on the critical square-root boundary. We briefly discuss two examples: One about fractional Brownian motion, and another about a family of linear stochastic partial differential equations.

Keywords: Boundary crossing probabilities; Gaussian processes; self-similarity; strong local non-determinism

References

[1] Alili, L. and Patie, P. (2014). Boundary crossing identities for Brownian motion and some nonlinear ODE's. *Proc. Amer. Math. Soc.* **142** 3811–3824. MR3251722 https://doi.org/10.1090/S0002-9939-2014-12194-0
[2] Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176. https://doi.org/10.2307/2032333
[3] Barczy, M. and Döring, L. (2013). On entire moments of self-similar Markov processes. *Stoch. Anal. Appl.* **31** 191–198. MR3021485 https://doi.org/10.1080/07362994.2013.741397
[4] Bass, R.F. and Burdzy, K. (1996). A critical case for Brownian slow points. *Probab. Theory Related Fields* **105** 85–108. MR1389733 https://doi.org/10.1007/BF01192072
[5] Berman, S.M. (1978). Gaussian processes with biconvex covariances. *J. Multivariate Anal.* **8** 30–44. MR0517591 https://doi.org/10.1016/0047-259X(78)90017-9
[6] Berman, S.M. (1987). Spectral conditions for local nondeterminism. *Stoch. Process. Appl.* **27** 73–84. MR0934530 https://doi.org/10.1016/0304-4149(87)90006-8
[7] Bertoin, J. (1996). *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge: Cambridge University Press. MR1406564
[8] Bertoin, J. and Kortchemski, I. (2016). Self-similar scaling limits of Markov chains on the positive integers. *Ann. Appl. Probab.* **26** 2556–2595. MR3543905 https://doi.org/10.1214/15-AAP1157
[9] Bertoin, J. and Yor, M. (2002). The entrance laws of self-similar Markov processes and exponential functionals of Lévy processes. *Potential Anal.* **17** 389–400. MR1918243 https://doi.org/10.1023/A:1016377720516
[10] Bertoin, J. and Yor, M. (2002). On the entire moments of self-similar Markov processes and exponential functionals of Lévy processes. *Ann. Fac. Sci. Toulouse Math. (6)* **11** 33–45. MR1986381
[11] Blackwell, D. and Freedman, D. (1964). A remark on the coin tossing game. *Ann. Math. Statist.* **35** 1345–1347. https://doi.org/10.1214/aoms/1177703292
[12] Borell, C. (1974). Convex measures on locally convex spaces. *Ark. Mat.* **12** 239–252. https://doi.org/10.1007/BF02384761

1350-7265

[13] Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216. https://doi.org/10.1007/BF01425510

[14] Breiman, L. (1967). First exit times from a square root boundary. In *Proc. Fifth Berkeley Sympos. Math. Statist. And Probability (Berkeley, Calif., 1965/66). Vol. II: Contributions to Probability Theory, Part 2* 9–16. Berkeley, CA: Univ. California Press. MR0212865

[15] Caballero, M.E. and Chaumont, L. (2006). Weak convergence of positive self-similar Markov processes and overshoots of Lévy processes. *Ann. Probab.* **34** 1012–1034. MR2243877 https://doi.org/10.1214/009117905000000611

[16] Chaumont, L., Kyprianou, A.E. and Pardo, J.C. (2009). Some explicit identities associated with positive self-similar Markov processes. *Stoch. Process. Appl.* **119** 980–1000. MR2499867 https://doi.org/10.1016/j.spa.2008.05.001

[17] Chaumont, L. and Pardo, J.C. (2006). The lower envelope of positive self-similar Markov processes. *Electron. J. Probab.* **11** 1321–1341. MR2268546 https://doi.org/10.1214/EJP.v11-382

[18] Chaumont, L., Kyprianou, A., Pardo, J.C. and Rivero, V. (2012). Fluctuation theory and exit systems for positive self-similar Markov processes. *Ann. Probab.* **40** 245–279. MR2917773 https://doi.org/10.1214/10-AOP612

[19] Chow, Y.S. and Teicher, H. (1966). On second moments of stopping rules. *Ann. Math. Statist.* **37** 388–392. https://doi.org/10.1214/aoms/1177699520

[20] Coeurjolly, J.-F. (2008). Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Ann. Statist.* **36** 1404–1434. https://doi.org/10.1214/009053607000000587

[21] Coti Zelati, M., Drivas, T.D. and Gvalani, R.S. (2024). Mixing by statistically self-similar Gaussian random fields. *J. Stat. Phys.* **191** 61. https://doi.org/10.1007/s10955-024-03277-w

[22] Cuzick, J. and DuPreez, J.P. (1982). Joint continuity of Gaussian local times. *Ann. Probab.* **10** 810–817. MR0659550

[23] Dalang, R.C. (1999). Extending the martingale measure stochastic integral with applications to spatially homogeneous s.p.d.e.'s. *Electron. J. Probab.* **4** 29. MR1684157 https://doi.org/10.1214/EJP.v4-43

[24] Dalang, R.C., Lee, C.Y., Mueller, C. and Xiao, Y. (2021). Multiple points of Gaussian random fields. *Electron. J. Probab.* **26** 17. https://doi.org/10.1214/21-EJP589

[25] Davis, B. (1976). On the L^p norms of stochastic integrals and other martingales. *Duke Math. J.* **43** 697–704. MR0418219

[26] Davis, B. (1983). On Brownian slow points. *Z. Wahrsch. Verw. Gebiete* **64** 359–367. https://doi.org/10.1007/BF00532967

[27] Davis, B. and Perkins, E. (1985). Brownian slow points: The critical case. *Ann. Probab.* **13** 779–803. MR0799422

[28] Dębicki, K. and Tabiś, K. (2020). Pickands-Piterbarg constants for self-similar Gaussian processes. *Probab. Math. Statist.* **40** 297–315. MR4206416 https://doi.org/10.37190/0208-4147.40.2.6

[29] Dereich, S., Döring, L. and Kyprianou, A.E. (2017). Real self-similar processes started from the origin. *Ann. Probab.* **45** 1952–2003. MR3650419 https://doi.org/10.1214/16-AOP1105

[30] Dobrushin, R.L. (1979). Gaussian and their subordinated self-similar random generalized fields. *Ann. Probab.* **7** 1–28.

[31] Döring, L. (2015). A jump-type SDE approach to real-valued self-similar Markov processes. *Trans. Amer. Math. Soc.* **367** 7797–7836. MR3391900 https://doi.org/10.1090/S0002-9947-2015-06270-9

[32] Döring, L. and Barczy, M. (2012). A jump type SDE approach to positive self-similar Markov processes. *Electron. J. Probab.* **17** 39. MR2994842 https://doi.org/10.1214/EJP.v17-2402

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

[34] Dudley, R.M. (1969). Random linear functionals. *Trans. Amer. Math. Soc.* **136** 1–24. MR0264726 https://doi.org/10.2307/1994697

[35] Dudley, R.M., Feldman, J. and Le Cam, L. (1971). On seminorms and probabilities, and abstract Wiener spaces. *Ann. of Math. (2)* **93** 390–408. https://doi.org/10.2307/1970780

[36] Fekete, M. (1923). Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.* **17** 228–249. MR1544613 https://doi.org/10.1007/BF01504345

[37] Fernique, X. (1970). Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris, Ser. A* **270** A1698–A1699. MR0266263

[38] Foondun, M., Khoshnevisan, D. and Mahboubi, P. (2015). Analysis of the gradient of the solution to a stochastic heat equation via fractional Brownian motion. *Stoch. Partial Differ. Equ. Anal. Comput.* **3** 133–158. MR3350450 https://doi.org/10.1007/s40072-015-0045-y

[39] Funaki, T. (1997). Singular limit for reaction-diffusion equation with self-similar Gaussian noise. In *New Trends in Stochastic Analysis (Charingworth, 1994)* 132–152. River Edge, NJ: World Sci. Publ. MR1654352

[40] Gnedin, A. and Pitman, J. (2005). Self-similar and Markov composition structures. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **326** 59–84. 280–281. MR2183216 https://doi.org/10.1007/s10958-007-0447-0

[41] Goodman, V. and Kuelbs, J. (1991). Rates of clustering for some Gaussian self-similar processes. *Probab. Theory Related Fields* **88** 47–75. MR1094077 https://doi.org/10.1007/BF01193582

[42] Graversen, S.E. and Vuolle-Apiala, J. (1986). α-Self-similar Markov processes. *Probab. Theory Related Fields* **71** 149–158. https://doi.org/10.1007/BF00366277

[43] Greenwood, P. and Perkins, E. (1983). A conditioned limit theorem for random walk and Brownian local time on square root boundaries. *Ann. Probab.* **11** 227–261. MR0690126

[44] Greenwood, P. and Perkins, E. (1984). Limit theorems for excursions from a moving boundary. *Teor. Veroyatn. Primen.* **29** 703–714. MR0773439

[45] Gross, L. (1967). Abstract Wiener spaces. In *Proc. Fifth Berkeley Sympos. Math. Statist. And Probability (Berkeley, Calif., 1965/66). Vol. II: Contributions to Probability Theory, Part 1* 31–42. Berkeley, CA: Univ. California Press. MR0212152

[46] Gu, Q., Hu, J. and Ngai, S.-M. (2020). Geometry of self-similar measures on intervals with overlaps and applications to sub-Gaussian heat kernel estimates. *Commun. Pure Appl. Anal.* **19** 641–676. MR4043760 https://doi.org/10.3934/cpaa.2020030

[47] Gundy, R.F. and Siegmund, D. (1967). On a stopping rule and the central limit theorem. *Ann. Math. Statist.* **38** 1915–1917. https://doi.org/10.1214/aoms/1177698627

[48] Hairer, M. and Pardoux, E. (2015). A Wong-Zakai theorem for stochastic PDEs. *J. Math. Soc. Japan* **67** 1551–1604. MR3417505 https://doi.org/10.2969/jmsj/06741551

[49] Harnett, D. and Nualart, D. (2017). Decomposition and limit theorems for a class of self-similar Gaussian processes. In *Stochastic Analysis and Related Topics*. *Progr. Probab.* **72** 99–116. Cham: Birkhäuser/Springer. MR3737626 https://doi.org/10.1007/978-3-319-59671-6_5

[50] Harnett, D. and Nualart, D. (2018). Central limit theorem for functionals of a generalized self-similar Gaussian process. *Stoch. Process. Appl.* **128** 404–425. MR3739502 https://doi.org/10.1016/j.spa.2017.04.014

[51] Hong, M., Liu, H. and Xu, F. (2024). Limit theorems for additive functionals of some self-similar Gaussian processes. *Ann. Appl. Probab.* **34** 5462–5497. MR4839629 https://doi.org/10.1214/24-aap2096

[52] Kalbasi, K. and Mountford, T. (2020). On the probability distribution of the local times of diagonally operator-self-similar Gaussian fields with stationary increments. *Bernoulli* **26** 1504–1534. MR4058376 https://doi.org/10.3150/19-BEJ1169

[53] Kent, J.T. and Wood, A.T.A. (1997). Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. Roy. Statist. Soc. Ser. B, Methodol.* **59** 679–699. MR1452033

[54] Khoshnevisan, D., Swanson, J., Xiao, Y. and Zhang, L. (2013). Weak Existence of a Solution to a Differential Equation Driven by a Very Rough fBm. Unpublished manuscript available at https://arxiv.org/abs/1309.3613.

[55] Kim, P., Song, R. and Vondraček, Z. (2023). Positive self-similar Markov processes obtained by resurrection. *Stoch. Process. Appl.* **156** 379–420. MR4520527 https://doi.org/10.1016/j.spa.2022.11.014

[56] Kolmogoroff, A.N. (1940). Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C. R. (Dokl.) Acad. Sci. URSS* **26** 115–118. MR0003441

[57] Kyprianou, A.E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. *Universitext.* Berlin: Springer-Verlag. MR2250061

[58] Kyprianou, A.E., Motala, M. and Rivero, V. (2024). Williams' path decomposition for self-similar Markov processes in $\mathbb{R}^d$. *Electron. J. Probab.* **29** 132. https://doi.org/10.1214/24-ejp1187

[59] Lai, D. (2010). Group sequential tests under fractional Brownian motion in monitoring clinical trials. *Stat. Methods Appl.* **19** 277–286. MR2651453 https://doi.org/10.1007/s10260-010-0138-8

[60] Lai, D., Moyé, L.A., Davis, B.R., Brown, L.E. and Sacks, F.M. (2001). Brownian motion and long-term clinical trial recruitment. *J. Statist. Plann. Inference* **93** 239–246. MR1822399 https://doi.org/10.1016/S0378-3758(00)00203-2

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

[62] Lamperti, J. (1967). Continuous state branching processes. *Bull. Amer. Math. Soc.* **73** 382–386. MR0208685 https://doi.org/10.1090/S0002-9904-1967-11762-2

[63] Lamperti, J. (1972). Semi-stable Markov processes. I. *Z. Wahrsch. Verw. Gebiete* **22** 205–225. https://doi.org/10.1007/BF00536091

[64] Latała, R. and Matlak, D. (2017). Royen's proof of the Gaussian correlation inequality. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **2169** 265–275. Cham: Springer. MR3645127

[65] Ledoux, M. (1996). Isoperimetry and Gaussian analysis. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1994)*. *Lecture Notes in Math.* **1648** 165–294. Berlin: Springer. MR1600888 https://doi.org/10.1007/BFb0095676

[66] Lee, C.Y. and Xiao, Y. (2023). Chung-type law of the iterated logarithm and exact moduli of continuity for a class of anisotropic Gaussian random fields. *Bernoulli* **29** 523–550. MR4497257 https://doi.org/10.3150/22-bej1467

[67] Lei, P. and Nualart, D. (2009). A decomposition of the bifractional Brownian motion and some applications. *Statist. Probab. Lett.* **79** 619–624. MR2499385 https://doi.org/10.1016/j.spl.2008.10.009

[68] Liao, M. and Wang, L. (2011). Isotropic self-similar Markov processes. *Stoch. Process. Appl.* **121** 2064–2071. MR2819241 https://doi.org/10.1016/j.spa.2011.05.008

[69] Lysy, M., Pillai, N.S., Hill, D.B., Forest, M.G., Mellnik, J.W.R., Vasquez, P.A. and McKinley, S.A. (2016). Model comparison and assessment for single particle tracking in biological fluids. *J. Amer. Statist. Assoc.* **111** 1413–1426. MR3601698 https://doi.org/10.1080/01621459.2016.1158716

[70] Makogin, V. and Mishura, Y. (2015). Example of a Gaussian self-similar field with stationary rectangular increments that is not a fractional Brownian sheet. *Stoch. Anal. Appl.* **33** 413–428. MR3339311 https://doi.org/10.1080/07362994.2014.1002042

[71] Mandelbrot, B.B. (1982). *The Fractal Geometry of Nature*. San Francisco, CA: W. H. Freeman and Co. MR0665254

[72] Mandelbrot, B.B. and Van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. https://doi.org/10.1137/1010093

[73] Mason, J.D. and Xiao, Y. (2001). Sample path properties of operator-self-similar Gaussian random fields. *Teor. Veroyatn. Primen.* **46** 94–116. MR1968707 https://doi.org/10.1137/S0040585X97978749

[74] Matsui, M. and Shieh, N.-R. (2014). The Lamperti transforms of self-similar Gaussian processes and their exponentials. *Stoch. Models* **30** 68–98. MR3175862 https://doi.org/10.1080/15326349.2014.868735

[75] Mattila, P. (1995). *Geometry of Sets and Measures in Euclidean Spaces*. *Cambridge Studies in Advanced Mathematics* **44**. Cambridge: Cambridge University Press. Fractals and rectifiability. MR1333890 https://doi.org/10.1017/CBO9780511623813

[76] Miclo, L., Patie, P. and Sarkar, R. (2022). Discrete self-similar and ergodic Markov chains. *Ann. Probab.* **50** 2085–2132. MR4499275 https://doi.org/10.1214/22-aop1577

[77] Minlos, R.A. (1959). Generalized random processes and their extension in measure. *Tr. Mosk. Mat. Obs.* **8** 497–518. MR0108851

[78] Molchan, G.M. (1997). Burgers equation with self-similar Gaussian initial data: Tail probabilities. *J. Stat. Phys.* **88** 1139–1150. MR1478064 https://doi.org/10.1007/BF02732428

[79] Monrad, D. and Pitt, L.D. (1987). Local nondeterminism and Hausdorff dimension. In *Seminar on Stochastic Processes, 1986 (Charlottesville, Va., 1986)*. *Progr. Probab. Statist.* **13** 163–189. Boston, MA: Birkhäuser Boston. MR0902433

[80] Mueller, C. and Wu, Z. (2009). A connection between the stochastic heat equation and fractional Brownian motion, and a simple proof of a result of Talagrand. *Electron. Commun. Probab.* **14** 55–65. MR2481666 https://doi.org/10.1214/ECP.v14-1403

[81] Muraoka, H. (1992). Self-similar isotropic Gaussian random fields. In *Probability Theory and Mathematical Statistics (Kiev, 1991)* 210–214. River Edge, NJ: World Sci. Publ. MR1212126

[82] Novikov, A.A. (1971). The stopping times of a Wiener process. *Teor. Veroyatn. Primen.* **16** 458–465. MR0315798

[83] Nualart, D. and Xu, F. (2019). Asymptotic behavior for an additive functional of two independent self-similar Gaussian processes. *Stoch. Process. Appl.* **129** 3981–4008. MR3997669 https://doi.org/10.1016/j.spa.2018.11.009

[84] Pardo, J.C. (2009). The upper envelope of positive self-similar Markov processes. *J. Theoret. Probab.* **22** 514–542. MR2501333 https://doi.org/10.1007/s10959-008-0152-z

[85] Pardo, J.C. and Rivero, V. (2013). Self-similar Markov processes. *Bol. Soc. Mat. Mexicana (3)* **19** 201–235. MR3183994

[86] Patie, P. (2012). Law of the absorption time of some positive self-similar Markov processes. *Ann. Probab.* **40** 765–787. MR2952091 https://doi.org/10.1214/10-AOP638

[87] Perkins, E. (1983). On the Hausdorff dimension of the Brownian slow points. *Z. Wahrsch. Verw. Gebiete* **64** 369–399. https://doi.org/10.1007/BF00532968

[88] Pitt, L.D. (1978). Local times for Gaussian vector fields. *Indiana Univ. Math. J.* **27** 309–330. https://doi.org/10.1512/iumj.1978.27.27024

[89] Rivero, V. (2003). A law of iterated logarithm for increasing self-similar Markov processes. *Stoch. Stoch. Rep.* **75** 443–472. MR2029617 https://doi.org/10.1080/10451120310001646014

[90] Rivero, V. (2005). Recurrent extensions of self-similar Markov processes and Cramér's condition. *Bernoulli* **11** 471–509. MR2146891 https://doi.org/10.3150/bj/1120591185

[91] Rivero, V. (2007). Recurrent extensions of self-similar Markov processes and Cramér's condition. II. *Bernoulli* **13** 1053–1070. MR2364226 https://doi.org/10.3150/07-BEJ6082

[92] Royen, T. (2014). A simple proof of the Gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East J. Theor. Stat.* **48** 139–145. MR3289621

[93] Sato, K.-I. (2013). *Lévy Processes and Infinitely Divisible Distributions*, revised ed. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge University Press. Translated from the 1990 Japanese original.

[94] Shepp, L.A. (1967). A first passage problem for the Wiener process. *Ann. Math. Statist.* **38** 1912–1914. https://doi.org/10.1214/aoms/1177698626

[95] Smith, N.R. and Majumdar, S.N. (2022). Condensation transition in large deviations of self-similar Gaussian processes with stochastic resetting. *J. Stat. Mech. Theory Exp.* **5** 053212. https://doi.org/10.1088/1742-5468/ac6f04

[96] Song, X. (2021). Large deviations for functionals of some self-similar Gaussian processes. *Stochastics* **93** 311–336. https://doi.org/10.1080/17442508.2020.1720018

[97] Talagrand, M. (1987). Regularity of Gaussian processes. *Acta Math.* **159** 99–149. https://doi.org/10.1007/BF02392556

[98] Talagrand, M. (1995). Hausdorff measure of trajectories of multiparameter fractional Brownian motion. *Ann. Probab.* **23** 767–775. MR1334170

[99] Teicher, H. and Zhang, C.-H. (1998). Moments of some stopping rules. *J. Lond. Math. Soc. (2)* **57** 503–512. MR1644253 https://doi.org/10.1112/S0024610798005912

[100] Verdier, H., Laurent, F., Cassé, A., Vestergaard, C.L. and Masson, J.-B. (2022). Variational inference of fractional Brownian motion with linear computational complexity. *Phys. Rev. E* **106** 055311. https://doi.org/10.1103/physreve.106.055311

[101] Verzani, J. (1995). Slow points in the support of historical Brownian motion. *Ann. Probab.* **23** 56–70. MR1330760

[102] Vidmar, M. (2022). Exit problems for positive self-similar Markov processes with one-sided jumps. In *Séminaire de Probabilités LI*. *Lecture Notes in Math.* **2301** 91–115. Cham: Springer. MR4461023 https://doi.org/10.1007/978-3-030-96409-2_3

[103] Walsh, J.B. (1986). An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*. *Lecture Notes in Math.* **1180** 265–439. Berlin: Springer. https://doi.org/10.1007/BFb0074920

[104] Xiao, Y. (1996). Hausdorff measure of the sample paths of Gaussian random fields. *Osaka J. Math.* **33** 895–913. MR1435460

[105] Xiao, Y. (1998). Asymptotic results for self-similar Markov processes. In *Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997)* 323–340. Amsterdam: North-Holland. MR1661490 https://doi.org/10.1016/B978-044450083-0/50022-8

[106] Zhang, Q. and Lai, D. (2011). Fractional Brownian motion and long term clinical trial recruitment. *J. Statist. Plann. Inference* **141** 1783–1788. MR2763208 https://doi.org/10.1016/j.jspi.2010.11.028

[107] Zunino, L., Pérez, D.G., Martín, M.T., Plastino, A., Garavaglia, M. and Rosso, O.A. (2007). Characterization of Gaussian self-similar stochastic processes using wavelet-based informational tools. *Phys. Rev. E* **3** 021115. MR2354015 https://doi.org/10.1103/PhysRevE.75.021115

Bernoulli **32**(3), 2026, 2123–2148
https://doi.org/10.3150/25-BEJ1947

Max-sliced Wasserstein concentration and uniform ratio bounds of empirical measures on RKHS

RUIYU HAN[1,a], CYNTHIA RUSH[2,b] and JOHANNES WIESEL[3,c]

[1] *Department of Mathematical Sciences, Carnegie Mellon University, Wean Hall, Hammerschlag Drive, Pittsburgh, US,* [a] *ruiyuh@andrew.cmu.edu*
[2] *Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York City, US,* [b] *cgr2130@columbia.edu*
[3] *Department of Mathematics, University of Copenhagen, Universitetsparken 5, Copenhagen, Denmark,* [c] *wiesel@math.ku.dk*

Optimal transport and the Wasserstein distance $\mathcal{W}_p$ have recently seen a number of applications in the fields of statistics, machine learning, data science, and the physical sciences. These applications are however severely restricted by the curse of dimensionality, meaning that the number of data points needed to estimate these problems accurately increases exponentially in the dimension. To alleviate this problem, a number of variants of $\mathcal{W}_p$ have been introduced. We focus here on one of these variants, namely the max-sliced Wasserstein metric $\overline{\mathcal{W}}_p$. This metric reduces the high-dimensional minimization problem given by $\mathcal{W}_p$ to a maximum of one-dimensional measurements in an effort to overcome the curse of dimensionality. In this note we derive concentration results and upper bounds on the expectation of $\overline{\mathcal{W}}_p$ between the true and empirical measure on unbounded reproducing kernel Hilbert spaces. We show that, under quite generic assumptions, probability measures concentrate uniformly fast in one-dimensional subspaces, at (nearly) parametric rates. Our results imply an improvement of currently known bounds for $\overline{\mathcal{W}}_p$ in the finite-dimensional case.

Keywords: (Max-sliced) Wasserstein distance; (projection robust) optimal transport; ratio limit theorem; reproducing kernel Hilbert space (RKHS)

References

Alexander, K.S. (1985). Rates of growth for weighted empirical processes. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer. Wadsworth Statistics/Probability Series* 475–493. Belmont, CA: Wadsworth. MR0822047

Alexander, K.S. (1987). The central limit theorem for weighted empirical processes indexed by sets. *J. Multivariate Anal.* **22** 313–339. MR0899666 https://doi.org/10.1016/0047-259X(87)90093-5

Alvarez-Melis, D. and Jaakkola, T.S. (2018). Gromov-Wasserstein alignment of word embedding spaces. In *EMNLP'18: Empirical Methods in Natural Language Processing*.

Arjovsky, M., Chintala, S. and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International Conference on Machine Learning* 214–223. PMLR.

Bartl, D. and Mendelson, S. (2025). A uniform Dvoretzky–Kiefer–Wolfowitz inequality. *Probab. Theory Related Fields* **193** 335–374. MR4970437 https://doi.org/10.1007/s00440-025-01408-2

Bartl, D. and Mendelson, S. (2025+). Structure preservation via the Wasserstein distance. *J. Funct. Anal.*

Bartlett, P. and Lugosi, G. (1999). An inequality for uniform deviations of sample averages from their means. *Statist. Probab. Lett.* **44** 55–62. MR1706315 https://doi.org/10.1016/S0167-7152(98)00291-0

Bayraktar, E. and Guoï, G. (2021). Strong equivalence between metrics of Wasserstein type. *Electron. Commun. Probab.* **26**. Paper No. 13, 13. MR4236683 https://doi.org/10.1214/21-ECP383

Ben-Hur, A. and Noble, W.S. (2005). Kernel methods for predicting protein–protein interactions. *Bioinformatics* **21** i38–i46.

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

Boedihardjo, M. (2025+). Sharp bounds for max-sliced Wasserstein distances. *Found. Comput. Math.*

Bogachev, V.I. (2018). *Weak Convergence of Measures 234*. Providence: American Mathematical Society.

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

Bonneel, N., Rabin, J., Peyré, G. and Pfister, H. (2015). Sliced and Radon Wasserstein barycenters of measures. *J. Math. Imaging Vision* **51** 22–45. MR3300482 https://doi.org/10.1007/s10851-014-0506-3

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

Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 https://doi.org/10.1214/16-AOS1450

Chizat, L., Roussillon, P., Léger, F., Vialard, F.-X. and Peyré, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. *Adv. Neural Inf. Process. Syst.* **33** 2257–2269.

Cortes, C., Haffner, P. and Mohri, M. (2003). Lattice kernels for spoken-dialog classification. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing 1* I–628 IEEE.

Cortes, C., Haffner, P. and Mohri, M. (2004). Rational kernels: Theory and algorithms. *J. Mach. Learn. Res.* **5** 1035–1062. MR2248008

Courty, N., Flamary, R., Habrard, A. and Rakotomamonjy, A. (2017). Joint distribution optimal transportation for domain adaptation. *Adv. Neural Inf. Process. Syst.* **30**.

Cuesta-Albertos, J.A., Matrán-Bea, C. and Tuero-Diaz, A. (1996). On lower bounds for the L^2-Wasserstein metric in a Hilbert space. *J. Theor. Probab.* **9** 263–283. MR1385397 https://doi.org/10.1007/BF02214649

Dereich, S., Scheutzow, M. and Schottstedt, R. (2013). Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183–1203. https://doi.org/10.1214/12-AIHP489

Deshpande, I., Zhang, Z. and Schwing, A.G. (2018). Generative modeling using the sliced Wasserstein distance. In *IEEE Conference on Computer Vision and Pattern Recognition* 3483–3491.

Deshpande, I., Hu, Y.-T., Sun, R., Pyrros, A., Siddiqui, N., Koyejo, S., Zhao, Z., Forsyth, D. and Schwing, A.G. (2019). Max-sliced Wasserstein distance and its use for GANs. In *IEEE Conference on Computer Vision and Pattern Recognition* 10648–10656.

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

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

Gelbrich, M. (1990). On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Math. Nachr.* **147** 185–203. MR1127323 https://doi.org/10.1002/mana.19901470121

Giné, E., Koltchinskii, V. and Wellner, J.A. (2003). Ratio limit theorems for empirical processes. In *Stochastic Inequalities and Applications* **56** 249–278. Basel: Birkhäuser. MR2073436

Girosi, F. (1998). An equivalence between sparse approximation and support vector machines. *Neural Comput.* **10** 1455–1480.

Goldfeld, Z. and Greenewald, K. (2020). Gaussian-smoothed optimal transport: Metric structure and statistical efficiency. In *International Conference on Artificial Intelligence and Statistics* 3327–3337. PMLR.

Goldfeld, Z., Greenewald, K., Niles-Weed, J. and Polyanskiy, Y. (2020). Convergence of smoothed empirical measures with applications to entropy estimation. *IEEE Trans. Inf. Theory* **66** 4368–4391. MR4130621 https://doi.org/10.1109/TIT.2020.2975480

Gramfort, A., Peyré, G. and Cuturi, M. (2015). Fast optimal transport averaging of neuroimaging data. In *Information Processing in Medical Imaging* 261–272. Springer.

Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B. and Smola, A. (2006). A kernel method for the two-sample-problem. In *Advances in Neural Information Processing Systems* (B. Schölkopf, J. Platt and T. Hoffman, eds.) **19**. MIT Press.

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. https://doi.org/10.1214/009053606000000272

Han, R. (2023). Sliced Wasserstein distance between probability measures on Hilbert spaces. arXiv preprint. Available at arXiv:2307.05802.

Han, R., Rush, C. and Wiesel, J. (2026). Supplement to "Max-sliced Wasserstein concentration and uniform ratio bounds of empirical measures on RKHS." https://doi.org/10.3150/25-BEJ1947SUPPA, https://doi.org/10.3150/25-BEJ1947SUPPB

Hofmann, T., Schölkopf, B. and Smola, A.J. (2008). Kernel methods in machine learning. *Ann. Statist.* **36** 1171–1220. https://doi.org/10.1214/009053607000000677

Kolouri, S., Park, S.R. and Rohde, G.K. (2016). The Radon cumulative distribution transform and its application to image classification. *IEEE Trans. Image Process.* **25** 920–934. MR3455456 https://doi.org/10.1109/TIP.2015.2509419

Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R. and Rohde, G. (2019). Generalized sliced Wasserstein distances. *Adv. Neural Inf. Process. Syst.* **32**.

Kondor, R. and Lafferty, J.D. (2002). Diffusion kernels on graphs and other discrete input spaces. In *International Conference on Machine Learning*.

Larsson, M., Park, J. and Wiesel, J. (2024). On concentration of the empirical measure for radial transport costs. *Stoch. Process. Appl.* **178**. Paper No. 104466, 20. MR4792529 https://doi.org/10.1016/j.spa.2024.104466

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

Lin, T., Zheng, Z., Chen, E., Cuturi, M. and Jordan, M.I. (2021). On projection robust optimal transport: Sample complexity and model misspecification. In *International Conference on Artificial Intelligence and Statistics* 262–270. PMLR.

Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N. and Watkins, C. (2002). Text classification using string kernels. *J. Mach. Learn. Res.* **2** 419–444.

Manole, T., Balakrishnan, S. and Wasserman, L. (2022). Minimax confidence intervals for the sliced Wasserstein distance. *Electron. J. Stat.* **16** 2252–2345. MR4402565 https://doi.org/10.1214/22-ejs2001

Mercer, J. (1909). XVI. Functions of positive and negative type, and their connection the theory of integral equations. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **209** 415–446.

Nath, J.S. and Jawanpuria, P. (2020). Statistical optimal transport posed as learning kernel embedding. *Adv. Neural Inf. Process. Syst.*

Nietert, S., Goldfeld, Z., Sadhu, R. and Kato, K. (2022). Statistical, robustness, and computational guarantees for sliced Wasserstein distances. *Adv. Neural Inf. Process. Syst.* **35** 28179–28193.

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

Oh, J.H., Pouryahya, M., Iyer, A., Apte, A.P., Deasy, J.O. and Tannenbaum, A. (2020). A novel kernel Wasserstein distance on Gaussian measures: An application of identifying dental artifacts in head and neck computed tomography. *Comput. Biol. Med.* **120** 103731. https://doi.org/10.1016/j.compbiomed.2020.103731

Olea, J.L.M., Rush, C., Velez, A. and Wiesel, J. (2022). On the generalization error of norm penalty linear regression models. arXiv preprint. Available at arXiv:2211.07608.

Paty, F.-P. and Cuturi, M. (2019). Subspace robust Wasserstein distances. In *International Conference on Machine Learning* 5072–5081. PMLR.

Paulsen, V.I. and Raghupathi, M. (2016). *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. *Cambridge Studies in Advanced Mathematics* **152**. Cambridge: Cambridge University Press. MR3526117 https://doi.org/10.1017/CBO9781316219232

Poggio, T. and Girosi, F. (1989). A Theory of Networks for Approximation and Learning Technical Report USA.

Pollard, D. (1981). Limit theorems for empirical processes. *Z. Wahrsch. Verw. Gebiete* **57** 181–195. MR626814 https://doi.org/10.1007/BF00535488

Rabin, J., Peyré, G., Delon, J. and Bernot, M. (2011). Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision* 435–446. Springer.

Rachev, S.T. and Rüschendorf, L. (1998). *Mass Transportation Problems. Vol. I: Theory*. *Probability and Its Applications*. New York: Springer-Verlag. MR1619170

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

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR2514435

Reed, M. and Simon, B. (1980). *Methods of Modern Mathematical Physics. I: Functional Analysis*, 2nd ed. New York: Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers]. MR751959

Rubner, Y., Tomasi, C. and Guibas, L.J. (2000). The Earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* **40** 99–121.

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

Schmidt, E. (1907). Zur Theorie der linearen und nichtlinearen Integralgleichungen. *Math. Ann.* **63** 433–476. MR1511415 https://doi.org/10.1007/BF01449770

Scholkopf, B., Sung, K.-K., Burges, C.J., Girosi, F., Niyogi, P., Poggio, T. and Vapnik, V. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans. Signal Process.* **45** 2758–2765.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Singh, S. and Póczos, B. (2018). Minimax distribution estimation in Wasserstein distance. arXiv preprint. Available at arXiv:1802.08855.

Van Gestel, T., Suykens, J.A., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., De Moor, B. and Vandewalle, J. (2004). Benchmarking least squares support vector machine classifiers. *Mach. Learn.* **54** 5–32.

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

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM.

Wang, J., Gao, R. and Xie, Y. (2022). Two-sample test with kernel projected Wasserstein distance. In *International Conference on Artificial Intelligence and Statistics*.

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

Wellner, J.A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45** 73–88. MR651392 https://doi.org/10.1007/BF00635964

Williams, C.K. (1998). Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In *Learning in Graphical Models* 599–621. Springer.

Yang, Z., Jin, C., Wang, Z., Wang, M. and Jordan, M.I. (2020). On function approximation in reinforcement learning: Optimism in the face of large state spaces. *Adv. Neural Inf. Process. Syst.*

Zhang, Z., Wang, M. and Nehorai, A. (2020). Optimal transport in reproducing kernel Hilbert spaces: Theory and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **42** 1741–1754. https://doi.org/10.1109/TPAMI.2019.2903050

Bernoulli **32**(3), 2026, 2149–2173
https://doi.org/10.3150/25-BEJ1948

Generalizing super/sub mot using weak ℓ^1 transport

ERHAN BAYRAKTAR[1,a] and DOMINYKAS NORGILAS[2,b]

[1]*Department of Mathematics, University of Michigan, Ann Arbor, MI, USA,* [a]*erhan@umich.edu*
[2]*Department of Mathematics, North Carolina State University, Raleigh, NC, USA,* [b]*dnorgil@ncsu.edu*

In this article we revisit the weak optimal transport (WOT) problem, introduced by Gozlan, Roberto, Samson and Tetali (*J. Funct. Anal.* **273** (2017) 3327–3405). We work on the real line, with barycentric cost functions, and, as our first result, give the following characterization of the set of optimal couplings for two probability measures μ and ν: every optimizer couples the left tails of μ and ν using a submartingale, the right tails using a supermartingale, while the central region is coupled using a martingale. We then consider a constrained optimal transport problem, where admissible transport plans are only those that are optimal for the WOT problem with L^1 costs. The constrained problem generalizes the (sub/super-) martingale optimal transport problems, studied by Beiglböck and Juillet (*Ann. Probab.* **44** (2016) 42–106), and Nutz and Stebegg (*Ann. Probab.* **46** (2018) 3351–3398) among others. Finally we introduce a generalized *shadow measure* and establish its connection to the WOT. This extends and generalizes the results obtained in (sub/super-) martingale settings.

Keywords: Convex order; couplings; optimal transport; martingale transport

References

[1] Alfonsi, A., Corbetta, J. and Jourdain, B. (2017). Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems. Available at SSRN 3072356.
[2] Alibert, J.-J., Bouchitté, G. and Champion, T. (2019). A new class of costs for optimal transport planning. *European J. Appl. Math.* **30** 1229–1263.
[3] Backhoff-Veraguas, J., Beiglböck, M. and Pammer, G. (2019). Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differential Equations* **58** 203.
[4] Backhoff-Veraguas, J., Beiglböck, M. and Pammer, G. (2020). Weak monotone rearrangement on the line. *Electron. Commun. Probab.* **25** 1–16.
[5] Backhoff-Veraguas, J. and Pammer, G. (2022a). Stability of martingale optimal transport and weak optimal transport. *Ann. Probab.* **32** 721–752.
[6] Backhoff-Veraguas, J. and Pammer, G. (2022b). Applications of weak transport theory. *Bernoulli* **28** 370–394.
[7] Backhoff-Veraguas, J., Bartl, D., Beiglböck, M. and Eder, M. (2020b). Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.* **24** 601–632.
[8] Backhoff-Veraguas, J., Beiglböck, M., Huesmann, M. and Källblad, S. (2020a). Martingale Benamou–Brenier: A probabilistic perspective. *Ann. Probab.* **48** 2258–2289.
[9] Bayraktar, E., Deng, S. and Norgilas, D. (2023). A potential-based construction of the increasing supermartingale coupling. *Ann. Appl. Probab.* **33** 3803–3834.
[10] Bayraktar, E., Deng, S. and Norgilas, D. (2024). Supermartingale shadow couplings: The decreasing case. *Bernoulli* **30** 143–169.
[11] Beiglböck, M., Hobson, D. and Norgilas, D. (2022). The potential of the shadow measure. *Electron. Commun. Probab.* **27** 1–12.
[12] Beiglböck, M. and Juillet, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106.
[13] Beiglböck, M. and Juillet, N. (2021). Shadow couplings. *Trans. Amer. Math. Soc.* **374** 4973–5002.

1350-7265

[14] Chacon, R.V. (1977). Potential processes. *Trans. Amer. Math. Soc.* **226** 39–58.

[15] Chacon, R.V. and Walsh, J.B. (1976). One-dimensional potential embedding. *Séemin. Probab.* **10** 19–23.

[16] Fathi, M. and Shu, Y. (2018). Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli* **24** 672–698.

[17] Gozlan, N. and Juillet, N. (2020). On a mixture of Brenier and Strassen theorems. *Proc. Lond. Math. Soc.* **120** 434–463.

[18] Gozlan, N., Roberto, C., Samson, P.-M. and Tetali, P. (2017). Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* **273** 3327–3405.

[19] Gozlan, N., Roberto, C., Samson, P.-M., Shu, Y. and Tetali, P. (2018). Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 1667–1693.

[20] Hirsch, F. and Roynette, B. (2012). A new proof of Kellerer's theorem. *ESAIM Probab. Stat.* **16** 48–60.

[21] Hobson, D.G. and Norgilas, D. (2019). The left-curtain martingale coupling in the presence of atoms. *Ann. Appl. Probab.* **29** 1904–1928.

[22] Marton, K. (1996a). A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** 556–571.

[23] Marton, K. (1996b). Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *Ann. Probab.* **24** 857–866.

[24] Nutz, M. and Stebegg, F. (2018). Canonical supermartingale couplings. *Ann. Probab.* **46** 3351–3398.

[25] Shu, Y. (2018). Hamilton-Jacobi equations on graph and applications. *Potential Anal.* **48** 125–157.

[26] Shu, Y. (2020). From Hopf–Lax formula to optimal weak transfer plan. *SIAM J. Math. Anal.* **52** 3052–3072.

[27] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

[28] Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.* **81** 73–205.

[29] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563.

Bernoulli **32**(3), 2026, 2174–2196
https://doi.org/10.3150/25-BEJ1949

Convergence of the pruning processes of stable Galton-Watson trees

GABRIEL BERZUNZA OJEDA[1,a] and ANITA WINTER[2,b]

[1]*Department of Mathematical Sciences, University of Liverpool, Liverpool, United Kingdom,*
[a]*gabriel.berzunza-ojeda@liverpool.ac.uk*
[2]*Fakultät für Mathematik, Universität Duisburg-Essen, Essen, Germany,* [b]*anita.winter@uni-due.de*

Since the work of Aldous and Pitman (*Ann. Inst. Henri Poincaré Probab. Stat.* **34** (1998a) 637–686), several authors have studied the pruning processes of Galton-Watson trees and their continuous analogue Lévy trees. Löhr, Voisin and Winter (*Ann. Inst. Henri Poincaré Probab. Stat.* **51** (2015) 1342–1368) introduced the space of bi-measure $\mathbb{R}$-trees equipped with the so-called *leaf sampling weak vague topology* which allows them to unify the discrete and the continuous picture by considering them as instances of the same Feller-continuous Markov process with different initial conditions. Moreover, the authors show that these so-called pruning processes converge in the Skorokhod space of càdlàg paths with values in the space of bi-measure $\mathbb{R}$-trees, whenever the initial bi-measure $\mathbb{R}$-trees converge. In this paper we provide an application to the above principle by verifying that a sequence of suitably rescaled critical conditioned Galton-Watson trees whose offspring distributions lie in the domain of attraction of a stable law of index $\alpha \in (1, 2]$ converge to the α-stable Lévy-tree in the leaf-sampling weak vague topology.

Keywords: Galton-Watson trees; Gromov-weak topology; pruning procedure; real trees; stable Lévy tree; tree-valued process

References

Abraham, R. and Delmas, J.-F. (2008). Fragmentation associated with Lévy processes using snake. *Probab. Theory Related Fields* **141** 113–154. MR2372967 https://doi.org/10.1007/s00440-007-0081-2

Abraham, R. and Delmas, J.-F. (2012). A continuum-tree-valued Markov process. *Ann. Probab.* **40** 1167–1211. MR2962090 https://doi.org/10.1214/11-AOP644

Abraham, R., Delmas, J.-F. and He, H. (2012). Pruning Galton-Watson trees and tree-valued Markov processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 688–705. MR2976559 https://doi.org/10.1214/11-AIHP423

Abraham, R., Delmas, J.-F. and Hoscheit, P. (2014). Exit times for an increasing Lévy tree-valued process. *Probab. Theory Related Fields* **159** 357–403. MR3201925 https://doi.org/10.1007/s00440-013-0509-9

Abraham, R., Delmas, J.-F. and Voisin, G. (2010). Pruning a Lévy continuum random tree. *Electron. J. Probab.* **15** 1429–1473. MR2727317 https://doi.org/10.1214/EJP.v15-802

Abraham, R. and Serlet, L. (2002). Poisson snake and fragmentation. *Electron. J. Probab.* **7**. MR1943890 https://doi.org/10.1214/EJP.v7-116

Aldous, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226

Aldous, D. and Pitman, J. (1998a). Tree-valued Markov chains derived from Galton-Watson processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **34** 637–686. MR1641670 https://doi.org/10.1016/S0246-0203(98)80003-4

Aldous, D. and Pitman, J. (1998b). The standard additive coalescent. *Ann. Probab.* **26** 1703–1726. MR1675063 https://doi.org/10.1214/aop/1022855879

Athreya, S., Löhr, W. and Winter, A. (2016). The gap between Gromov-vague and Gromov-Hausdorff-vague topology. *Stochastic Process. Appl.* **126** 2527–2553. MR3522292 https://doi.org/10.1016/j.spa.2016.02.009

Billingsley, P. (1999). *Convergence of Probability Measures*, Second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. MR700749 https://doi.org/10.1002/9780470316962

1350-7265

Curien, N. and Kortchemski, I. (2014). Random stable looptrees. *Electron. J. Probab.* **19**. MR3286462 https://doi.org/10.1214/EJP.v19-2732

Dieuleveut, D. (2015). The vertex-cut-tree of Galton-Watson trees converging to a stable tree. *Ann. Appl. Probab.* **25** 2215–2262. MR3349006 https://doi.org/10.1214/14-AAP1047

Dress, A., Moulton, V. and Terhalle, W. (1996). *T*-Theory: An overview. *European J. Combin.* **17** 161–175. MR1379369 https://doi.org/10.1006/eujc.1996.0015

Duquesne, T. (2003). A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.* **31** 996–1027. MR1964956 https://doi.org/10.1214/aop/1048516543

Duquesne, T. and Le Gall, J.-F. (2005). Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields* **131** 553–603. MR2147221 https://doi.org/10.1007/s00440-004-0385-4

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. MR838085 https://doi.org/10.1002/9780470316658

Evans, S.N. and Winter, A. (2006). Subtree prune and regraft: A reversible real tree-valued Markov process. *Ann. Probab.* **34** 918–961. MR2243874 https://doi.org/10.1214/009117906000000034

Greven, A., Pfaffelhuber, P. and Winter, A. (2009). Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields* **145** 285–322. MR2520129

Gromov, M. (2007). *Metric Structures for Riemannian and Non-Riemannian Spaces*, English ed. *Modern Birkhäuser Classics*. Boston, MA: Birkhäuser Boston. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates. MR2307192

He, H. and Winkel, M. (2014). Invariance principles for pruning processes of Galton-Watson trees. ArXiv e-prints.

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

Kortchemski, I. (2013). A simple proof of Duquesne's theorem on contour processes of conditioned Galton-Watson trees. In *Séminaire de Probabilités XLV*. *Lecture Notes in Math.* **2078** 537–558. Cham: Springer. MR3185928 https://doi.org/10.1007/978-3-319-00321-4_20

Kortchemski, I. (2014). Random stable laminations of the disk. *Ann. Probab.* **42** 725–759. MR3178472 https://doi.org/10.1214/12-AOP799

Le Gall, J.-F. (2005). Random trees and applications. *Probab. Surv.* **2** 245–311. MR2203728 https://doi.org/10.1214/154957805100000140

Löhr, W., Voisin, G. and Winter, A. (2015). Convergence of bi-measure $\mathbb{R}$-trees and the pruning process. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1342–1368. MR3414450 https://doi.org/10.1214/14-AIHP628

Miermont, G. (2005). Self-similar fragmentations derived from the stable tree. II. Splitting at nodes. *Probab. Theory Related Fields* **131** 341–375. MR2123249 https://doi.org/10.1007/s00440-004-0373-8

Neveu, J. (1986). Arbres et processus de Galton-Watson. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 199–207. MR0850756

Whitt, W. (1980). Some useful functions for functional limit theorems. *Math. Oper. Res.* **5** 67–85. MR561155 https://doi.org/10.1287/moor.5.1.67

Bernoulli **32**(3), 2026, 2197–2219
https://doi.org/10.3150/25-BEJ1950

A law of large numbers for kinetic interacting diffusions

CARLO BELLINGERI[1,a] and FABIO COPPINI[2,b]

[1] *IRIMAS, Unversité de Haute-Alsace, 18 Rue des Frères Lumières, 68200 Mulhouse, France,*
[a] *carlo.bellingeri@uha.fr*
[2] *Mathematisch Instituut, Universiteit Utrecht, Budapestlaan 6, 3584 CD Utrecht, The Netherlands,*
[b] *f.coppini@uu.nl*

We study the convergence of the empirical distribution associated with a system of interacting kinetic particles subject to independent Brownian forcing in a finite horizon setting, using some recent progress on kinetic non-linear partial differential equations. Under general assumptions that require only weak convergence on the initial datum -without assuming independence or moment conditions- we prove convergence in probability to the corresponding non-linear Fokker-Planck PDE.

Keywords: Anisotropic Sobolev spaces; interacting particle system; kinetic non-linear Fokker-Planck equation

References

[1] Bechtold, F. and Coppini, F. (2021). A law of large numbers for interacting diffusions via a mild formulation. *Electron. J. Probab.* **26** 1–27. https://doi.org/10.1214/21-EJP671
[2] Bertini, L., Giacomin, G. and Poquet, C. (2014). Synchronization and random long time dynamics for mean-field plane rotators. *Probab. Theory Related Fields* **160** 593–653. https://doi.org/10.1007/s00440-013-0536-6
[3] Braun, W. and Hepp, K. (1977). The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Comm. Math. Phys.* **56** 101–113. https://doi.org/10.1007/BF01611497
[4] Cardaliaguet, P., Delarue, F., Lasry, J.-M. and Lions, P.-L. (2019). *The Master Equation and the Convergence Problem in Mean Field Games*. *Ann. Math. Stud.* **201**. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9780691193717
[5] Coghi, M., Deuschel, J.-D., Friz, P.K. and Maurelli, M. (2020). Pathwise McKean-Vlasov theory with additive noise. *Ann. Appl. Probab.* **30** 2355–2392. https://doi.org/10.1214/20-AAP1560
[6] Coppini, F. (2022). A note on Fokker–Planck equations and graphons. *J. Stat. Phys.* **187** 15. https://doi.org/10.1007/s10955-022-02905-7
[7] Delarue, F., Lacker, D. and Ramanan, K. (2019). From the master equation to mean field game limit theory: A central limit theorem. *Electron. J. Probab.* **24** 1–54. https://doi.org/10.1214/19-EJP298
[8] Delattre, S., Giacomin, G. and Luçon, E. (2016). A note on dynamical models on random graphs and Fokker-Planck equations. *J. Stat. Phys.* **165** 785–798. https://doi.org/10.1007/s10955-016-1652-3
[9] Dietert, H. (2016). Contributions to mixing and hypocoercivity in kinetic models PhD thesis, University of Cambridge.
[10] Dobrushin, R.L. (1979). Vlasov equations. *Funct. Anal. Appl.* **13** 115–123. https://doi.org/10.1007/BF01077243
[11] Fernandez, B. and Méléard, S. (1997). A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stoch. Process. Appl.* **71** 33–53. https://doi.org/10.1016/S0304-4149(97)00067-7
[12] Flandoli, F., Leimbach, M. and Olivera, C. (2019). Uniform convergence of proliferating particles to the FKPP equation. *J. Math. Anal. Appl.* **473** 27–52. https://doi.org/10.1016/j.jmaa.2018.12.013
[13] Hao, Z., Wu, M. and Zhang, X. (2020). Schauder estimates for nonlocal kinetic equations and applications. *J. Math. Pures Appl. (9)* **140** 139–184. https://doi.org/10.1016/j.matpur.2020.06.003

[14] Hao, Z., Zhang, X., Zhu, R. and Zhu, X. (2024). Singular kinetic equations and applications. *Ann. Probab.* **52** 576–657. https://doi.org/10.1214/23-AOP1666

[15] Jabin, P.-E. and Wang, Z. (2018). Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Invent. Math.* **214** 523–591. https://doi.org/10.1007/s00222-018-0808-y

[16] Jeans, J.H. (1915). On the theory of star-streaming and the structure of the universe. *Mon. Not. R. Astron. Soc.* **76** 70–84. https://doi.org/10.1093/mnras/76.2.70

[17] Jeans, J.H. (1916). On the theory of star-streaming and the structure of the universe. (Second paper.). *Mon. Not. R. Astron. Soc.* **76** 552–572. https://doi.org/10.1093/mnras/76.7.552

[18] Léonard, C. (1986). Une loi des grands nombres pour des systèmes de diffusions avec interaction et à coefficients non bornés. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 237–262.

[19] McKean, H.P. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** 1907–1911. https://doi.org/10.1073/pnas.56.6.1907

[20] Metivier, M. (1987). Weak convergence of measure valued processes using Sobolev-imbedding techniques. In *Stochastic Partial Differential Equations and Applications, Proc. Conf., Trento/Italy 1985. Lect. Notes Math.*. **1236** 172–183.

[21] Neunzert, H. (1975). *Neuere Qualitative und Numerische Methoden in der Plasmaphysik*. Paderborn: Vorlesungsmanuskript.

[22] Neunzert, H. (1984). An introduction to the nonlinear Boltzmann-Vlasov equation. In *Kinetic Theories and the Boltzmann Equation* (C. Cercignani, ed.). *Lecture Notes in Mathematics* 60–110. Berlin, Heidelberg: Springer. https://doi.org/10.1007/BFb0071878

[23] Oelschläger, K. (1984). A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.* **12** 458–479. https://doi.org/10.1214/aop/1176993301

[24] Oliveira, R.I. and Reis, G.H. (2019). Interacting diffusions on random graphs with diverging average degrees: Hydrodynamics and large deviations. *J. Stat. Phys.* **176** 1057–1087. https://doi.org/10.1007/s10955-019-02332-1

[25] Parise, F. and Ozdaglar, A. (2023). Graphon games: A statistical framework for network games and interventions. *Econometrica* **91** 191–225. https://doi.org/10.3982/ECTA17564

[26] Pascucci, A. and Pesce, A. (2024). Sobolev embeddings for kinetic Fokker-Planck equations. *J. Funct. Anal.* **286** 110344. https://doi.org/10.1016/j.jfa.2024.110344

[27] Stroock, D.W. and Varadhan, S.R.S. (2006). *Multidimensional Diffusion Processes*. *Grundlehren der Mathematischen Wissenschaften* **233**. Berlin; New York: Springer.

[28] Sznitman, A.-S. (1991). Topics in propagation of chaos. In *Ecole d'Eté de Probabilités de Saint-Flour XIX - 1989* (P.-L. Hennequin, ed.) **1464** 165–251. Berlin Heidelberg: Springer.

[29] Tanaka, H. (1984). Limit theorems for certain diffusion processes with interaction. In *North-Holland Mathematical Library* (K. Itô, ed.). *Stochastic Analysis* **32** 469–488. Elsevier. https://doi.org/10.1016/S0924-6509(08)70405-7

[30] Triebel, H. (2006). *Theory of Function Spaces III*. *Monographs in Mathematics* **100**. Basel: Birkhäuser. https://doi.org/10.1007/3-7643-7582-5

[31] Villani, C. (2009). *Hypocoercivity*. *Mem. Am. Math. Soc. 950*. American Mathematical Society (AMS), Providence, RI. https://doi.org/10.1090/S0065-9266-09-00567-5

[32] Vlasov, A.A. (1968). The vibrational properties of an electron gas. *Sov. Phys., Usp.* **10** 721. https://doi.org/10.1070/PU1968v010n06ABEH003709

[33] Zhang, X. and Zhang, X. (2024). Cauchy problem of stochastic kinetic equations. *Ann. Appl. Probab.* **34** 148–202. https://doi.org/10.1214/23-AAP1961

Bernoulli **32**(3), 2026, 2220–2243
https://doi.org/10.3150/25-BEJ1951

Convergence in χ^2 distance to the normal distribution for sums of independent random variables

VYTAS ZACHAROVAS[1,2,a]

[1] *Institute of Computer Science, Vilnius University, Vilnius, Lithuania,* [a]*vytas.zacharovas@mif.vu.lt*
[2] *Institute of Statistical Science, Academia Sinica, Taipei, Taiwan*

We study convergence in the Central Limit Theorem under the χ^2-distance. Suppose n independent random variables $X_1, X_2, \dots, X_n$ have zero mean and equal variance. We prove that if the average of χ^2 distances between these variables and the normal distribution is bounded by a sufficiently small constant, then the χ^2 distance between their standardized sum and the normal distribution is $O(1/n)$.

Keywords: Normal approximation; χ^2-distance; Hermite polynomials; subgaussian; Stein's approach; Parseval's identity

References

Artstein, S., Ball, K.M., Barthe, F. and Naor, A. (2004). On the rate of convergence in the entropic central limit theorem. *Probab. Theory Related Fields* **129** 381–390. MR2128238 https://doi.org/10.1007/s00440-003-0329-4

Barron, A.R. (1986). Entropy and the central limit theorem. *Ann. Probab.* **14** 336–342. MR0815975

Bobkov, S.G., Chistyakov, G.P. and Götze, F. (2019). Rényi divergence and the central limit theorem. *Ann. Probab.* **47** 270–323. https://doi.org/10.1214/18-AOP1261

Bobkov, S.G., Chistyakov, G.P. and Götze, F. (2024). Strictly subgaussian probability distributions. *Electron. J. Probab.* **29** Paper No. 62, 28. MR4736269 https://doi.org/10.1214/24-ejp1120

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

Fomin, S.V. (1982). The central limit theorem: Convergence in the norm $\| u \| = (\int_{-\infty}^{\infty} u^2(x) e^{x^2/2}\, dx)^{1/2}$. *Zap. Nauchn. Sem. Leningr. Otdel. Mat. Inst. Steklov. (LOMI)* **119** 218–229, 242, 245. Problems of the theory of probability distribution, VII. MR666101

Hardy, G.H., Littlewood, J.E. and Pólya, G. (1952). *Inequalities*, 2d ed. Cambridge: the University Press. MR0046395

Johnson, O. and Barron, A. (2004). Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields* **129** 391–409. MR2128239 https://doi.org/10.1007/s00440-004-0344-0

Steele, J.M. (2004). *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. *AMS/MAA Problem Books Series*. Washington, DC: Mathematical Association of America. MR2062704 https://doi.org/10.1017/CBO9780511817106

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

Szegő, G. (1975). *Orthogonal Polynomials*, 4th ed. *American Mathematical Society Colloquium Publications* **XXIII**. Providence, RI: American Mathematical Society. MR0372517

Zacharovas, V. (2023). Poisson approximation in $\chi 2$ distance by the Stein-Chen approach. *Bernoulli* **29** 1600–1614. MR4550237 https://doi.org/10.3150/22-bej1512

Zacharovas, V. and Hwang, H.K. (2010). A Charlier-Parseval approach to Poisson approximation and its applications. *Lith. Math. J.* **50** 88–119. MR2607681 https://doi.org/10.1007/s10986-010-9073-5

Zolotarev, V.M. (1976). Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Teor. Veroyatn. Primen.* **21** 741–758. MR0517338

Bernoulli **32**(3), 2026, 2244–2270
https://doi.org/10.3150/25-BEJ1952

Minimax optimal rates of convergence in monotone shuffled and unlinked regression models under vanishing noise

CECILE DUROT[1,a] and DEBARGHYA MUKHERJEE[2,b]

[1]*MODAL'X, UMR CNRS 9023, Université Paris Nanterre, Nanterre, France,* [a]*cecile.durot@parisnanterre.fr*
[2]*Department of Mathematics and Statistics, Boston University, Boston, USA,* [b]*mdeb@bu.edu*

Shuffled regression and unlinked regression represent intriguing challenges that have garnered considerable attention in many fields, including ecological regression, multi-target tracking problems, image denoising, and others. However, a notable gap exists in the existing literature, particularly in vanishing noise, i.e., how the rate of estimation of the underlying signal scales with the error variance. This paper aims to bridge this gap by delving into the monotone function estimation problem under vanishing noise variance, i.e., we allow the error variance to go to 0 as the number of observations increases. Our investigation reveals that, asymptotically, the shuffled regression problem is comparatively simpler than the unlinked regression; if the error variance is smaller than a threshold, then the minimax risk of the shuffled regression is smaller than that of the unlinked regression. On the other hand, the minimax estimation error is of the same order in the two problems if the noise level is larger than that threshold. Our analysis is quite general in that we do not assume smoothness assumptions on the link function m_0 (which only needs to be left-continuous, but may even have jump discontinuities). Because these problems are related to deconvolution, we also provide bounds for deconvolution in a similar context. Through this exploration, we contribute to understanding the intricate relationships between these statistical problems and shed light on their behaviors under vanishing noise.

Keywords: Deconvolution; minimax rate of estimation; shuffled regression; unlinked regression; vanishing noise

References

Azadkia, M. and Balabdaoui, F. (2024). Linear regression with unmatched data: A deconvolution perspective. *J. Mach. Learn. Res.* **25** 1–55.

Balabdaoui, F., Doss, C.R. and Durot, C. (2021). Unlinked monotone regression. *J. Mach. Learn. Res.* **22** 1–60.

Bobkov, S. and Ledoux, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261**.

Brown, P.J. and Payne, C.D. (1986). Aggregate data, ecological regression, and voting transitions. *J. Amer. Statist. Assoc.* **81** 452–460.

Carpentier, A. and Schlüter, T. (2016). Learning relationships between data obtained independently. In *AISTATS* 658–666. PMLR.

Comte, F. and Lacour, C. (2010). Pointwise deconvolution with unknown error distribution. *C. R. Math. Acad. Sci. Paris* **348** 323–326.

Crewe, I. and Payne, C. (1976). Another game with nature: An ecological regression model of the British two-party vote ratio in 1970. *Br. J. Polit. Sci.* **6** 43–81.

Czekanowski, J. (1909). *Zur Differentialdiagnose der Neandertalgruppe*. Vieweg, Wiesbaden.

Dedecker, J., Fischer, A. and Michel, B. (2015). Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one. *Electron. J. Stat.* **9** 234–265.

Dedecker, J. and Michel, B. (2013). Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension. *J. Multivariate Anal.* **122** 278–291.

1350-7265

DeGroot, M.H., Feder, P.I. and Goel, P.K. (1971). Matchmaking. *Ann. Math. Stat.* **42** 578–593.

Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statist. Sinica* 1025–1045.

Delaigle, A. and Van Keilegom, I. (2021). Deconvolution with unknown error distribution. Chapman and Hall/CRC 245–270.

Durot, C. and Mukherjee, D. (2026). Supplement to "Minimax optimal rates of convergence in monotone shuffled and unlinked regression models under vanishing noise." https://doi.org/10.3150/25-BEJ1952SUPP

Durrant-Whyte, H. and Bailey, T. (2006). Simultaneous localization and mapping: Part I. *IEEE Robot. Autom. Mag.* **13** 99–110.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhya, Ser. A* **53** 97–110.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* 1257–1272.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Canad. J. Statist.* **20** 155–169.

Flammarion, N., Mao, C. and Rigollet, P. (2019). Optimal rates of statistical seriation. *Bernoulli* **25** 623–653.

Forsyth, E. and Katz, L. (1946). A matrix approach to the analysis of sociometric data: Preliminary report. *Sociometry* **9** 340–347.

Freedman, D.A., Klein, S.P., Sacks, J., Smyth, C.A. and Everett, C.G. (1991). Ecological regression and voting rights. *Eval. Rev.* **15** 673–711.

Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Jiang, W., King, G., Schmaltz, A. and Tanner, M.A. (2020). Ecological regression with partial identification. *Polit. Anal.* **28** 65–86.

Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.* **37** 2301–2323.

Kousser, J.M. (1973). Ecological regression and the analysis of past politics. *J. Interdiscip. Hist.* **4** 237–262.

Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer.

Ma, R., Cai, T.T. and Li, H. (2021). Optimal estimation of bacterial growth rates based on a permuted monotone matrix. *Biometrika* **108** 693–708.

Petrie, W.F. (1899). Sequences in prehistoric remains. *J. Anthropol. Inst. G. B. Irel.* 295–301.

Poore, A.B. and Gadaleta, S. (2006). Some assignment problems arising from multiple target tracking. *Math. Comput. Modelling* **43** 1074–1091.

Rigollet, P. and Weed, J. (2019). Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Inf. Inference* **8** 691–717.

Slawski, M. and Sen, B. (2024). Permuted and unlinked monotone regression in $\mathbb{R}^d$: An approach based on mixture modeling and optimal transport. *J. Mach. Learn. Res.* **25** 1–57.

Sokal, R.R. (1963). The principles and practice of numerical taxonomy. *Taxon* **12** 190–199.

Szeliski, R. (2022). *Computer Vision: Algorithms and Applications*. Springer Nature.

Wakefield, J. (2003). Sensitivity analyses for ecological regression. *Biometrics* **59** 9–17.

Bernoulli **32**(3), 2026, 2271–2297
https://doi.org/10.3150/25-BEJ1953

Learning topic hierarchies by tree-directed latent variable models

SUNRIT CHAKRABORTY[a], RAYLEIGH LEI[b] and XUANLONG NGUYEN[c]

Department of Statistics, University of Michigan, Ann Arbor, MI, USA
E-mails:[a]*sunritc@umich.edu,* [b]*rayleigh@umich.edu,* [c]*xuanlong@umich.edu*

We study a parametric family of latent variable models, namely topic models, equipped with a hierarchical structure among the topic variables. These models may be viewed as a finite mixture of the latent Dirichlet allocation (LDA) induced distributions, but the LDA components are constrained by a latent hierarchy, specifically a rooted and directed tree structure, which enables the learning of interpretable and latent topic hierarchies of interest. A mathematical framework is developed in order to establish the identifiability of the latent topic hierarchy under suitable regularity conditions and to derive bounds for posterior contraction rates of the model and its parameters. We demonstrate the usefulness of such models and validate their theoretical properties through careful simulation studies and a real data example using the New York Times articles.

Keywords: Consistency; contraction rate; directed tree; identifiability; inverse bound; Latent Dirichlet Allocation; topic hierarchy; topic model

References

Anandkumar, A., Foster, D.P., Hsu, D., Kakade, S.M. and Liu, Y.-K. (2015). A spectral algorithm for latent Dirichlet allocation. *Algorithmica* **72** 193–214.

Arora, S., Ge, R. and Moitra, A. (2012). Learning topic models–going beyond SVD. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science 1–10*. Los Alamitos, CA: IEEE Computer Society.

Baio, G. and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *J. Appl. Stat.* **37** 253–264.

Bing, X., Bunea, F. and Wegkamp, M. (2020a). A fast algorithm with minimax optimal guarantees for topic models with an unknown number of topics. *Bernoulli* **26** 1765–1796.

Bing, X., Bunea, F. and Wegkamp, M. (2020b). Optimal estimation of sparse topic models. *J. Mach. Learn. Res.* **21** 177.

Blei, D.M., Griffiths, T.L. and Jordan, M.I. (2010). The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies. *J. ACM* **57** 7.

Blei, D. and Lafferty, J. (2005). Correlated topic models. In *Proceedings of the 19th International Conference on Neural Information Processing Systems 18* 147–154. Cambridge, MA: MIT Press.

Blei, D.M. and Lafferty, J.D. (2006). Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning 113–120*. New York, NY: ACM Press.

Blei, D.M., Andrew, Y.A. and Jordan, M.I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3** 993–1022.

cAllenby, G.M. and Rossi, P.E. (2006). Hierarchical Bayes models. *The Handbook of Marketing Research: Uses, misuses, and future advances* 418–440.

Camerlenghi, F., Lijoi, A., Orbanz, P. and Prünster, I. (2019). Distribution theory for hierarchical processes. *Ann. Statist.* **47** 67–92.

Catalano, M., De Blasi, P., Lijoi, A. and Prünster, I. (2022). Posterior asymptotics for boosted hierarchical Dirichlet process mixtures. *J. Mach. Learn. Res.* **23** 80.

Catalano, M., Del Sole, C., Lijoi, A. and Prünster, I. (2024). A unified approach to hierarchical random measures. *Sankhya A* **86** 255–287.

Chakraborty, S., Lei, R. and Nguyen, X. (2026). Supplement to "Learning topic hierarchies by tree-directed latent variable models." https://doi.org/10.3150/25-BEJ1953SUPP

Chen, Y., He, S., Yang, Y. and Liang, F. (2023). Learning topic models: Identifiability and finite-sample analysis. *J. Amer. Statist. Assoc.* **118** 2860–2875.

D'Angelo, L. and Denti, F. (2024). A finite-infinite shared atoms nested model for the Bayesian analysis of large grouped data sets. *Bayesian Anal.* **1** 1–34.

Das, R., Zaheer, M. and Dyer, C. (2015). Gaussian LDA for topic models with word embeddings. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* 795–804. Beijing: Association for Computational Linguistics.

Ghosal, S. and van der Vaart, A.W. (2017). *Fundamentals of Nonparametric Bayesian Inference 44*. Cambridge: Cambridge University Press.

Gu, Y. and Dunson, D.B. (2023). Bayesian pyramids: Identifiable multilayer discrete latent structure models for discrete data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 399–426.

Guha, A., Ho, N. and Nguyen, X. (2021). On posterior contraction of parameters and interpretability in Bayesian mixture modeling. *Bernoulli* **27** 2159–2188.

Harary, F. (1969). *J. Graph Theory*. Reading, MA: Addison-Wesley Publishing Company.

Ho, N. and Nguyen, X. (2015). Identifiability and optimal rates of convergence for parameters of multiple types in finite mixtures.

Holder, M. and Lewis, P.O. (2003). Phylogeny estimation: Traditional and Bayesian approaches. *Nat. Rev. Genet.* **4** 275–284.

Holmes, C., Denison, D.T., Ray, S. and Mallick, B. (2005). Bayesian prediction via partitioning. *J. Comput. Graph. Statist.* **14** 811–830.

Javadi, H. and Montanari, A. (2019). Nonnegative matrix factorization via archetypal analysis. *J. Amer. Statist. Assoc.* **115** 896–907.

Ke, Z.T. and Wang, M. (2024). Using SVD for topic modeling. *J. Amer. Statist. Assoc.* **119** 434–449.

Lijoi, A., Prünster, I. and Rebaudo, G. (2023). Flexible clustering via hidden hierarchical Dirichlet priors. *Scand. J. Stat.* **50** 213–234.

Lijoi, A., Prünster, I. and Rigon, T. (2024). Finite-dimensional discrete random structures and Bayesian clustering. *J. Amer. Statist. Assoc.* **119** 929–941.

McGlothlin, A.E. and Viele, K. (2018). Bayesian hierarchical models. *J. Amer. Med. Assoc.* **320** 2365–2366.

Nguyen, X. (2010). Inference of global clusters from locally distributed data. *Bayesian Anal.* **5** 817–845.

Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400.

Nguyen, X. (2015). Posterior contraction of the population polytope in finite admixture models. *Bernoulli* **21** 618–646.

Nguyen, X. (2016). Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure. *Bernoulli* **22**.

Orbanz, P. and Roy, D.M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.

Papaspiliopoulos, O., Stumpf-Fétizon, T. and Zanella, G. (2023). Scalable Bayesian computation for crossed and nested hierarchical models. *Electron. J. Stat.* **17** 3575–3612.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155** 945–959.

Rouder, J.N. and Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychon. Bull. Rev.* **12** 573–604.

Schwaller, L., Robin, S. and Stumpf, M. (2017). A Closed-Form Approach to Bayesian Inference in Tree-Structured Graphical Models.

Shalit, U., Weinshall, D. and Chechik, G. (2013). Modeling musical influence with topic models. In *Proceedings of the 30th International Conference on International Conference on Machine Learning-Volume* **28** 244–252. JMLR.org.

Tang, J., Meng, Z., Nguyen, X., Mei, Q. and Zhang, M. (2014). Understanding the limiting factors of topic modeling via posterior contraction analysis. In *Proceedings of the 31st International Conference on Machine Learning* **32** 190–198. JMLR.org.

Teh, Y.W., Jordan, M.I., Beal, M.J. and Blei, D.M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581.

Valle, F., Osella, M. and Caselle, M. (2020). A topic modeling analysis of tcga breast and lung cancer transcriptomic data. *Cancers* **12** 3799.

Veenman, M., Stefan, A.M. and Haaf, J.M. (2024). Bayesian hierarchical modeling: An introduction and reassessment. *Behav. Res. Methods* **56** 4600–4631.

Wang, Y. (2019). Convergence rates of latent topic models under relaxed identifiability conditions. *Electron. J. Stat.* **13** 37–66.

Webb, J.A., Stewardson, M.J. and Koster, W.M. (2010). Detecting ecological responses to flow variation using Bayesian hierarchical models. *Freshw. Biol.* **55** 108–126.

Bernoulli **32**(3), 2026, 2298–2322
https://doi.org/10.3150/25-BEJ1954

Liberating dimension and spectral norm: A universal approach to spectral properties of sample covariance matrices

YANQING YIN[a]

School of Statistics and Data Science, Joint Lab for Statistics and Finance, Nanjing Audit University, Nanjing, China, [a]*yinyq799@nenu.edu.cn*

In this paper, our primary objective is to elucidate a guiding principle that governs the spectral properties of the sample covariance matrix. This principle exhibits a harmonious behavior across various limiting frameworks, eliminating the necessity for constraints on the rates of dimension p and sample size n as long as both tend to infinity. We achieve this by employing a well-suited normalization technique on the original sample covariance matrix. Subsequently, we establish a robust central limit theorem for linear spectral statistics within this expansive framework, extending the Bai-Silverstein theorem (Ann. Probab. **32** (2004) 553–605). This accomplishment effectively eliminates the need for a bounded spectral norm on the population covariance matrix and relaxes constraints on the rates of dimension p and sample size n. As a result, our findings significantly broaden the applicability of these results in the realm of high-dimensional statistics. To demonstrate the potency of our established results, we provide an illustrative example involving the test for covariance structure under high dimensionality. This illustrative example extends the findings in the work of Ledoit and Wolf (Ann. Stat. **30** (2002) 1081–1102) and Qiu, Li, and Yao (Ann. Stat. **51** (2023) 1427–1451) by liberating both p and n. Extensive numerical analyses are conducted to thoroughly investigate the robustness of our theoretical findings.

Keywords: Central limit theorem; limiting spectral distribution; linear spectral statistics; M-P law; ultra-high dimension covariance matrix

References

Bai, Z.D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statist. Sinica* **9** 611–677.

Bai, Z.D. and Silverstein, J.W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605.

Bai, Z.D. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, Second ed. Springer Verlag.

Bai, Z.D. and Yin, Y.Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875.

Bai, Z.D. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442.

Bai, Z.D., Jiang, D.D., Yao, J.F. and Zheng, S.R. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. https://doi.org/10.1214/09-AOS694

Bao, Z., Ding, X., Wang, J. and Wang, K. (2022). Statistical inference for principal components of spiked covariance matrices. *Ann. Statist.* **50** 1144–1169. https://doi.org/10.1214/21-AOS2143

Bloemendal, A., Knowles, A., Yau, H.-T. and Yin, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. https://doi.org/10.1007/s00440-015-0616-x

Chen, B. and Pan, G. (2015). CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli* **21** 1089–1133. https://doi.org/10.3150/14-BEJ599

Ding, X. and Wang, Z. (2023). Global and local CLTs for linear spectral statistics of general sample covariance matrices when the dimension is much larger than the sample size with applications. arXiv preprint, arXiv:2308.08646.

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

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

Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. https://doi.org/10.1214/13-AOS1134

Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. https://doi.org/10.1214/aos/1009210544

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102.

Liu, Z., Hu, J., Bai, Z. and Song, H. (2023). A CLT for the LSS of large-dimensional sample covariance matrices with diverging spikes. *Ann. Statist.* **51** 2246–2271.

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

Nagao, H. (1973). Asymptotic expansions of the distributions of Bartlett's test and sphericity test under the local alternatives. *Ann. Inst. Statist. Math.* **25** 407–422.

Qiu, J., Li, Z. and Yao, J. (2023). Asymptotic normality for eigenvalue statistics of a general sample covariance matrix when $p/n \to \infty$ and applications. *Ann. Statist.* **51** 1427–1451. https://doi.org/10.1214/23-AOS2300

Silverstein, J.W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. https://doi.org/10.1006/jmva.1995.1083

Tracy, C.A. and Widom, H. (2009). The distributions of random matrix theory and their applications. In *New Trends in Mathematical Physics* 753–765.

Wachter, K.W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1–18.

Wang, L. and Paul, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.* **126** 25–52. https://doi.org/10.1016/j.jmva.2013.12.015

Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

Yao, J.F., Zheng, S.R. and Bai, Z.D. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press.

Yin, Y. (2022). Spectral statistics of high dimensional sample covariance matrix with unbounded population spectral norm. *Bernoulli* **28** 1729–1756. https://doi.org/10.3150/21-bej1391

Yin, Y. (2026). Supplement to "Liberating dimension and spectral norm: A universal approach to spectral properties of sample covariance matrices." https://doi.org/10.3150/25-BEJ1954SUPP

Yin, Y.-Q., Bai, Z.-D. and Krishnaiah, P.R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521.

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

Bernoulli **32**(3), 2026, 2323–2350
https://doi.org/10.3150/25-BEJ1955

The football model, stochastic ordering and martingale transport

GAOYUE GUO[1,a], NICOLAS JUILLET[2,3,b] and WENPIN TANG[4,c]

[1] *Université Paris-Saclay CentraleSupélec, Laboratoire MICS and CNRS FR-3487, France,* [a] *gaoyue.guo@centralesupelec.fr*
[2] *Université de Haute Alsace, IRIMAS (Institut de Recherche en Informatique, Mathématiques, Automatique et Signal) UR 7499, F-68 100 Mulhouse, France,* [b] *nicolas.juillet@uha.fr*
[3] *Université de Strasbourg, France*
[4] *Department of Industrial Engineering and Operations Research, Columbia University, New York, USA,* [c] *wt2319@columbia.edu*

Tournaments are competitions between a number of teams, the outcome of which determines the relative strength or rank of each team. In many cases, the strength of a team in the tournament is given by a score. Perhaps the most fundamental mathematical result in the theory of random tournaments is Moon's theorem, which provides a necessary and sufficient condition for a feasible score sequence via majorization. To give a probabilistic interpretation of Moon's result, Aldous and Kolesnik introduced the football model, the existence of which gives a short proof of Moon's theorem. However, the proof of Aldous and Kolesnik is "noncanonical", leading to the question of a canonical construction of the football model. The purpose of this paper is to provide explicit constructions of the football model with an additional stochastic ordering constraint, which can be formulated by martingale transport. Two solutions are given: one is by solving an entropy optimization problem via Sinkhorn's algorithm, and the other relies on the idea of shadow couplings. It turns out that both constructions yield the property of strong stochastic transitivity. A nontransitive version of the football model is also considered.

Keywords: Entropy optimization; football model; martingale transport; pairwise comparison; score sequence; shadow coupling; Sinkhorn's algorithm; stochastic ordering; strong stochastic transitivity; tournament

References

Acciaio, B., Marini, A. and Pammer, G. (2025). Calibration of the bass local volatility model. *SIAM J. Financial Math.* **16** 803–833.

Adler, I., Cao, Y., Karp, R., Peköz, E.A. and Ross, S.M. (2017). Random knockout tournaments. *Oper. Res.* **65** 1589–1596.

Aldous, D.J. (2021). A prediction tournament paradox. *Amer. Statist.* **75** 243–248.

Aldous, D.J. and Kolesnik, B. (2022). To stay discovered: On tournament mean score sequences and the Bradley-Terry model. *Stoch. Process. Appl.* **150** 844–852.

Avellaneda, M. (1998). Minimum entropy calibration of asset pricing models. *Int. J. Theor. Appl. Finance* **1** 447472.

Axelrod, R. and Hamilton, W.D. (1981). The evolution of cooperation. *Science* **211** 1390–1396.

Backhoff, J., Beiglböck, M., Huesmann, M. and Källblad, S. (2020). Martingale Benamou–Brenier: A probabilistic perspective. *Ann. Probab.* **48** 2258–2289.

Bassan, M., Donderwinkel, S. and Kolesnik, B. (2024). Tournament score sequences, Erdös-Ginzburg-Ziv numbers, and the Lévy-Khintchine method. arXiv:2407.01441.

Bauschke, H.H. and Lewis, A.S. (2000). Dykstra's algorithm with Bregman projections: A convergence proof. *Optimization* **48** 409–427.

1350-7265 This research was funded, in whole or in part, by Agence Nationale de la Recherche, ANR-23_CE40-0017 and Institut Europlace de Finance, Bourse de l'Institut Europlace de Finance.

Bayraktar, E., Deng, S. and Norgilas, D. (2023). A potential-based construction of the increasing supermartingale coupling. *Ann. Appl. Probab.* **33** 3803–3834.

Beiglböck, M., Huesmann, M. and Stebegg, F. (2016). Root to Kellerer. In *Séminaire de Probabilités XLVIII. Lecture Notes in Math.* **2168** 1–12. Cham: Springer.

Beiglböck, M. and Juillet, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106.

Beiglböck, M. and Juillet, N. (2021). Shadow couplings. *Trans. Amer. Math. Soc.* **374** 4973–5002.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L. and Peyré, G. (2015). Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.* **37** A1111–A1138.

Boyd, S.P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Bradley, R.A. (1976). Science, statistics, and paired comparisons. *Biometrics* **32** 213–232.

Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39** 324–345.

Brualdi, R.A. (1984). The doubly stochastic matrices of a vector majorization. *Linear Algebra Appl.* **61** 141–154. https://doi.org/10.1016/0024-3795(84)90027-2

Brualdi, R.A., Hwang, S.-G. and Pyo, S.-S. (1997). Vector majorization via positive definite matrices. *Linear Algebra Appl.* **257** 105–120. https://doi.org/10.1016/S0024-3795(96)00042-0

Cartier, P., Fell, J.M.G. and Meyer, P.-A. (1964). Comparaison des mesures portées par un ensemble convexe compact. *Bull. Soc. Math. France* **92** 435–445.

Chao, K.-M. and Wong, C.S. (1992). Applications of M-matrices to majorization. *Linear Algebra Appl.* **169** 31–40. https://doi.org/10.1016/0024-3795(92)90168-A

Chen, H., Zhao, H., Lam, H., Yao, D. and Tang, W. (2025). MallowsPO: Fine-tune your LLM with preference dispersions. In *ICLR 13*.

Claesson, A., Dukes, M., Franklín, A.F. and Stefánsson, S. u. O. (2023). Counting tournament score sequences. *Proc. Amer. Math. Soc.* **151** 3691–3704.

Conze, A. and Henry-Labordere, P. (2021). Bass construction with multi-marginals: Lightspeed computation in a new local volatility model. Available at SSRN 3853085.

David, H.A. (1959). Tournaments and paired comparisons. *Biometrika* **46** 139–149.

De March, H. and Henry-Labordere, P. (2019). Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants. arXiv preprint, arXiv:1902.04456.

Donderwinkel, S. and Kolesnik, B. (2024). Tournaments and random walks. arXiv:2403.12940.

Ford, L.R. Jr. (1957). Solution of a ranking problem from binary comparisons. *Amer. Math. Monthly* **64** 28–33.

Graham, R.L. and Spencer, J.H. (1971). A constructive solution to a tournament problem. *Canad. Math. Bull.* **14** 45–48.

Harary, F. and Moser, L. (1966). The theory of round Robin tournaments. *Amer. Math. Monthly* **73** 231–246.

Hardy, G.H., Littlewood, J.E. and Pólya, G. (1952). *Inequalities*, 2d ed. Cambridge: University Press.

Hirsch, F., Profeta, C., Roynette, B. and Yor, M. (2011). *Peacocks and Associated Martingales, with Explicit Constructions*. *Bocconi & Springer Series* **3**. Milan; Bocconi University Press, Milan: Springer.

Isaev, M., Iyer, T. and McKay, B.D. (2020). Asymptotic enumeration of orientations of a graph as a function of the out-degree sequence. *Electron. J. Combin.* **27** 30. research paper p1.26. https://doi.org/10.37236/8929

Joe, H. (1988). Majorization, entropy and paired comparisons. *Ann. Statist.* **16** 915–925.

Juillet, N. (2016a). Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1823–1843.

Juillet, N. (2016b). Peacocks parametrised by a partially ordered set. In *Séminaire de Probabilités XLVIII. Lecture Notes in Math.* **2168** 13–32. Cham: Springer. MR3618125

Kellerer, H.G. (1973). Integraldarstellung von Dilationen. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech. Univ. Prague, Prague, 1971; dedicated to the memory of Antonín Špaček)* 341–374. Prague: Academia [Publishing House of the Czechoslovak Academy of Sciences].

Kendall, M.G. (1955). Further contributions to the theory of paired comparisons. *Biometrics* **11** 43–62.

Kolesnik, B. (2023). The asymptotic number of score sequences. *Combinatorica* **43** 827–844.

Kolesnik, B. and Sanchez, M. (2023). Coxeter tournaments. arXiv preprint, arXiv:2302.14002.

Landau, H.G. (1953). On dominance relations and the structure of animal societies. III. The condition for a score structure. *Bull. Math. Biophys.* **15** 143–148.

Laslier, J.-F. (1997). *Tournament Solutions and Majority Voting 7*. Springer.

Lazear, E.P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *J. Polit. Econ.* **89** 841–864.

Lowther, G. (2009). Limits of one-dimensional diffusions. *Ann. Probab.* **37** 78–106.

Luce, R.D. (1959). *Individual Choice Behavior a Theoretical Analysis*. John Wiley & Sons Inc.

MacMahon, P.A. (1920). An American tournament treated by the calculus of symmetric functions. *Quart. J. Math.* **49** 1–36.

Mallows, C.L. (1957). Non-null ranking models. I. *Biometrika* **44** 114–130.

Marshall, A.W., Olkin, I. and Arnold, B.C. (2011). *Inequalities: Theory of Majorization and Its Applications*, second ed. *Springer Series in Statistics*. New York: Springer.

McCullagh, P. and Nelder, J.A. (1983). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

Moon, J.W. (1963). An extension of Landau's theorem on tournaments. *Pacific J. Math.* **13** 1343–1345.

Moon, J.W. (1968). *Topics on Tournaments*. New York-Montreal, Que.-London: Holt, Rinehart and Winston.

Moon, J.W. and Pullman, N.J. (1970). On generalized tournament matrices. *SIAM Rev.* **12** 384–399.

Mosteller, F. (1951). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika* **16** 207–218.

Müller, A. and Rüschendorf, L. (2001). On the optimal stopping values induced by general dependence structures. *J. Appl. Probab.* **38** 672–684.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. *Wiley Series in Probability and Statistics*. Chichester: John Wiley & Sons, Ltd.

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

Pinto, P. and Pischke, N. (2024). On Dykstra's algorithm with Bregman projections. Available at: https://www2.mathematik.tu-darmstadt.de/~pinto/articles/B-Dykstra.pdf.

Plackett, R. (1975). The analysis of permutations. *Appl. Stat.* 193–202.

Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C.D. and Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model. In *Neurips* **36**.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. New York: Springer.

Sinkhorn, R. and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.* **21** 343–348.

Steinhaus, H. and Trybuła, S. (1959). On a paradox in applied probabilities. *Bull. Acad. Pol. Sci.* **7** 108.

Stob, M. (1984). A supplement to: "A mathematician's guide to popular sports" [Amer. Math. Monthly **90** (1983), no. 4, 246–266; MR0700265 (85i:05122a)]. *Amer. Math. Monthly* **91** 277–282.

Stockmeyer, P.K. (2023). Counting various classes of tournament score sequences. *J. Integer Seq.* **26**. Art. 23.5.2, 9.

Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

Suck, R. (2002). Independent random utility representations. *Math. Social Sci.* **43** 371–389. https://doi.org/10.1016/S0165-4896(02)00020-3

Takács, L. (1991). A Bernoulli excursion and its various applications. *Adv. Appl. Probab.* **23** 557–585.

Thurstone, L. (1927). A law of comparative judgment. *Psychol. Rev.* **34** 273.

Thurstone, L. (1931). Rank order as a psycho-physical method. *J. Exp. Psychol.* **14** 187.

Trybuła, S. (1960/61). On the paradox of three random variables. *Zastos. Mat.* **5** 321–332.

Vuksanovic, P. and Hildebrand, A.J. (2021). On cyclic and nontransitive probabilities. *Involve* **14** 327–348. https://doi.org/10.2140/involve.2021.14.327

Zermelo, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Math. Z.* **29** 436–460.

Bernoulli **32**(3), 2026, 2351–2378
https://doi.org/10.3150/25-BEJ1956

Hessian stability and convergence rates for entropic and Sinkhorn potentials via semiconcavity

GIACOMO GRECO[1,a] and LUCA TAMANINI[2,b]

[1]*RoMaDS - Department of Mathematics, Università degli Studi di Roma Tor Vergata, Rome, Italy,* [a]*greco@mat.uniroma2.it*

[2]*Department of Mathematics and Physics "Niccolò Tartaglia", Università Cattolica del Sacro Cuore, Brescia, Italy,* [b]*luca.tamanini@unicatt.it*

In this paper we determine quantitative stability bounds for the Hessian of entropic potentials, *i.e.*, the dual solution to the entropic optimal transport problem. To the authors' knowledge this is the first work addressing this second-order quantitative stability estimate in general unbounded settings. Our proof strategy relies on semiconcavity properties of entropic potentials and on the representation of entropic transport plans as laws of forward and backward diffusion processes, known as Schrödinger bridges. Moreover, our approach allows to deduce a stochastic proof of quantitative stability estimates for entropic transport plans and for gradients of entropic potentials as well. Finally, as a direct consequence of these stability bounds, we deduce exponential convergence rates for gradient and Hessian of Sinkhorn iterates along Sinkhorn's algorithm, a problem that was still open in unbounded settings. Our rates have a polynomial dependence on the regularization parameter.

Keywords: Entropic Optimal Transport; Hamilton-Jacobi-Bellman; Hessian stability; Schrödinger bridges; Sinkhorn's algorithm

References

Akyildiz, O.D., Del Moral, P. and Miguez, J. (2026+a). Gaussian entropic optimal transport: Schrödinger bridges and the Sinkhorn algorithm. *Found. Data Sci.* Forthcoming. https://doi.org/10.3934/fods.2026004

Akyildiz, O.D., Del Moral, P. and Miguez, J. (2026+b). On the contraction properties of Sinkhorn semigroups. *Found. Data Sci.* Forthcoming. https://doi.org/10.3934/fods.2026006

Aubin-Frankowski, P.-C., Korba, A. and Léger, F. (2022). Mirror descent with relative smoothness in measure spaces, with application to Sinkhorn and EM. In *Advances in Neural Information Processing Systems* (A.H. Oh, A. Agarwal, D. Belgrave and K. Cho, eds.).

Bakry, D., Gentil, I. and Ledoux, M. (2013). *Analysis and Geometry of Markov Diffusion Operators 348*. Springer Science & Business Media.

Bauerschmidt, R., Bodineau, T. and Dagallier, B. (2024). Stochastic dynamics and the Polchinski equation: An introduction. *Probab. Surv.* **21** 200–290.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L. and Peyré, G. (2015). Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.* **37** A1111–A1138.

Berman, R.J. (2020). The Sinkhorn algorithm, parabolic optimal transport and geometric Monge–Ampère equations. *Numer. Math.* **145** 771–836.

Bernton, E., Ghosal, P. and Nutz, M. (2022). Entropic optimal transport: Geometry and large deviations. *Duke Math. J.* **171** 3363–3400. https://doi.org/10.1215/00127094-2022-0035

Bortoli, V.D., Thornton, J., Heng, J. and Doucet, A. (2021). Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems* (A. Beygelzimer, Y. Dauphin, P. Liang and J.W. Vaughan, eds.).

Borwein, J.M., Lewis, A.S. and Nussbaum, R. (1994). Entropy minimization, DAD problems, and doubly stochastic kernels. *J. Funct. Anal.* **123** 264–307.

Carlier, G. (2022). On the linear convergence of the multimarginal Sinkhorn algorithm. *SIAM J. Optim.* **32** 786–794.

Carlier, G., Chizat, L. and Laborde, M. (2024). Displacement smoothness of entropic optimal transport. *ESAIM Control Optim. Calc. Var.* **30** 25.

Carlier, G. and Laborde, M. (2020). A differential approach to the multi-marginal Schrödinger system. *SIAM J. Math. Anal.* **52** 709–717.

Chaintron, L.-P. and Conforti, G. (2026+). Regularity and stability for the Gibbs conditioning principle on path space via McKean-Vlasov control. *Probab. Theory Related Fields*. Forthcoming. https://doi.org/10.1007/s00440-025-01447-9

Chen, Y. and Eldan, R. (2022). Localization schemes: A framework for proving mixing bounds for Markov chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* 110–122. IEEE.

Chen, Y., Georgiou, T. and Pavon, M. (2016). Entropic and displacement interpolation: A computational approach using the Hilbert metric. *SIAM J. Appl. Math.* **76** 2375–2396.

Chewi, S. and Pooladian, A.-A. (2023). An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *C. R. Math.* **361** 1471–1482. https://doi.org/10.5802/crmath.486

Chiarini, A., Conforti, G., Greco, G. and Tamanini, L. (2023). Gradient estimates for the Schrödinger potentials: Convergence to the Brenier map and quantitative stability. *Comm. Partial Differential Equations* **48** 895–943. https://doi.org/10.1080/03605302.2023.2215527

Chiarini, A., Conforti, G., Greco, G. and Tamanini, L. (2024). A semiconcavity approach to stability of entropic plans and exponential convergence of Sinkhorn's algorithm. arXiv preprint. Available at arXiv:2412.09235.

Chizat, L., Delalande, A. and Vaškevičius, T. (2025). Sharper exponential convergence rates for Sinkhorn's algorithm in continuous settings. *Math. Program.* 1–50.

Conforti, G. (2024). Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges. *Probab. Theory Related Fields* **189** 1045–1071. https://doi.org/10.1007/s00440-024-01264-6

Conforti, G., Durmus, A.O. and Greco, G. (2023). Quantitative contraction rates for Sinkhorn algorithm: beyond bounded costs and compact marginals. arXiv preprint. Available at arXiv:2304.04451.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 2292–2300.

Deb, N., Kim, Y.-H., Pal, S. and Schiebinger, G. (2023). Wasserstein mirror gradient flow as the limit of the Sinkhorn algorithm. Preprint. Available at arXiv:2307.16421.

Del Moral, P. (2025). Stability of Schrödinger bridges and Sinkhorn semigroups for log-concave models. arXiv preprint. Available at arXiv:2503.15963.

Deligiannidis, G., de Bortoli, V. and Doucet, A. (2024). Quantitative uniform stability of the iterative proportional fitting procedure. *Ann. Appl. Probab.* **34** 501–516. https://doi.org/10.1214/23-AAP1970

Di Marino, S. and Gerolin, A. (2020). An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *J. Sci. Comput.* **85** 27.

Divol, V., Niles-Weed, J. and Pooladian, A.-A. (2025). Tight stability bounds for entropic Brenier maps. *Int. Math. Res. Not.* **2025** rnaf078. https://doi.org/10.1093/imrn/rnaf078

Eckstein, S. (2025). Hilbert's projective metric for functions of bounded growth and exponential convergence of Sinkhorn's algorithm. *Probab. Theory Related Fields*. https://doi.org/10.1007/s00440-025-01366-9

Eckstein, S. and Lakhal, A. (2025). Exponential convergence of general iterative proportional fitting procedures. arXiv preprint. Available at arXiv:2502.20264.

Eckstein, S. and Nutz, M. (2022). Quantitative stability of regularized optimal transport and convergence of Sinkhorn's algorithm. *SIAM J. Math. Anal.* **54** 5922–5948.

Fathi, M., Gozlan, N. and Prodhomme, M. (2020). A proof of the Caffarelli contraction theorem via entropic regularization. *Calc. Var. Partial Differential Equations* **59**.

Franklin, J. and Lorenz, J. (1989). On the scaling of multidimensional matrices. *Linear Algebra Appl.* **114–115** 717–735.

Ghosal, P. and Nutz, M. (2026+). On the convergence rate of Sinkhorn's algorithm. *Math. Oper. Res.* Forthcoming. https://doi.org/10.1287/moor.2024.0427

Ghosal, P., Nutz, M. and Bernton, E. (2022). Stability of entropic optimal transport and Schrödinger bridges. *J. Funct. Anal.* **283** 109622. https://doi.org/10.1016/j.jfa.2022.109622

Greco, G. (2024). The Schrödinger problem: where analysis meets stochastics. Phd Thesis (Graduation TU/e), Mathematics and Computer Science Proefschrift.

Greco, G. and Tamanini, L. (2026). Supplement to "Hessian stability and convergence rates for entropic and Sinkhorn potentials via semiconcavity." https://doi.org/10.3150/25-BEJ1956SUPP

Greco, G., Noble, M., Conforti, G. and Durmus, A. (2023). Non-asymptotic convergence bounds for Sinkhorn iterates and their gradients: A coupling approach. In *Proceedings of Thirty Sixth Conference on Learning Theory* (G. Neu and L. Rosasco, eds.). *Proceedings of Machine Learning Research* **195** 716–746. PMLR.

Kitagawa, J., Letrouit, C. and Mérigot, Q. (2025). Stability of optimal transport maps on Riemannian manifolds. arXiv preprint. Available at arXiv:2504.05412.

Léger, F. (2021). A gradient descent perspective on Sinkhorn. *Appl. Math. Optim.* **84** 1843–1855.

Léger, F. and Aubin-Frankowski, P.-C. (2023). Gradient descent with a general cost. Preprint. Available at arXiv: 2305.04917.

Léonard, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. https://doi.org/10.3934/dcds.2014.34.1533

Letrouit, C. and Mérigot, Q. (2024). Gluing methods for quantitative stability of optimal transport maps. arXiv preprint. Available at arXiv:2411.04908.

Ma, X.-N., Trudinger, N.S. and Wang, X.-J. (2005). Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* **177** 151–183.

Malamut, H. and Sylvestre, M. (2025). Convergence rates of the regularized optimal transport: Disentangling suboptimality and entropy. *SIAM J. Math. Anal.* **57** 2533–2558. https://doi.org/10.1137/23M1591554

Mikami, T. (2004). Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields* **129** 245–260.

Nutz, M. (2021). Introduction to Entropic Optimal Transport. Available at http://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.

Nutz, M. and Wiesel, J. (2022). Entropic optimal transport: Convergence of potentials. *Probab. Theory Related Fields* **184** 401–424. https://doi.org/10.1007/s00440-021-01096-8

Pooladian, A.-A. and Niles-Weed, J. (2021). Entropic estimation of optimal transport maps. arXiv preprint. Available at arXiv:2109.12004.

Ruschendorf, L. (1995). Convergence of the iterative proportional fitting procedure. *Ann. Statist.* 1160–1174.

Sander, M.E., Ablin, P., Blondel, M. and Peyré, G. (2022). Sinkformers: Transformers with doubly stochastic attention. In *International Conference on Artificial Intelligence and Statistics* 3515–3530. PMLR.

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. https://doi.org/10.1007/978-3-319-20828-2

Schrödinger, E. (1931). Über die Umkehrung der Naturgesetze. *Sitz.ber. Preuss. Akad. Wiss. Berl.* **144** 144–153. https://doi.org/10.1063/1.527002

Schrödinger, E. (1932). La théorie relativiste de l'électron et l' interprétation de la mécanique quantique. *Ann. Inst. Henri Poincaré* **2** 269–310. https://doi.org/10.1016/j.spa.2005.04.010

Shi, Y., De Bortoli, V., Deligiannidis, G. and Doucet, A. (2022). Conditional simulation using diffusion Schrödinger bridges. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence* (J. Cussens and K. Zhang, eds.). *Proceedings of Machine Learning Research* **180** 1792–1802. PMLR.

Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Stat.* **35** 876–879.

Sinkhorn, R. and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.* **21** 343–348.

Wang, G., Jiao, Y., Xu, Q., Wang, Y. and Yang, C. (2021). Deep generative learning via Schrödinger bridge. In *Proceedings of the 38th International Conference on Machine Learning* (M. Meila and T. Zhang, eds.). *Proceedings of Machine Learning Research* **139** 10794–10804. PMLR.

Bernoulli **32**(3), 2026, 2379–2402
https://doi.org/10.3150/25-BEJ1957

Limiting laws for spiked eigenvalues and largest non-spiked eigenvalues of sample covariance matrices in elliptical distributions

JIAHUI XIE[1,a], LONG YU[2,c] and WANG ZHOU[1,b]

[1] *Department of Statistics and Data Science, National University of Singapore,* [a]*jiahui.xie@u.nus.edu,* [b]*wangzhou@nus.edu.sg*
[2] *School of Statistics and Data Science, and Institute of Data Science and Statistics, Shanghai University of Finance and Economics,* [c]*yulong@mail.shufe.edu.cn*

We investigate the asymptotic behavior of the spiked eigenvalues and the largest non-spiked eigenvalues of the sample covariance matrix TXX^*T^*, where the entries of X follow the elliptical distributions. We discuss this model under a general framework that allows for the existence of both divergent spiked eigenvalues and bounded ones, while the total number of spikes can grow to infinity. We show that some well-known results which have been proved for X with independent components still hold under elliptical distributions. Specifically, the divergent spiked sample eigenvalues converge in distribution to Gaussian limits after proper centralization and scaling. The asymptotic means depend not only on the population spikes but also on the non-spikes, while the asymptotic variances depend solely on the radius of X. For the bounded sample spiked eigenvalues, they converge in probability to certain typical locations of the limiting spectral distribution. We also provide a central limit theorem for a class of bounded spiked eigenvalues for potential statistical use. Additionally, for the largest non-spiked sample eigenvalues, the limiting Tracy-Widom law is obtained. We note that the results above are derived under the assumption that the radius of the entries in X has finite fourth moments.

Keywords: Bounded spikes; central limit theorem; divergent spikes; elliptical distributions; sample covariance matrices; Tracy-Widom distribution

References

Bai, Z.D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. MR1711663

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273

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

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

Bai, Z. and Yao, J. (2021). Erratum: Central limit theorems for eigenvalues in a spiked population model [Annales de l'Institut Henri Poincaré—Probabilités et Statistiques 2008, Vol. 44, No. 3, 447–474]. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 272. MR4255175 https://doi.org/10.1214/20-aihp1078

Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

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

Baik, J. and Silverstein, J.W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

1350-7265

Benaych-Georges, F. and Knowles, A. (2016). Lectures on the local semicircle law for Wigner matrices. Preprint. Available at arXiv:1601.04055.

Bloemendal, A., Knowles, A., Yau, H.-T. and Yin, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. MR3449395 https://doi.org/10.1007/s00440-015-0616-x

Cai, T.T., Han, X. and Pan, G. (2020). Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Ann. Statist.* **48** 1255–1280. MR4124322 https://doi.org/10.1214/18-AOS1798

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

Ding, X. (2019). Asymptotics of empirical eigen-structure for high dimensional sample covariance matrices of general form. Preprint. Available at arXiv:1708.06296.

Ding, X. and Xie, J. (2025). Tracy-Widom distribution for the edge eigenvalues of elliptical model. *Inf. Inference* **14**. Paper No. iaaf004. MR4894837 https://doi.org/10.1093/imaiai/iaaf004.

Ding, X., Xie, J., Yu, L. and Zhou, W. (2023). Extreme eigenvalues of sample covariance matrices under generalized elliptical models with applications. Preprint. Available at arXiv:2303.03532.

Dobriban, E. and Liu, S. (2019). A new theory for sketching in linear regression. Preprint. Available at arXiv:1810.06089.

Dobriban, E. and Sheng, Y. (2021). Distributed linear regression by averaging. *Ann. Statist.* **49** 918–943. MR4255113 https://doi.org/10.1214/20-aos1984

El Karoui, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248 https://doi.org/10.1214/08-AAP548

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

El Karoui, N., Bean, D., Bickel, P.J., Lim, C. and Yu, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562.

Erdësh, L. (2011). Universality of Wigner random matrices: A survey of recent results. *Uspekhi Mat. Nauk* **66** 67–198. MR2859190 https://doi.org/10.1070/RM2011v066n03ABEH004749

Fama, E.F. and French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* **33** 3–56.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653 https://doi.org/10.1111/rssb.12016

Fan, J., Liu, H. and Wang, W. (2018). Large covariance estimation through elliptical factor models. *Ann. Statist.* **46** 1383–1414. MR3819104 https://doi.org/10.1214/17-AOS1588

Fang, K.T. and Anderson, T.W., eds. (1990). *Statistical Inference in Elliptically Contoured and Related Distributions*. New York: Allerton Press. MR1066887

Gupta, A.K., Varga, T. and Bodnar, T. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory*, Second ed. New York: Springer. MR3112145 https://doi.org/10.1007/978-1-4614-8154-6

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

Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

Jung, S. and Marron, J.S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454 https://doi.org/10.1214/09-AOS709

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–438. MR0287628

Li, W. and Yao, J. (2018). On structure testing for component covariance matrices of a high dimensional mixture. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 293–318. MR3763693 https://doi.org/10.1111/rssb.12248

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

Silverstein, J.W. (2009). The Stieltjes transform and its role in eigenvalue behavior of large dimensional random matrices. In *Random Matrix Theory and Its Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **18** 1–25. Hackensack, NJ: World Sci. Publ. MR2603192 https://doi.org/10.1142/9789814273121\protect \relax \T1\textunderscore 0001

Stock, J.H. and Watson, M.W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* **97** 1167–1179. MR1951271 https://doi.org/10.1198/016214502388618960

Wang, W. and Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

Wen, J., Xie, J., Yu, L. and Zhou, W. (2022). Tracy-Widom limit for the largest eigenvalue of high-dimensional covariance matrices in elliptical distributions. *Bernoulli* **28** 2941–2967. MR4474568 https://doi.org/10.3150/21-bej1443

Xie, J., Yu, L. and Zhou, W. (2026). Supplement to "Limiting laws for spiked eigenvalues and largest non-spiked eigenvalues of sample covariance matrices in elliptical distributions." https://doi.org/10.3150/25-BEJ1957SUPP

Yang, X., Zheng, X. and Chen, J. (2021). Testing high-dimensional covariance matrices under the elliptical distribution and beyond. *J. Econometrics* **221** 409–423. MR4215033 https://doi.org/10.1016/j.jeconom.2020.05.017

Yu, L., He, Y. and Zhang, X. (2019). Robust factor number specification for large-dimensional elliptical factor model. *J. Multivariate Anal.* **174** 104543. MR4001652 https://doi.org/10.1016/j.jmva.2019.104543

Bernoulli **32**(3), 2026, 2403–2426
https://doi.org/10.3150/25-BEJ1958

Comparing moments of real log-concave random variables

DANIEL MURAWSKI[a]

University of Warsaw, Warsaw, Poland, [a]*daniel.murawski99@gmail.com*

We show that for every mean zero log-concave real random variable X one has $\|X\|_p \leq \frac{p}{q}\|X\|_q$ for $p \geq q \geq 1$, going beyond the well-known case of symmetric random variables. We also prove that in the class of arbitrary log-concave real random variables for $p > q > 0$ the quantity $\|X\|_p / \|X\|_q$ is maximized for some shifted exponential distribution. Building upon this we derive the bound $\|X\|_p \leq C_0 \frac{p}{q}\|X\|_q$ for arbitrary log-concave X, with best possible absolute constant $C_0 = e^{W(1/e)} \approx 1.3211$ in front of $\frac{p}{q}$, where W stands for the Lambert function.

Keywords: Degrees of freedom; log-concavity; moment inequalities

References

[1] Andrews, G.E., Askey, R. and Roy, R. (1999). *Special Functions*. *Encyclopedia of Mathematics and Its Applications* **71**. Cambridge: Cambridge University Press.
[2] Barlow, R.E., Marshall, A.W. and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist.* **34** 375–389.
[3] Białobrzeski, M. and Nayar, P. (2021). Rényi entropy and variance comparison for symmetric log-concave random variables. Available at arXiv:2108.10100.
[4] Borell, C. (1975). Convex set functions in d-space. *Period. Math. Hungar.* **6** 111–136.
[5] Brazitikos, S., Giannopoulos, A., Valettas, P. and Vritsiou, B.H. (2014). *Geometry of Isotropic Convex Bodies*. *Mathematical Surveys and Monographs* **196**. Providence, RI: American Mathematical Society.
[6] Eitan, Y. (2021). The centered convex body whose marginals have the heaviest tails. Available at arXiv:2110.14382.
[7] Eskenazis, A., Nayar, P. and Tkocz, T. (2018). Sharp comparison of moments and the log-concave moment problem. *Adv. Math.* **334** 389–416.
[8] Fradelizi, M. and Guédon, O. (2004). The extreme points of subsets of s-concave probabilities and a geometric localization theorem. *Discrete Comput. Geom.* **31** 327–335.
[9] Fradelizi, M. and Guédon, O. (2006). A generalized localization theorem and geometric inequalities for convex bodies. *Adv. Math.* **204** 509–529.
[10] Grünbaum, B. (1960). Partitions of mass-distributions and of convex bodies by hyperplanes. *Pacific J. Math.* **10** 1257–1261.
[11] Khintchine, A. (1923). Über dyadische Brüche. *Math. Z.* **18** 109–116.
[12] Latała, R. (2011). Weak and strong moments of random vectors. *Banach Center Publ.* **95.1** 115–121.
[13] Latała, R. and Wojtaszczyk, J.O. (2008). On the infimum convolution inequality. *Studia Math.* **189** 147–187.
[14] Ledoux, M. (2001). *The Concentration of Measure Phenomenom*. *Mathematical Surveys and Monographs* **89**. Providence, RI: American Mathematical Society.
[15] Ledoux, M. and Talagrand, M. (2011). *Probability in Banach Spaces: Isoperimetry and Processes*. *Classics in Mathematics*. Berlin: Springer-Verlag. Reprint of the 1991 edition.
[16] Madiman, M., Nayar, P. and Tkocz, T. (2021). Sharp moment-entropy inequalities and capacity bounds for log-concave distributions. *IEEE Trans. Inf. Theory* **67** no. 1, 81–94.

1350-7265

Bernoulli **32**(3), 2026, 2427–2450
https://doi.org/10.3150/25-BEJ1959

The limits of assumption-free tests for algorithm performance

YUETIAN LUO[1,a] and RINA FOYGEL BARBER[2,b]

[1]*Data Science Institute, University of Chicago, Chicago, USA,* [a]*yl2785@stat.rutgers.edu*
[2]*Department of Statistics, University of Chicago, Chicago, USA,* [b]*rina@uchicago.edu*

Algorithm evaluation and comparison are fundamental questions in machine learning and statistics—how well does an algorithm perform at a given modeling task, and which algorithm performs best? Many methods have been developed to assess algorithm performance, often based around cross-validation type strategies, retraining the algorithm of interest on different subsets of the data and assessing its performance on the held-out data points. Despite the broad use of such procedures, the theoretical properties of these methods are not yet fully understood. In this work, we explore some fundamental limits for answering these questions with limited amounts of data. In particular, we make a distinction between two questions: how good is an algorithm $\mathcal{A}$ at the problem of learning from a training set of size n, versus, how good is a particular fitted model produced by running $\mathcal{A}$ on a particular training data set of size n? Our main results prove that, for any test that treats the algorithm $\mathcal{A}$ as a "black box" (i.e., we can only study the behavior of $\mathcal{A}$ empirically), there is a fundamental limit on our ability to carry out inference on the performance of $\mathcal{A}$, unless the number of available data points N is many times larger than the evaluation sample size n of interest. On the other hand, evaluating the performance of a particular fitted model can be easy as long as the loss function is bounded and a holdout data set is available—that is, as long as $N - n$ is not too small. We also ask whether an assumption of algorithmic stability might be sufficient to circumvent this hardness result. Surprisingly, we find that the same hardness result still holds for the problem of evaluating the performance of $\mathcal{A}$, aside from a high-stability regime where fitted models are essentially nonrandom. Finally, we also establish similar hardness results for the problem of comparing multiple algorithms.

Keywords: Algorithm evaluation; algorithm risk; algorithmic stability; distribution-free inference

References

Angelopoulos, A.N., Barber, R.F. and Bates, S. (2024). Theoretical foundations of conformal prediction. arXiv preprint. Available at arXiv:2411.11824.

Angelopoulos, A.N., Bates, S., Fisch, A., Lei, L. and Schuster, T. (2024). Conformal risk control. In *The Twelfth International Conference on Learning Representations*.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Stat. Surv.* **4** 40–79.

Austern, M. and Zhou, W. (2020). Asymptotics of cross-validation. arXiv preprint. Available at arXiv:2001.11111.

Bahadur, R.R. and Savage, L.J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27** 1115–1122.

Barber, R.F. (2020). Is distribution-free inference possible for binary regression? *Electron. J. Stat.* **14** 3487–3524.

Barber, R.F., Candes, E.J., Ramdas, A. and Tibshirani, R.J. (2021a). The limits of distribution-free conditional predictive inference. *Inf. Inference* **10** 455–482.

Barber, R.F., Candès, E.J., Ramdas, A. and Tibshirani, R.J. (2021b). Predictive inference with the jackknife+. *Ann. Statist.* **49**.

Bates, S., Hastie, T. and Tibshirani, R. (2023). Cross-validation: What does it estimate and how well does it do it? *J. Amer. Statist. Assoc.* 1–12.

Bayle, P., Bayle, A., Janson, L. and Mackey, L. (2020). Cross-validation confidence intervals for test error. *Adv. Neural Inf. Process. Syst.* **33** 16339–16350.

Bouckaert, R.R. (2003). Choosing between two learning algorithms based on calibrated tests. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning* 51–58.

Bousquet, O. and Elisseeff, A. (2002). Stability and generalization. *J. Mach. Learn. Res.* **2** 499–526.

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* **24** 123–140.

Celisse, A. and Guedj, B. (2016). Stability revisited: new generalisation bounds for the leave-one-out. arXiv preprint. Available at arXiv:1608.06412.

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7** 1–30.

Devroye, L. and Wagner, T. (1979). Distribution-free inequalities for the deleted and holdout error estimates. *IEEE Trans. Inf. Theory* **25** 202–207.

Dietterich, T.G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput.* **10** 1895–1923.

Elisseeff, A., Evgeniou, T., Pontil, M. and Kaelbing, L.P. (2005). Stability of randomized learning algorithms. *J. Mach. Learn. Res.* **6**.

Garcia, S. and Herrera, F. (2008). An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons. *J. Mach. Learn. Res.* **9**.

Geisser, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70** 320–328.

Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction 2*. Springer.

Hinton, G., Neal, R., Tibshirani, R., Revow, M., Rasmussen, C., van Camp, D., Kustra, R. and Ghahramani, Z. (1995). Assessing learning procedures using DELVE Technical Report, Technical report, University of Toronto.

Iyengar, G., Lam, H. and Wang, T. (2025). Is cross-validation the gold standard to estimate out-of-sample model performance? *Adv. Neural Inf. Process. Syst.* **37** 94736–94775.

Kim, B. and Barber, R.F. (2023). Black-box tests for algorithmic stability. *Inf. Inference* **12** 2690–2719.

Kim, I., Neykov, M., Balakrishnan, S. and Wasserman, L. (2022). Local permutation tests for conditional independence. *Ann. Statist.* **50** 3388–3414.

Kumar, R., Lokshtanov, D., Vassilvitskii, S. and Vattani, A. (2013). Near-optimal bounds for cross-validation via loss stability. In *International Conference on Machine Learning* 27–35. PMLR.

Li, J. (2023). Asymptotics of K-fold cross validation. *J. Artificial Intelligence Res.* **78** 491–526.

Luo, Y. and Barber, R.F. (2024). Is algorithmic stability testable? a unified framework under computational constraints. arXiv preprint. Available at arXiv:2405.15107.

Luo, Y. and Barber, R. F. (2026). Supplement to "The limits of assumption-free tests for algorithm performance." https://doi.org/10.3150/25-BEJ1959SUPP

Medarametla, D. and Candès, E. (2021). Distribution-free conditional median inference. *Electron. J. Stat.* **15** 4625–4658.

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473.

Mourtada, J. and Gaiffas, S. (2022). An improper estimator with optimal excess risk in misspecified density estimation and logistic regression. *J. Mach. Learn. Res.* **23** 1–49.

Mourtada, J., Vaskevicius, T. and Zhivotovskiy, N. (2022). Distribution-free robust linear regression. *Math. Stat. Learn.* **4** 253–292.

Myers, G.J., Sandler, C. and Badgett, T. (2011). *The Art of Software Testing*. Wiley.

Nadeau, C. and Bengio, Y. (2003). Inference for the generalization error. *Mach. Learn.* **52** 239–281.

Neal, R.M. (1998). Assessing relevance determination methods using DELVE. *NATO ASI Ser., Ser. F: Comput. Syst. Sci.* **168** 97–132.

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint. Available at arXiv:1811.12808.

Rinaldo, A., Wasserman, L. and G'Sell, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469.

Rosset, S. and Tibshirani, R.J. (2019). From fixed-X to random-X regression: Bias-variance decompositions, covariance penalties, and prediction error estimation. *J. Amer. Statist. Assoc.*

Salzberg, S.L. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Min. Knowl. Discov.* **1** 317–328.

Shah, R.D. and Peters, J. (2020). The hardness of conditional independence testing and the generalized covariance measure. *Ann. Statist.* **48** 1514–1538.

Shalev-Shwartz, S., Shamir, O., Srebro, N. and Sridharan, K. (2010). Learnability, stability and uniform convergence. *J. Mach. Learn. Res.* **11** 2635–2670.

Shao, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc.* **88** 486–494.

Soloff, J.A., Barber, R.F. and Willett, R. (2024). Bagging provides assumption-free stability. *J. Mach. Learn. Res.* **25** 1–35.

Steinberger, L. and Leeb, H. (2023). Conditional predictive inference for high-dimensional stable algorithms. *Ann. Statist.* **51** 290–311.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Statist. Soc. Ser. B, Methodol.* **36** 111–133.

Trippe, B.L., Deshpande, S.K. and Broderick, T. (2023). Confidently comparing estimates with the c-value. *J. Amer. Statist. Assoc.* 1–12.

Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **7** 1–8.

Vovk, V., Gammerman, A. and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer.

Wibisono, A., Rosasco, L. and Poggio, T. (2009). Sufficient conditions for uniform stability of regularization algorithms. Computer Science and Artificial Intelligence Laboratory Technical Report, MIT-CSAIL-TR-2009-060 156.

Xu, H., Caramanis, C. and Mannor, S. (2011). Sparse algorithms are not stable: A no-free-lunch theorem. *IEEE Trans. Pattern Anal. Mach. Intell.* **34** 187–193.

Yang, Y. (2007). Consistency of cross validation for comparing regression procedures. *Ann. Statist.* 2450–2473.

Yu, B. (2013). Stability. *Bernoulli* **19** 1484–1500.

Bernoulli **32**(3), 2026, 2451–2472
https://doi.org/10.3150/25-BEJ1960

Optimal level set estimation for non-parametric tournament and crowdsourcing problems

MAXIMILIAN GRAF[1,a], ALEXANDRA CARPENTIER[1,b] and NICOLAS VERZELEN[2,c]

[1]*Universität Potsdam, Potsdam, Germany,* [a]*graf9@uni-potsdam.de,* [b]*carpentier@uni-potsdam.de*
[2]*INRAE, MISTEA, Univ. Montpellier, Montpellier, France,* [c]*nicolas.verzelen@inrae.fr*

Motivated by crowdsourcing, we consider a problem where we partially observe the correctness of the answers of n experts on d questions. In this paper, we assume that both the experts and the questions can be ordered, namely that the matrix M containing the probability that expert i answers correctly to question j is bi-isotonic up to a permutation of it rows and columns. When $n = d$, this also encompasses the strongly stochastic transitive (SST) model from the tournament literature. Here, we focus on the relevant problem of deciphering small entries of M from large entries of M, which is key in crowdsourcing for efficient allocation of workers to questions. More precisely, we aim at recovering a (or several) level set p of the matrix up to a precision h, namely recovering resp. the sets of positions (i, j) in M such that $M_{ij} > p + h$ and $M_{ij} < p - h$. We consider, as a loss measure, the number of misclassified entries. As our main result, we construct an efficient polynomial-time algorithm that turns out to be minimax optimal for this classification problem. This heavily contrasts with existing literature in the SST model where, for the stronger reconstruction loss, statistical-computational gaps have been conjectured. More generally, this sheds light on the nature of statistical-computational gaps for permutations models.

Keywords: Bivariate isotonic matrices; crowdsourcing; minimax estimation; noisy sorting; statistical computational gap; tournament problem

References

[1] Ballinger, T.P. and Wilcox, N.T. (1997). Decisions, error and heterogeneity. *Econ. J.* **107** 1090–1105.
[2] Bengs, V., Busa-Fekete, R., El Mesaoudi-Paul, A. and Hüllermeier, E. (2021). Preference-based online learning with dueling bandits: A survey. *J. Mach. Learn. Res.* **22** 278–385.
[3] Bong, H. and Rinaldo, A. (2022). Generalized results for the existence and consistency of the MLE in the Bradley-Terry-Luce model. In *International Conference on Machine Learning* 2160–2177. PMLR.
[4] Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39** 324–345.
[5] Braverman, M. and Mossel, E. (2008). Noisy sorting without resampling. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms* 268–276.
[6] Buhlmann, H. and Huber, P.J. (1963). Pairwise comparison and ranking in tournaments. *Ann. Math. Stat.* **34** 501–510.
[7] Chatterjee, S. and Mukherjee, S. (2019). Estimation in tournaments and graphs under monotonicity constraints. *IEEE Trans. Inf. Theory* **65** 3525–3539.
[8] Flammarion, N., Mao, C. and Rigollet, P. (2019). Optimal rates of statistical seriation. *Bernoulli* **25** 623–653.
[9] Fürnkranz, J. and Hüllermeier, E. (2010). Preference learning and ranking by pairwise comparison. In *Preference Learning* 65–82. Springer.
[10] Gao, C., Shen, Y. and Zhang, A.Y. (2023). Uncertainty quantification in the Bradley–Terry–Luce model. *Inf. Inference* **12** 1073–1140.
[11] Graf, M., Carpentier, A. and Verzelen, N. (2026). Supplement to "Optimal level set estimation for non-parametric tournament and crowdsourcing problems." https://doi.org/10.3150/25-BEJ1960SUPP

[12] Hunter, D.R. (2004). MM algorithms for generalized Bradley-Terry models. *Ann. Statist.* **32** 384–406.

[13] Liu, H., Gao, C. and Samworth, R.J. (2021). Minimax rates in sparse, high-dimensional change point detection. *Ann. Statist.* **49** 1081–1112.

[14] Liu, A. and Moitra, A. (2020). Better algorithms for estimating non-parametric models in crowd-sourcing and rank aggregation. In *Conference on Learning Theory* 2780–2829. PMLR.

[15] Luce, R.D. (2012). *Individual Choice Behavior: A Theoretical Analysis*. Courier Corporation.

[16] Mao, C., Pananjady, A. and Wainwright, M.J. (2018). Breaking the $1/\sqrt{n}$ barrier: Faster rates for permutation-based models in polynomial time. In *Conference on Learning Theory* 2037–2042. PMLR.

[17] Mao, C., Pananjady, A. and Wainwright, M.J. (2020). Towards optimal estimation of bivariate isotonic matrices with unknown permutations. *Ann. Statist.* **48** 3183–3205.

[18] Mao, C., Weed, J. and Rigollet, P. (2018). Minimax rates and efficient algorithms for noisy sorting. In *Algorithmic Learning Theory* 821–847. PMLR.

[19] McLaughlin, D.H. and Luce, R.D. (1965). Stochastic transitivity and cancellation of preferences between bitter-sweet solutions. *Psychon. Sci.* **2** 89–90.

[20] Pananjady, A. and Samworth, R.J. (2022). Isotonic regression with unknown permutations: Statistics, computation and adaptation. *Ann. Statist.* **50** 324–350.

[21] Pananjady, A., Mao, C., Muthukumar, V., Wainwright, M.J. and Courtade, T.A. (2020). Worst-case versus average-case design for estimation from partial pairwise comparisons. *Ann. Statist.* **48** 1072–1097.

[22] Pilliat, E., Carpentier, A. and Verzelen, N. (2023). Optimal permutation estimation in CrowdSourcing problems. *Ann. Statist.* **51** 935–961.

[23] Pilliat, E., Carpentier, A. and Verzelen, N. (2024). Optimal rates for ranking a permuted isotonic matrix in polynomial time. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* 3236–3273. SIAM.

[24] Raykar, V.C. and Yu, S. (2011). Ranking annotators for crowdsourced labeling tasks. In *Advances in Neural Information Processing Systems* **24**.

[25] Saad, E.M., Verzelen, N. and Carpentier, A. (2023). Active ranking of experts based on their performances in many tasks. In *International Conference on Machine Learning* 29490–29513. PMLR.

[26] Shah, N.B., Balakrishnan, S. and Wainwright, M.J. (2019). Feeling the Bern: Adaptive estimators for Bernoulli probabilities of pairwise comparisons. *IEEE Trans. Inf. Theory* **65** 4854–4874.

[27] Shah, N.B., Balakrishnan, S. and Wainwright, M.J. (2020). A permutation-based model for crowd labeling: Optimal estimation and robustness. *IEEE Trans. Inf. Theory* **67** 4162–4184.

[28] Shah, N.B. and Wainwright, M.J. (2018). Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.* **18** 1–38.

[29] Shah, N., Balakrishnan, S., Bradley, J., Parekh, A., Ramchandran, K. and Wainwright, M. (2015). Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. In *Artificial Intelligence and Statistics* 856–865. PMLR.

[30] Shah, N.B., Balakrishnan, S., Guntuboyina, A. and Wainwright, M.J. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Trans. Inf. Theory* **63** 934–959.

Bernoulli **32**(3), 2026, 2473–2500
https://doi.org/10.3150/25-BEJ1961

Functional central limit theorem for the subgraph count of the voter model on dynamic random graphs

SIMONE BALDASSARRI[1,a] and NIKOLAI KRIUKOV[2,b]

[1]*Gran Sasso Science Institute, Viale Francesco Crispi 7, 67100 L'Aquila, Italy,* [a]*simone.baldassarri@gssi.it*
[2]*Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands,* [b]*n.kriukov@uva.nl*

In this paper we consider two-opinion voter models on dynamic random graphs, in which the joint dynamics of opinions and graphs acts as *one-way feedback*, i.e., edges appear and disappear over time depending on the opinions of the two connected vertices, while the opinion dynamics is not affected by the graph structure. Our goal is to investigate the joint evolution of the entries of a *voter subgraph count vector*, i.e., vector of subgraphs where each vertex has a specific opinion, in the regime that the number of vertices grows large. The main result of this paper is a functional central limit theorem. In particular, we prove that, under a proper centering and scaling, the joint functional of the vector of subgraph counts converges to a specific multidimensional Gaussian process.

Keywords: Dynamic random graphs; functional central limit theorem; subgraph count; voter model

References

[1] Avena, L., Güldaş, H., van der Hofstad, R. and den Hollander, F. (2018). Mixing times of random walks on dynamic configuration models. *Ann. Appl. Probab.* **28** 1977–2002.
[2] Avena, L., Güldaş, H., van der Hofstad, R. and den Hollander, F. (2019). Random walks on dynamic configuration models: A trichotomy. *Stoch. Process. Appl.* **129** 3360–3375.
[3] Avena, L., Baldasso, R., Hazra, R.S., den Hollander, F. and Quattropani, M. (2024a). Discordant edges for the voter model on regular random graphs. *ALEA Lat. Amer. J. Probab. Math. Stat.* **21** 431–464.
[4] Avena, L., Capannoli, F., Hazra, R.S. and Quattropani, M. (2024b). Meeting, coalescence and consensus time on random directed graphs. *Ann. Appl. Probab.* **34** 4940–4997.
[5] Avena, L., Baldasso, R., Hazra, R.S., den Hollander, F. and Quattropani, M. (2025). The voter model on random regular graphs with random rewiring. Preprint at arXiv:2501.08703.
[6] Baldassarri, S. and Bet, G. (2022). Asymptotic normality of degree counts in a general preferential attachment model. *Markov Process. Related Fields* **28** 577–603.
[7] Baldassarri, S. and Wang, J. (2025). Parameter estimation in interacting particle systems on dynamic random networks. *Phys. Rev. E* **112**.
[8] Baldassarri, S., Braunsteins, P., den Hollander, F. and Mandjes, M. (2024). Opinion dynamics on dense dynamic random graphs. Preprint at arXiv:2410.14618.
[9] Basu, R. and Sly, A. (2017). Evolving voter model on dense random graphs. *Ann. Appl. Probab.* **27** 1235–1288.
[10] Bhattacharya, B.B., Chatterjee, A. and Janson, S. (2023). Fluctuations of subgraph counts in graphon based random graphs. *Combin. Probab. Comput.* **32** 428–464.
[11] Billingsley, P. (2013). *Convergence of Probability Measures*. Wiley.
[12] Billingsley, P. (2017). *Probability and Measure*. Wiley.
[13] Bloznelis, M., Karjalainen, J. and Leskelä, L. (2024). Normal and stable approximation to subgraph counts in superpositions of Bernoulli random graphs. *J. Appl. Probab.* **61** 401–419.
[14] Braunsteins, P., den Hollander, F. and Mandjes, M. (2022). Graphon-valued processes with vertex-level fluctuations. Preprint at arXiv:2209.01544.

1350-7265

[15] Capannoli, F. (2025). Evolution of discordant edges in the voter model on random sparse digraphs. *Electron. J. Probab.* **30** 1–24.
[16] Chatterjee, A., Dan, S. and Bhattacharya, B.B. (2024). Higher-order graphon theory: fluctuations, degeneracies, and inference. Preprint at arXiv:2404.13822.
[17] Chatterjee, S., Sivakoff, D. and Wascher, M. (2022). The effect of avoiding known infected neighbors on the persistence of a recurring infection process. *Electron. J. Probab.* **27** 1–40.
[18] Cooper, C., Frieze, A. and Radzik, T. (2010). Multiple random walks in random regular graphs. *SIAM J. Discrete Math.* **23** 1738–1761.
[19] Durrett, R., Gleeson, J.P., Lloyd, A.L., Mucha, P.J., Shi, F., Sivakoff, D., Socolar, J.E.S. and Varghese, C. (2012). Graph fission in an evolving voter model. *Proc. Natl. Acad. Sci. USA* **109** 3682–3687.
[20] Erdös, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17–61.
[21] Hazra, R.S., Kriukov, N. and Mandjes, M. (2024). Functional central limit theorem for the principal eigenvalue of dynamic Erdös–Rényi random graphs. Preprint at arXiv:2407.02686.
[22] Hazra, R.S., Kriukov, N. and Mandjes, M. (2025). Functional central limit theorem for the simultaneous subgraph count of dynamic Erdős–Rényi random graphs. *Electron. J. Probab.* **30** 1–38.
[23] Hladký, J., Pelekis, C. and Šileikis, M. (2021). A limit theorem for small cliques in inhomogeneous random graphs. *J. Graph Theory* **97** 578–599.
[24] Holley, R.A. and Liggett, T.M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.* **3** 643–663.
[25] Holme, P. and Newman, M.E.J. (2006). Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E* **74**.
[26] Jacob, E. and Mörters, P. (2017). The contact process on scale-free networks evolving by vertex updating. *R. Soc. Open Sci.* **4** 1–14.
[27] Janson, S. and Nowicki, K. (1991). The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probab. Theory Related Fields* **90** 341–375.
[28] Kaur, G. and Röllin, A. (2021). Higher-order fluctuations in dense random graph models. *Electron. J. Probab.* **26** 1–36.
[29] Liggett, T.M. (1985). *Interacting Particle Systems*. New York, USA: Springer.
[30] Nowicki, K. (1989). Asymptotic normality of graph statistics. *J. Statist. Plann. Inference* **21** 209–222.
[31] Oliveira, R.I. (2013). Mean field conditions for coalescing random walks. *Ann. Probab.* **41** 3420–3461.
[32] Owada, T. (2017). Functional central limit theorem for subgraph counting processes. *Electron. J. Probab.* **22** 1–38.
[33] Peres, Y., Sousi, P. and Steif, J.E. (2020). Mixing time for random walk on supercritical dynamical percolation. *Probab. Theory Related Fields* **176** 809–849.
[34] Peres, Y., Stauffer, A. and Steif, J.E. (2015). Random walks on dynamical percolation: Mixing times, mean squared displacement and hitting times. *Probab. Theory Related Fields* **162** 487–530.
[35] Resnick, S.I. and Samorodnitsky, G. (2016). Asymptotic normality of degree counts in a preferential attachment model. *Adv. Appl. Probab.* **48** 283–299.
[36] Ruciński, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** 1–10.
[37] Sousi, P. and Thomas, S. (2020). Cutoff for random walk on dynamical Erdös–Rényi graph. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2745–2773.

Bernoulli **32**(3), 2026, 2501–2523
https://doi.org/10.3150/25-BEJ1962

On the breakdown point of transport-based quantiles

MARCO AVELLA MEDINA[a] and ALBERTO GONZÁLEZ-SANZ[b]

Department of Statistics, Columbia University, New York, United States, [a]*marco.avella@columbia.edu,* [b]*alberto.gonzalezsanz@columbia.edu*

Recent work has used optimal transport ideas to generalize the notion of (center-outward) quantiles to dimension $d \geq 2$. We study the robustness properties of these transport-based quantiles by deriving their breakdown point, roughly, the smallest amount of contamination required to make these quantiles take arbitrarily aberrant values. We prove that the transport median defined in Chernozhukov et al. (2017) and Hallin et al. (2021) has breakdown point of $1/2$. Moreover, a point in the transport depth contour of order $\tau \in [0, 1/2]$ has breakdown point of τ. This shows that the multivariate transport depth shares the same breakdown properties as its univariate counterpart. Our proof relies on a general argument connecting the breakdown point of transport maps evaluated at a point to the Tukey depth of that point in the reference measure.

Keywords: Center-outward quantiles; contamination model; multivariate medians; optimal transport; robustness; Tukey depth

References

[1] Adrover, J. and Yohai, V. (2002). Projection estimates of multivariate location. *Ann. Statist.* **30** 1760–1781.
[2] Bernholt, T. (2006). Robust estimators are hard to compute Technical Report, Universität Dortmund.
[3] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability & Mathematical Statistics: Probability & Mathematical Statistics*. Nashville, TN: Wiley.
[4] Brualdi, R.A. (2006). *Combinatorial Matrix Classes*. *Encyclopedia of Mathematics and Its Applications* **108**. Cambridge: Cambridge University Press. MR2266203
[5] Brunel, V.-E. (2018). Concentration of the empirical level sets of Tukey's halfspace depth. *Probab. Theory Related Fields* **173** 1165–1196.
[6] Caffarelli, L.A. (1992). The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5** 99–104. MR1124980
[7] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872.
[8] Chen, M., Gao, C. and Ren, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960.
[9] Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256.
[10] Cordero-Erausquin, D. and Figalli, A. (2019). Regularity of monotone transport maps between unbounded domains. *Discrete Contin. Dyn. Syst. Ser. A* **39** 7101–7112.
[11] Davies, L.P. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* 1269–1292.
[12] Deb, N., Bhattacharya, B.B. and Sen, B. (2025). Pitman efficiency lower bounds for multivariate distribution-free tests based on optimal transport. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **00** 1–28.
[13] Deb, N. and Sen, B. (2023). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* **118** 192–207.
[14] del Barrio, E. and González-Sanz, A. (2024). Regularity of center-outward distribution functions in non-convex domains 24 880–894.

1350-7265

[15] del Barrio, E., González-Sanz, A. and Hallin, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.* **180** 104671.

[16] del Barrio, E., González-Sanz, A. and Hallin, M. (2024). Nonparametric multiple-output center-outward quantile regression. *J. Amer. Statist. Assoc.* **120** 818–832.

[17] Depersin, J. and Lecué, G. (2023). On the robustness to adversarial corruption and to heavy-tailed data of the Stahel–Donoho median of means. *Inf. Inference* **12** 814–850.

[18] Donoho, D.L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

[19] Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point 157184 157–184. Belmont CA, Wadsworth.

[20] Figalli, A. (2017). *The Monge–Ampère Equation and Its Applications*. EMS Press.

[21] Figalli, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.* **177** 413–421.

[22] Gangbo, W. and McCann, R.J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161.

[23] Ghosal, P. and Sen, B. (2022). Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *Ann. Statist.* **50** 1012–1037.

[24] González-Sanz, A., Hallin, M. and Yao, Y. (2025). Nonparametric Vector Quantile Autoregression. arXiv preprint arXiv:2510.03166.

[25] González-Sanz, A. and Sheng, S. (2024). Linearization of Monge-Ampère equations and data science applications. arXiv preprint arXiv:2408.06534.

[26] Hallin, M. and Konen, D. (2024). Multivariate quantiles: Geometric and measure-transportation-based contours. In *Applications of Optimal Transport to Economics and Related Topics* 61–78. Springer.

[27] 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

[28] Hampel, F.R. (1968). Contributions to the theory of robust estimation PhD dissertation, University of California, Berkeley.

[29] Hampel, F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.* **42** 1887–1896.

[30] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

[31] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions 196*. Wiley.

[32] Huang, Z. and Sen, B. (2023). Multivariate symmetry: Distribution-free testing via optimal transport. arXiv preprint arXiv:2305.01839.

[33] Huber, P.J. (1984). Finite sample breakdown of M-and P-estimators. *Ann. Statist.* **12** 119–126.

[34] Huber, P.J. and Ronchetti, E.M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley.

[35] Kent, J.T. and Tyler, D.E. (1996). Constrained M-estimation for multivariate location and scatter. *Ann. Statist.* **24** 1346–1370.

[36] Koltchinskii, V. (1994). Spatial quantiles and their Bahadur-Kiefer representations. In *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium, Held from September 4–9, 1993* 361–367. Springer.

[37] Konen, D. (2025). PDE characterization of geometric distribution functions and quantiles. *Bernoulli* **31** 2077–2104.

[38] Konen, D. and Paindaveine, D. (2022). Multivariate ρ-quantiles: A spatial approach. *Bernoulli* **28** 1912–1934.

[39] Konen, D. and Paindaveine, D. (2023). Spatial quantiles on the hypersphere. *Ann. Statist.* **51** 2221–2245.

[40] Konen, D. and Paindaveine, D. (2025). On the robustness of spatial quantiles. *Ann. Inst. Henri Poincaré Probab. Stat.* **62** 548–581.

[41] Lopuhaa, H.P. and Rousseeuw, P.J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* 229–248.

[42] Manole, T., Balakrishnan, S., Niles-Weed, J. and Wasserman, L. (2024). Plugin estimation of smooth optimal transport maps. *Ann. Statist.* **52** 966–998.

[43] Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* 51–67.

[44] Maronna, R.A. and Yohai, V.J. (1995). The behavior of the Stahel-Donoho robust multivariate estimator. *J. Amer. Statist. Assoc.* **90** 330–341.

[45] Maronna, R.A., Martin, R.D., Yohai, V.J. and Salibián-Barrera, M. *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. Theory and methods (with R).
[46] Massé, J.-C. (2002). Asymptotics for the Tukey median. *J. Multivariate Anal.* **81** 286–300.
[47] McCann, R.J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323.
[48] Minsker, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335.
[49] Minsker, S. and Strawn, N. (2024). The geometric median and applications to robust mean estimation. *SIAM J. Math. Data Sci.* **6** 504–533.
[50] Mosler, K. (2013). *Depth Statistics in Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather 17–34*. Berlin: Springer.
[51] Paindaveine, D. and Passeggeri, R. (2024). On the robustness of semi-discrete optimal transport. arXiv preprint arXiv:2410.19596.
[52] Peyré, G. and Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **11** 355–607.
[53] Ramsay, K., Durocher, S. and Leblanc, A. (2021). Robustness and asymptotics of the projection median. *J. Multivariate Anal.* **181** 104678.
[54] Rockafellar, R.T. (1970). On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.* **33** 209–216.
[55] Rockafellar, R.T. (1970). *Convex Analysis*. Princeton, NJ: Princeton University Press.
[56] Rockafellar, R.T. and Wets, R.J.B. (1998). *Variational Analysis*. *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **317**. Berlin: Springer-Verlag.
[57] Ronchetti, E. (2023). Robustness aspects of optimal transport. In *Research Papers in Statistical Inference for Time Series and Related Models: Essays in Honor of Masanobu Taniguchi* 445–453. Springer.
[58] Rousseeuw, P.J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.
[59] Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg, 1983*, 256–272. Springer.
[60] Segers, J. (2022). Graphical and uniform consistency of estimated optimal transport plans.
[61] Shi, H., Hallin, M., Drton, M. and Han, F. (2022). On universally consistent and fully distribution-free rank tests of vector independence. *Ann. Statist.* **50**.
[62] Tukey, J.W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* 523–531.
[63] Tyler, D.E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22** 1024–1044.
[64] Villani, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Providence, RI: American Mathematical Society.
[65] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.* 642–656.
[66] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490.
[67] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.* 461–482.

Bernoulli **32**(3), 2026, 2524–2542
https://doi.org/10.3150/25-BEJ1965

Stability of Khintchine inequalities with optimal constants between the second and the p-th moment for $p \geq 3$

JACEK JAKIMIUK[a]

Institute of Mathematics, University of Warsaw, 02-097 Warsaw, Poland, [a]*jj406165@mimuw.edu.pl*

We give a strengthening of the classical Khintchine inequality between the second and the p-th moment for $p \geq 3$ with optimal constant by adding a deficit depending on the vector of coefficients of the Rademacher sum.

Keywords: Convex functions; Khintchine inequality; Rademacher sum; Schur order

References

[1] Ahle, T.D. (2022). Sharp and simple bounds for the raw moments of the binomial and Poisson distributions. *Statist. Probab. Lett.* **182** 109306.
[2] Chasapis, G., Nayar, P. and Tkocz, T. (2022). Slicing ℓ_p-balls reloaded: Stability, planar sections in l_1. *Ann. Probab.* **50** 2344–2372.
[3] Czerwiński, W. (2008). Khintchine inequalities (in Polish). Master Thesis, University of Warsaw.
[4] De, A., Diakonikolas, I. and Servedio, R.A. (2016). A robust Khintchine inequality, and algorithms for computing optimal constants in Fourier analysis and high-dimensional geometry. *SIAM J. Discrete Math.* **30** 1058–1094.
[5] Dzindzalieta, D. and Götze, F. (2020). Half-spaces with influential variable. *Theory Probab. Appl.* **65** 114–120.
[6] Eaton, M. (1970). A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* **41** 1223–1226.
[7] Eskenazis, A., Nayar, P. and Tkocz, T. (2024). Distributional stability of the Szarek and Ball inequalities. *Math. Ann.* **389** 1161–1185.
[8] Eskenazis, A., Nayar, P. and Tkocz, T. (2024). Resilience of cube slicing in ℓ_p. *Duke Math. J.* **173** 3377–3412.
[9] Figiel, T., Hitczenko, P., Johnson, W.B., Schechtman, G. and Zinn, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349** 997–1027.
[10] Haagerup, U. (1981). The best constants in the Khintchine inequality. *Studia Math.* **70** 231–283.
[11] Khintchine, A. (1923). Über dyadische Brüche. *Math. Z.* **18** 109–116.
[12] Khot, S., Kindler, G., Mossel, E. and O'Donnell, R. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.* **37** 319–357.
[13] Komorowski, R. (1988). On the best possible constants in the Khintchine inequality for p ≥ 3. *Bull. Lond. Math. Soc.* **20** 73–75.
[14] Latała, R. and Oleszkiewicz, K. (1994). On the best constant in the Khinchine-Kahane inequality. *Studia Math.* **109** 101–104.
[15] Littlewood, J.E. (1930). On a certain bilinear form. *Quart. J. Math. Oxf. Ser.* **1** 164–174.
[16] Melbourne, J. and Roberto, C. (2022). Quantitative form of Ball's cube slicing in R^n and equality cases in min-entropy power inequality. *Proc. Amer. Math. Soc.* **150** 3595–3611.
[17] Mordhorst, O. (2017). The optimal constants in Khintchine's inequality for the case $2 < p < 3$. *Colloq. Math.* **147** 203–216.
[18] Nayar, P. and Oleszkiewicz, K. (2012). Khintchine type inequalities with optimal constants via ultra log-concavity. *Positivity* **16** 359–371.

1350-7265

[19] Nayar, P. and Tkocz, T. (2023). Extremal sections and projections of certain convex bodies: A survey. In *Harmonic Analysis and Convexity* (A. Koldobsky and A. Volberg, eds.) 343–390. Berlin, Boston: De Gruyter.

[20] Nazarov, F.L. and Podkorytov, A.N. (2000). Ball, Haagerup, and distribution functions. In *Complex Analysis, Operators, and Related Topics*. *Oper. Theory Adv. Appl.* **113** 247–267. Basel: Birkhäuser.

[21] Oleszkiewicz, K. (1999). Comparison of moments via Poincaré-type inequality. In *Advances in Stochastic Inequalities (Atlanta, GA, 1997)*. *Contemp. Math.* **234** 135–148. Providence, RI: Amer. Math. Soc.

[22] Pinelis, I. (1994). Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.* **22** 357–368.

[23] Szarek, S.J. (1976). On the best constants in the Khinchine inequality. *Studia Math.* **58** 197–208.

[24] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent random variables. *Theory Probab. Appl.* **5** 302–305.

[25] Young, R.M.G. (1976). On the best possible constants in the Khintchine inequality. *J. Lond. Math. Soc.* **14** 496–504.