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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Special Section of Memorial Articles for Willem van Zwet

Regular Articles

continued

Vol. 49, No. 5—October 2021

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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

Vol. 49, No. 5—October 2021

THE ANNALS OF STATISTICS Vol. 49, No. 5, pp. 2431–3041 October 2021

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 49, Number 5, October 2021. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

The Annals of Statistics
2021, Vol. 49, No. 5, 2431
https://doi.org/10.1214/21-AOS2112

PREFACE: SECTION OF MEMORIAL ARTICLES FOR WILLEM VAN ZWET

BY RICHARD J. SAMWORTH[1,??] AND MING YUAN[??]

[1]*Statistical Laboratory, University of Cambridge,* [??]*r.samworth@statslab.cam.ac.uk*
[2]*Department of Statistics, Columbia University,* [??]*ming.yuan@columbia.edu*

REFERENCES

BICKEL, P., FIOCCO, M., DE GUNST, M. and GÖTZE, F. (2021). Willem van Zwet's research. *Ann*. *Statist*. this issue.

FISHER, N. I. and SMITH, A. F. M. (2021). Willem van Zwet's contributions to the profession. *Ann*. *Statist*. this issue.

SAMWORTH, R. J. and YUAN, M. (2021). Supplement to "Preface: Section of memorial articles for Willem van Zwet." https://doi.org/10.1214/21-AOS2112SUPPA, https://doi.org/10.1214/21-AOS2112SUPPB

VAN ZWET, W. R. (1999). No complaints so far. Speech given by the author on the occasion of his retirement as Professor of Mathematical Statistics at the University of Leiden. *Nieuw Arch*. *Wiskd*. (4) **17** 268–279. MR1722868

VAN DE GEER, A. and KLAASSEN, C. A. J. (2021). Willem van Zwet, teacher and thesis advisor. *Ann*. *Statist*. this issue.

The Annals of Statistics
2021, Vol. 49, No. 5, 2432–2438
https://doi.org/10.1214/21-AOS2053

WILLEM VAN ZWET'S CONTRIBUTIONS TO THE PROFESSION

BY NICHOLAS I. FISHER[1,??] AND ADRIAN F. M. SMITH[2,??]

[1]*School of Mathematics & Statistics, University of Sydney,* [??]*Nicholas.Fisher@sydney.edu.au*
[2]*The Alan Turing Institute,* [??]*InstituteDirector@turing.ac.uk*

Bill van Zwet's contributions to his profession are, in their own way, comparable to his very substantial research, research supervision and teaching contributions. He had several leadership roles in professional societies, edited leading theoretical journals, started up or revived important conference series, founded a research center and played a huge role in providing colleagues in Central and Eastern Europe access to the western world of probability and statistics. And to each of these several activities, he brought farsightedness, energy and wit, qualities that shone through when he was interviewed about what he'd done (*Statist. Sci.* **24** (2009) 87–115). In fact, because Bill provided a wealth of detail in "the interview," this article focuses more on the recollections of some of the people with whom he interacted.

REFERENCES

BERAN, R. J. and FISHER, N. I. (2009). An evening spent with Bill van Zwet. *Statist. Sci.* **24** 87–115. MR2561130 https://doi.org/10.1214/08-STS261

DEN HOLLANDER, W. T. F., SENDEN, W. J. M. and VAN ZWET, W. R. (2008). EURANDOM: A decade of European stochastics. *Stat. Neerl.* **62** 256–265. MR2526272 https://doi.org/10.1111/j.1467-9574.2008.00398.x

VAN DER VAART, A.(2020). Obituary: Willem van Zwet. *IMS Bulletin* **49** 12–13.

MSC2020 subject classifications. 01A70.
Key words and phrases. Bernoulli Society, EURANDOM, Institute of Mathematical Statistics, International Statistical Institute.

The Annals of Statistics
2021, Vol. 49, No. 5, 2439–2447
https://doi.org/10.1214/21-AOS2060

WILLEM VAN ZWET'S RESEARCH

BY PETER BICKEL[1,??], MARTA FIOCCO[2,??], MATHISCA DE GUNST[3,??] AND FRIEDRICH GÖTZE[4,??]

[1]*Department of Statistics, University of California, Berkeley,* [??]*bickel.peter@gmail.com*
[2]*Mathematical Institute, Leiden University,* [??]*m.fiocco@math.leidenuniv.nl*
[3]*Department of Mathematics, Vrije Universiteit Amsterdam,* [??]*m.c.m.de.gunst@vu.nl*
[4]*Faculty of Mathematics, University of Bielefeld,* [??]*goetze@math.uni-bielefeld.de*

Willem van Zwet made deep and influential contributions to probability and statistics, which we review in this paper. Bickel and Götze collaborated with him on his major contributions to higher order asymptotics of nonlinear statistics and on resampling and the bootstrap. We relate this work to his remarkable development of the properties of the Hoeffding expansion for symmetric statistics as well as Fourier analytic tools. Fiocco and De Gunst were his students. We describe how in their theses and subsequent papers with him they developed statistical inference in two subtle stochastic models, the contact process and cell population development under a plausible regime. We also touch on his solutions of intriguing problems not related directly to his main interests.

REFERENCES

[1] ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Ann. Statist.* **4** 108–156. MR0391373

[2] BENTKUS, V., GÖTZE, F. and VAN ZWET, W. R. (1997). An Edgeworth expansion for symmetric statistics. *Ann. Statist.* **25** 851–896. MR1439326 https://doi.org/10.1214/aos/1031833676

[3] BHATTACHARYA, R. N. and RAO, R. R. (2010). *Normal Approximation and Asymptotic Expansions*. *Classics in Applied Mathematics* **64**. SIAM, Philadelphia, PA. Updated reprint of the 1986 edition [MR0855460], corrected edition of the 1976 original [MR0436272]. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

[4] BICKEL, P. J. (1974). Edgeworth expansions in non parametric statistics. *Ann. Statist.* **2** 1–20. MR0350952

[5] BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1986). The Edgeworth expansion for U-statistics of degree two. *Ann. Statist.* **14** 1463–1484. MR0868312 https://doi.org/10.1214/aos/1176350170

[6] BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31. MR1441142

[7] BICKEL, P. J. and VAN ZWET, W. R. (1978). Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Ann. Statist.* **6** 937–1004. MR0499567

[8] BICKEL, P. J. and VAN ZWET, W. R. (1980). On a theorem of Hoeffding. In *Asymptotic Theory of Statistical Tests and Estimation* (*Proc. Adv. Internat. Sympos., Univ. North Carolina, Chapel Hill, N.C.*, 1979) 307–324. Academic Press, New York. MR0571346

[9] BLOZNELIS, M. and GÖTZE, F. (2021). Edgeworth approximations for distributions of symmetric statistics. Available at https://arxiv.org/abs/2102.03589.

[10] CHIBISOV, D. M. and VAN ZWET, W. R. (1984). On the Edgeworth expansion for the logarithm of the likelihood ratio. I. *Teor. Veroyatn. Primen.* **29** 417–439. MR0761128

[11] CHIBISOV, D. M. and VAN ZWET, W. R. (1984). On the Edgeworth expansion for the logarithm of the likelihood ratio. II. In *Asymptotic Statistics*, 2 (*Kutná Hora*, 1983) 451–461. Elsevier, Amsterdam. MR0785424

[12] CHRISTENSEN, J. P. R. and RESSEL, P. (1981). A probabilistic characterisation of negative definite and completely alternating functions. *Z. Wahrsch. Verw. Gebiete* **57** 407–417. MR0629534 https://doi.org/10.1007/BF00534832

MSC2020 subject classifications. Primary 01A70; secondary 62-03.
Key words and phrases. Van Zwet, asymptotic approximation, resampling, contact process.

[13] DE GUNST, M. C. M. and VAN ZWET, W. R. (1992). A non-Markovian model for cell population growth: Speed of convergence and central limit theorem. *Stochastic Process. Appl.* **41** 297–324. MR1164182 https://doi.org/10.1016/0304-4149(92)90129-E

[14] DE GUNST, M. C. M. and VAN ZWET, W. R. (1993). A non-Markovian model for cell population growth: Tail behavior and duration of the growth process. *Ann. Appl. Probab.* **3** 1112–1144. MR1241037

[15] DI BUCCHIANICO, A., HUŠKOVÁ, M., KLÁŠTERECKÝ, P. and VAN ZWET, W. R. (2004). Performance of control charts for specific alternative hypotheses. In *COMPSTAT* 2004—*Proceedings in Computational Statistics* 903–910. Physica, Heidelberg. MR2173090

[16] DIKS, C., VAN ZWET, W. R., TAKENS, F. and DEGOEDE, J. (1996). Detecting differences between delay vector distributions. *Phys. Rev. E* (3) **53** 2169–2176. MR1388210 https://doi.org/10.1103/PhysRevE.53.2169

[17] FIOCCO, M. and VAN ZWET, W. R. (1998). On the shape theorem for the supercritical contact process. *Prague Stochastics*. Union of Czech Mathematicians and Physicists, Prague.

[18] FIOCCO, M. and VAN ZWET, W. R. (2000). Statistics for the contact process. In *Symposium in Honour of Ole E. Barndorff–Nielsen* (*Aarhus*, 2000). *Memoirs* **16** 102–109. Univ. Aarhus, Aarhus. MR1985915

[19] FIOCCO, M. and VAN ZWET, W. R. (2003). Decaying correlations for the supercritical contact process conditioned on survival. *Bernoulli* **9** 763–781. MR2047685 https://doi.org/10.3150/bj/1066418877

[20] FIOCCO, M. and VAN ZWET, W. R. (2003). Parameter estimation for the supercritical contact process. *Bernoulli* **9** 1071–1092. MR2046818 https://doi.org/10.3150/bj/1072215201

[21] FIOCCO, M. and VAN ZWET, W. R. (2004). Maximum likelihood estimation for the contact process. In *A Festschrift for Herman Rubin*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **45** 309–318. IMS, Beachwood, OH. MR2126906 https://doi.org/10.1214/lnms/1196285399

[22] FIOCCO, M. and VAN ZWET, W. R. (2006). Maximum likelihood for the fully observed contact process. *J. Comput. Appl. Math.* **186** 117–129. MR2190301 https://doi.org/10.1016/j.cam.2005.01.037

[23] GÖTZE, F. (1981). On Edgeworth expansions in Banach spaces. *Ann. Probab.* **9** 852–859. MR0628877

[24] GÖTZE, F. (1993). *IMS Bulletin*. Special Invited Lecture in Annual joint IMS/ASA meeting on Statistics and Probability, San Francisco, Aug. 1993.

[25] GÖTZE, F. and VAN ZWET, W. R. (2006). An expansion for a discrete non-lattice distribution. In *Frontiers in Statistics* 257–274. Imp. Coll. Press, London. MR2326005 https://doi.org/10.1142/9781860948886_0012

[26] GUNST, M. D., HARKES, P., VAL, J., ZWET, W. R. and LIBBENGA, K. (1990). Modelling the growth of a batch culture of plant cells: A corpuscular approach. *Enzyme Microb. Technol.* **12** 61–71. https://doi.org/10.1016/0141-0229(90)90182-P

[27] HÁJEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Stat.* **39** 325–346. MR0222988 https://doi.org/10.1214/aoms/1177698394

[28] HELMERS, R. and VAN ZWET, W. R. (1982). The Berry–Esseen bound for U-statistics. In *Statistical Decision Theory and Related Topics*, *III*, *Vol.* 1 (*West Lafayette*, *Ind.*, 1981) 497–512. Academic Press, New York. MR0705302

[29] HODGES, J. L. JR. and LEHMANN, E. L. (1970). Deficiency. *Ann. Math. Stat.* **41** 783–801. MR0272092 https://doi.org/10.1214/aoms/1177696959

[30] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

[31] HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Stat.* **27** 713–721. MR0080391 https://doi.org/10.1214/aoms/1177728178

[32] LITVAK, N. and VAN ZWET, W. R. (2004). On the minimal travel time needed to collect n items on a circle. *Ann. Appl. Probab.* **14** 881–902. MR2052907 https://doi.org/10.1214/105051604000000152

[33] MASON, D. M. and VAN ZWET, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884. MR0893903

[34] MASON, D. M. and VAN ZWET, W. R. (1987). A note on the strong approximation to the renewal process. *Publ. Inst. Stat. Univ. Paris* **32** 81–91. MR0899236

[35] MISES, R. V. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.* **18** 309–348. MR0022330 https://doi.org/10.1214/aoms/1177730385

[36] PFANZAGL, J. (1979). First order efficiency implies second order efficiency. In *Contributions to Statistics* 167–196. Reidel, Dordrecht, MA. MR0561268

[37] PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory*. *Lecture Notes in Statistics* **13**. Springer, New York. With the assistance of W. Wefelmeyer. MR0675954

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

[39] PUTTER, H. and VAN ZWET, W. R. (1996). Resampling: Consistency of substitution estimators. *Ann. Statist.* **24** 2297–2318. MR1425955 https://doi.org/10.1214/aos/1032181156

[40] PUTTER, H. and VAN ZWET, W. R. (1998). Empirical Edgeworth expansions for symmetric statistics. *Ann. Statist.* **26** 1540–1569. MR1647697 https://doi.org/10.1214/aos/1024691253

[41] PYKE, R. and VAN ZWET, W. R. (2004). Weak convergence results for the Kakutani interval splitting procedure. *Ann. Probab.* **32** 380–423. MR2040787 https://doi.org/10.1214/aop/1078415840

[42] RÉNYI, A. (1970). *Probability Theory*. North-Holland, Amsterdam. Translated by László Vekerdi, North-Holland Series in Applied Mathematics and Mechanics, Vol. 10. MR0315747

[43] VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. *Mathematical Centre Tracts* **7**. Mathematisch Centrum, Amsterdam. MR0176511

[44] VAN ZWET, W. R. (1977). Asymptotic expansions for the distribution functions of linear combinations of order statistics. In *Statistical Decision Theory and Related Topics*, *II* (*Proc. Sympos.*, *Purdue Univ.*, *Lafayette*, *Ind.*, 1976) 421–437. MR0438586

[45] VAN ZWET, W. R. (1978). A proof of Kakutani's conjecture on random subdivision of longest intervals. *Ann. Probab.* **6** 133–137. MR0478307 https://doi.org/10.1214/aop/1176995617

[46] VAN ZWET, W. R. (1979). The Edgeworth expansion for linear combinations of uniform order statistics. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics* (*Hradec Králové*, 1978) 93–101. North-Holland, Amsterdam. MR0571178

[47] VAN ZWET, W. R. (1980). A strong law for linear functions of order statistics. *Ann. Probab.* **8** 986–990. MR0586781

[48] VAN ZWET, W. R. (1982). On the Edgeworth expansion for the simple linear rank statistic. In *Nonparametric Statistical Inference*, *Vol. I*, *II* (*Budapest*, 1980). *Colloquia Mathematica Societatis János Bolyai* **32** 889–909. North-Holland, Amsterdam. MR0719749

[49] VAN ZWET, W. R. (1984). A Berry–Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete* **66** 425–440. MR0751580 https://doi.org/10.1007/BF00533707

[50] VAN ZWET, W. R. (1994). The asymptotic distribution of point charges on a conducting sphere. In *Statistical Decision Theory and Related Topics*, *V* (*West Lafayette*, *IN*, 1992) 427–430. Springer, New York. MR1286318

[51] VAN ZWET, W. R. and OOSTERHOFF, J. (1967). On the combination of independent test statistics. *Ann. Math. Stat.* **38** 659–680. MR0216645 https://doi.org/10.1214/aoms/1177698861

The Annals of Statistics
2021, Vol. 49, No. 5, 2448–2456
https://doi.org/10.1214/21-AOS2069

WILLEM VAN ZWET, TEACHER AND THESIS ADVISOR

EDITED BY SARA VAN DE GEER[1,??] AND CHRIS A. J. KLAASSEN[2,??]

[1]*Department of Mathematics, ETH Zürich,* ??*sara.vandegeer@stat.math.ethz.ch*
[2]*Korteweg-de Vries Institute for Mathematics, University of Amsterdam,* ??*c.a.j.klaassen@uva.nl*

Willem van Zwet was supervisor of sixteen PhD students. All of them pursued academic careers and most of them became full professor. Below are some stories of PhD students Wim Albers, Cees Diks, Ronald Does, Marta Fiocco, Sara van de Geer, Mathisca de Gunst, Chris Klaassen, Hein Putter, Aad van der Vaart, Marten Wegkamp and Martien van Zuijlen with in addition a contribution by Nelly Litvak who was guided by Willem after her PhD.

MSC2020 subject classifications. Primary 01A70; secondary 62-03.
Key words and phrases. Obituary, research supervision, PhD students, teaching.

The Annals of Statistics
2021, Vol. 49, No. 5, 2457–2490
https://doi.org/10.1214/20-AOS2041

OPTIMAL RATES FOR INDEPENDENCE TESTING VIA U-STATISTIC PERMUTATION TESTS

BY THOMAS B. BERRETT[1,??], IOANNIS KONTOYIANNIS[2,*] AND RICHARD J. SAMWORTH[2,†]

[1]*Department of Statistics, University of Warwick,* [??]*tom.berrett@warwick.ac.uk*
[2]*Statistical Laboratory, Centre for Mathematical Sciences,* [*]*yiannis@maths.cam.ac.uk,* [†]*r.samworth@statslab.cam.ac.uk*

We study the problem of independence testing given independent and identically distributed pairs taking values in a σ-finite, separable measure space. Defining a natural measure of dependence $D(f)$ as the squared L^2-distance between a joint density f and the product of its marginals, we first show that there is no valid test of independence that is uniformly consistent against alternatives of the form $\{f : D(f) \geq \rho^2\}$. We therefore restrict attention to alternatives that impose additional Sobolev-type smoothness constraints, and define a permutation test based on a basis expansion and a U-statistic estimator of $D(f)$ that we prove is minimax optimal in terms of its separation rates in many instances. Finally, for the case of a Fourier basis on $[0, 1]^2$, we provide an approximation to the power function that offers several additional insights. Our methodology is implemented in the `R` package `USP`.

REFERENCES

ALBERT, M. (2015). Tests of independence by bootstrap and permutation: An asymptotic and non-asymptotic study. Application to neurosciences. Ph.D. thesis. Available at https://tel.archives-ouvertes.fr/tel-01274647/file/2015NICE4079.pdf.

ALBERT, M., BOURET, Y., FROMONT, M. and REYNAUD-BOURET, P. (2015). Bootstrap and permutation tests of independence for point processes. *Ann. Statist.* **43** 2537–2564. MR3405603 https://doi.org/10.1214/15-AOS1351

ANTOCH, J. and HUŠKOVÁ, M. (2001). Permutation tests in change point analysis. *Statist. Probab. Lett.* **53** 37–46. MR1843339 https://doi.org/10.1016/S0167-7152(01)00009-8

BERRETT, T. B., GROSE, D. J. and SAMWORTH, R. J. (2018). IndepTest: Nonparametric independence tests based on entropy estimation. *R Package version 0.2.0*. Available at https://cran.r-project.org/web/packages/IndepTest/index.html.

BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2020). USP: U-statistic permutation tests of independence for all data types, with improvement on Pearson's chi-squared test for discrete data. *R package version 0.1.0*. Available at https://cran.r-project.org/web/packages/USP/index.html.

BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2021). Supplement to "Optimal rates for independence testing via U-statistic permutation tests." https://doi.org/10.1214/20-AOS2041SUPP

BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. MR4060981

BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. With a foreword by Michel Ledoux. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

CHUNG, E. and ROMANO, J. P. (2013). Exact and asymptotically robust permutation tests. *Ann. Statist.* **41** 484–507. MR3099111 https://doi.org/10.1214/13-AOS1090

CHUNG, E. and ROMANO, J. P. (2016). Asymptotically valid and exact permutation tests based on two-sample U-statistics. *J. Statist. Plann. Inference* **168** 97–105. MR3412224 https://doi.org/10.1016/j.jspi.2015.07.004

MSC2020 subject classifications. 62G09, 62G10.
Key words and phrases. Independence testing, permutation tests, minimax separation rates, U-statistics, Stein's method.

DE JONG, P. (1990). A central limit theorem for generalized multilinear forms. *J. Multivariate Anal.* **34** 275–289. MR1073110 https://doi.org/10.1016/0047-259X(90)90040-O

DEB, N. and SEN, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. Available at arXiv:1909.08733.

DIAKONIKOLAS, I. and KANE, D. M. (2016). A new approach for testing properties of discrete distributions. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 685–694. IEEE Comput. Soc., Los Alamitos, CA. MR3631031

DÖBLER, C. and PECCATI, G. (2017). Quantitative de Jong theorems in any dimension. *Electron. J. Probab.* **22** Paper No. 2. MR3613695 https://doi.org/10.1214/16-EJP19

DÖBLER, C. and PECCATI, G. (2019). Quantitative CLTs for symmetric U-statistics using contractions. *Electron. J. Probab.* **24** Paper No. 5. MR3916325 https://doi.org/10.1214/19-EJP264

ERMAKOV, M. S. (1990). Asymptotically minimax tests for nonparametric hypotheses concerning the distribution density. *J. Sov. Math.* **52** 2891–2898.

FISHER, R. A. (1935). *The Design of Experiments*, 1st ed. Oliver & Boyd, Edinburgh.

GABRYS, R. and KOKOSZKA, P. (2007). Portmanteau test of independence for functional observations. *J. Amer. Statist. Assoc.* **102** 1338–1348. MR2412554 https://doi.org/10.1198/016214507000001111

GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert–Schmidt norms. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909 https://doi.org/10.1007/11564089_7

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

HELLER, R., HELLER, Y., KAUFMAN, S., BRILL, B. and GORFINE, M. (2016). Consistent distribution-free K-sample and independence tests for univariate random variables. *J. Mach. Learn. Res.* **17** Paper No. 29. MR3491123

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

HOFERT, M., KOJADINOVIC, I., MÄCHLER, M. and YAN, J. (2017). copula: Multivariate dependence with copulas. *R Package version 0.999-18*. Available at https://cran.r-project.org/web/packages/copula/index.html.

INGSTER, Y. I. (1989). Asymptotic minimax testing of independence hypothesis. *J. Sov. Math.* **44** 466–476.

INGSTER, Y. I. (1996). Minimax testing of the hypothesis of independence for ellipsoids in ℓ_p. *J. Math. Sci.* **81** 2406–2420.

JANSSEN, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239–253. MR1762910 https://doi.org/10.1214/aos/1016120371

KAHANE, J.-P. (1997). A century of interplay between Taylor series, Fourier series and Brownian motion. *Bull. Lond. Math. Soc.* **29** 257–279. MR1435557 https://doi.org/10.1112/S0024609396002913

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–89.

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Minimax optimality of permutation tests. Available at https://arxiv.org/abs/2003.13208.

KOJADINOVIC, I. and HOLMES, M. (2009). Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process. *J. Multivariate Anal.* **100** 1137–1154. MR2508377 https://doi.org/10.1016/j.jmva.2008.10.013

LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981 https://doi.org/10.1214/aos/1032894458

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

LI, T. and YUAN, M. (2019). On the optimality of Gaussian kernel based nonparametric tests against smooth alternatives. Available at arXiv:1909.03302.

MEYNAOUI, A., ALBERT, M., LAURENT, B. and MARREL, A. (2019). Adaptive test of independence based on HSIC measures. Available at arXiv:1902.06441.

NGUYEN, D. and EISENSTEIN, J. (2017). A kernel independence test for geographical language variation. *Comput. Linguist.* **43** 567–592. MR3708620 https://doi.org/10.1162/COLI_a_00293

PATEFIELD, W. M. (1981). Algorithm AS159. An efficient method of generating $r \times c$ tables with given row and column totals. *J. Roy. Statist. Soc. Ser. C* **30** 91–97.

PEARSON, K. (1920). Notes on the history of correlation. *Biometrika* **13** 25–45.

PESARIN, F. and SALMASO, L. (2010). *Permutation Tests for Complex Data. Theory, Applications and Software*. Wiley, Chichester, UK.

PFISTER, N. and PETERS, J. (2017). dHSIC: Independence testing via Hilbert Schmidt independence criterion. *R Package version 2.0*. https://cran.r-project.org/web/packages/dHSIC/index.html.

PFISTER, N., BÜHLMANN, P., SCHÖLKOPF, B. and PETERS, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. MR3744710 https://doi.org/10.1111/rssb.12235

PITMAN, E. J. G. (1938). Significance tests which may be applied to samples from any populations: III. The analysis of variance test. *Biometrika* **29** 322–335.

RESHEF, D. N., RESHEF, Y. A., FINUCANE, H. K., GROSSMAN, S. R., MCVEAN, G., TURNBAUGH, P. J., LANDER, E. S., MITZENMACHER, M. and SABETI, P. C. (2011). Detecting novel associations in large data sets. *Science* **334** 1518–1524.

RINOTT, Y. and ROTAR, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7** 1080–1105. MR1484798 https://doi.org/10.1214/aoap/1043862425

RIZZO, M. L. and SZEKELY, G. J. (2017). energy: E-statistics: Multivariate inference via the energy of data. *R Package version 1.7-2*. Available at: https://cran.r-project.org/web/packages/energy/index.html.

ROMANO, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17** 141–159. MR0981441 https://doi.org/10.1214/aos/1176347007

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. MR3127866 https://doi.org/10.1214/13-AOS1140

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0595165

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

SHI, H., DRTON, M. and HAN, F. (2020). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2020.1782223

SONG, L., SMOLA, A., GRETTON, A., BEDO, J. and BORGWARDT, K. (2012). Feature selection via dependence maximization. *J. Mach. Learn. Res.* **13** 1393–1434. MR2930643

SPEARMAN, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* **15** 72–101.

STEUER, R., KURTHS, J., DAUB, C. O., WEISE, J. and SELBIG, J. (2002). The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics* **18** 231–240.

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

The Annals of Statistics
2021, Vol. 49, No. 5, 2491–2505
https://doi.org/10.1214/20-AOS2043

VARIABLE SELECTION CONSISTENCY OF GAUSSIAN PROCESS REGRESSION

BY SHENG JIANG* AND SURYA T. TOKDAR†

Department of Statistical Science, Duke University, *sheng.jiang@duke.edu, †surya.tokdar@duke.edu

Bayesian nonparametric regression under a rescaled Gaussian process prior offers smoothness-adaptive function estimation with near minimax-optimal error rates. Hierarchical extensions of this approach, equipped with stochastic variable selection, are known to also adapt to the unknown intrinsic dimension of a sparse true regression function. But it remains unclear if such extensions offer variable selection consistency, that is, if the true subset of important variables could be consistently learned from the data. It is shown here that variable consistency may indeed be achieved with such models at least when the true regression function has finite smoothness to induce a polynomially larger penalty on inclusion of false positive predictors. Our result covers the high-dimensional asymptotic setting where the predictor dimension is allowed to grow with the sample size. The proof utilizes Schwartz theory to establish that the posterior probability of wrong selection vanishes asymptotically. A necessary and challenging technical development involves providing sharp upper and lower bounds to small ball probabilities at all rescaling levels of the Gaussian process prior, a result that could be of independent interest.

REFERENCES

[1] AURZADA, F., IBRAGIMOV, I. A., LIFSHITS, M. A. and VAN ZANTEN, K. Y. (2009). Small deviations of smooth stationary Gaussian processes. *Theory Probab. Appl.* **53** 697–707.

[2] BERTIN, K. and LECUÉ, G. (2008). Selection of variables and dimension reduction in high-dimensional non-parametric regression. *Electron. J. Stat.* **2** 1224–1241. MR2461900 https://doi.org/10.1214/08-EJS327

[3] BHATTACHARYA, A., PATI, D. and DUNSON, D. (2014). Anisotropic function estimation using multi-bandwidth Gaussian processes. *Ann. Statist.* **42** 352–381. MR3189489 https://doi.org/10.1214/13-AOS1192

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

[5] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

[6] BULL, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490–1516. MR2988456 https://doi.org/10.1214/12-EJS720

[7] BUTUCEA, C. and STEPANOVA, N. (2017). Adaptive variable selection in nonparametric sparse additive models. *Electron. J. Stat.* **11** 2321–2357. MR3656494 https://doi.org/10.1214/17-EJS1275

[8] CASTILLO, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* **2** 1281–1299. MR2471287 https://doi.org/10.1214/08-EJS273

[9] CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

[10] COMMINGES, L. and DALALYAN, A. S. (2012). Tight conditions for consistency of variable selection in the context of high dimensionality. *Ann. Statist.* **40** 2667–2696. MR3097616 https://doi.org/10.1214/12-AOS1046

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

MSC2020 subject classifications. Primary 62G08, 62G20; secondary 60G05.
Key words and phrases. Gaussian process priors, high-dimensional regression, nonparametric variable selection, Bayesian inference, adaptive estimation.

[12] GHOSAL, S., LEMBER, J. and VAN DER VAART, A. (2008). Nonparametric Bayesian model selection and averaging. *Electron. J. Stat.* **2** 63–89. MR2386086 https://doi.org/10.1214/07-EJS090

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

[14] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[15] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738

[16] HAHN, P. R. and CARVALHO, C. M. (2015). Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective. *J. Amer. Statist. Assoc.* **110** 435–448. MR3338514 https://doi.org/10.1080/01621459.2014.993077

[17] JIANG, S. and TOKDAR, S. T. (2021). Supplement to "Variable selection consistency of Gaussian process regression." https://doi.org/10.1214/20-AOS2043SUPP.

[18] KUELBS, J. and LI, W. V. (1993). Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.* **116** 133–157. MR1237989 https://doi.org/10.1006/jfan.1993.1107

[19] LAFFERTY, J. and WASSERMAN, L. (2008). Rodeo: Sparse, greedy nonparametric regression. *Ann. Statist.* **36** 28–63. MR2387963 https://doi.org/10.1214/009053607000000811

[20] LI, W. V. and LINDE, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27** 1556–1578. MR1733160 https://doi.org/10.1214/aop/1022677459

[21] NARISETTY, N. N. and HE, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. MR3210987 https://doi.org/10.1214/14-AOS1207

[22] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704

[23] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

[24] RAY, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. MR3737900 https://doi.org/10.1214/16-AOS1533

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

[26] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270

[27] TOKDAR, S. T. (2011). Dimension adaptability of Gaussian process models with variable selection and projection. Preprint . Available at arXiv:1112.0716.

[28] VAN DER VAART, A. and VAN ZANTEN, H. (2011). Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.* **12** 2095–2119. MR2819028

[29] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Inst. Math. Stat. (IMS) Collect.* **3** 200–222. IMS, Beachwood, OH. MR2459226 https://doi.org/10.1214/074921708000000156

[30] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.* **37** 2655–2675. MR2541442 https://doi.org/10.1214/08-AOS678

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

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

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

[34] YANG, Y. and TOKDAR, S. T. (2015). Minimax-optimal nonparametric regression in high dimensions. *Ann. Statist.* **43** 652–674. MR3319139 https://doi.org/10.1214/14-AOS1289

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

[36] ZHANG, P. (1991). Variable selection in nonparametric regression with continuous covariates. *Ann. Statist.* **19** 1869–1882. MR1135153 https://doi.org/10.1214/aos/1176348375

The Annals of Statistics
2021, Vol. 49, No. 5, 2506–2530
https://doi.org/10.1214/20-AOS2044

OPTIMALITY OF SPECTRAL CLUSTERING IN THE GAUSSIAN MIXTURE MODEL

BY MATTHIAS LÖFFLER[1,??], ANDERSON Y. ZHANG[2,??], AND HARRISON H. ZHOU[3,??]

[1]*Seminar for Statistics, ETH Zürich,* [??]*matthias.loeffler@stat.math.ethz.ch*
[2]*Department of Statistics, University of Pennsylvania,* [??]*ayz@wharton.upenn.edu*
[3]*Department of Statistics and Data Science, Yale University,* [??]*huibin.zhou@yale.edu*

Spectral clustering is one of the most popular algorithms to group high-dimensional data. It is easy to implement and computationally efficient. Despite its popularity and successful applications, its theoretical properties have not been fully understood. In this paper, we show that spectral clustering is minimax optimal in the Gaussian mixture model with isotropic covariance matrix, when the number of clusters is fixed and the signal-to-noise ratio is large enough. Spectral gap conditions are widely assumed in the literature to analyze spectral clustering. On the contrary, these conditions are not needed to establish optimality of spectral clustering in this paper.

REFERENCES

[1] ABBE, E., FAN, J. and WANG, K. (2020). An ℓ_p-theory of PCA and spectral clustering. Preprint.
[2] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854
[3] ALPERT, C. J. and YAO, S. (1995). Spectral partitioning: The more eigenvectors, the better. In 32*nd Design Automation Conference* 195–200. IEEE, New York.
[4] ANANDKUMAR, A., GE, R., HSU, D. and KAKADE, S. M. (2014). A tensor approach to learning mixed membership community models. *J. Mach. Learn. Res.* **15** 2239–2312. MR3231594
[5] BACH, F. R. and JORDAN, M. I. (2006). Learning spectral clustering, with application to speech separation. *J. Mach. Learn. Res.* **7** 1963–2001. MR2274430
[6] BALAKRISHNAN, S., XU, M., KRISHNAMURTHY, A. and SINGH, A. (2011). Noise thresholds for spectral clustering. In *Advances in Neural Information Processing Systems* 954–962.
[7] BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.
[8] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *Conference on Learning Theory* 35.1–35.23.
[9] CHEN, X. and YANG, Y. (2020). Cutoff for exact recovery of Gaussian mixture models. Preprint.
[10] COJA-OGHLAN, A. (2010). Graph partitioning via adaptive spectral techniques. *Combin. Probab. Comput.* **19** 227–284. MR2593622 https://doi.org/10.1017/S0963548309990514
[11] DHILLON, I. S. (2001). Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 269–274. ACM, New York.
[12] DING, C., HE, X. and SIMON, H. D. (2005). On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the* 2005 *SIAM International Conference on Data Mining* 606–610. SIAM, Philadelphia.
[13] DING, C. H. Q., HE, X., ZHA, H., GU, M. and SIMON, H. D. (2001). A min-max cut algorithm for graph partitioning and data clustering. In *Proceedings* 2001 *IEEE International Conference on Data Mining* 107–114. IEEE, New York.
[14] DING, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli* **26** 387–417. MR4036038 https://doi.org/10.3150/19-BEJ1129
[15] DONATH, W. E. and HOFFMAN, A. J. (2003). Lower bounds for the partitioning of graphs. In *Selected Papers of Alan J. Hoffman*: *With Commentary* 437–442. World Scientific, Singapore.

MSC2020 subject classifications. 62H30.
Key words and phrases. Spectral clustering, K-means, Gaussian mixture model, Spectral perturbation.

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

[17] FIEDLER, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Math. J.* **23** 298–305. MR0318007

[18] FISHKIND, D. E., SUSSMAN, D. L., TANG, M., VOGELSTEIN, J. T. and PRIEBE, C. E. (2013). Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. *SIAM J. Matrix Anal. Appl.* **34** 23–39. MR3032990 https://doi.org/10.1137/120875600

[19] FURUI, S. (1989). Unsupervised speaker adaptation based on hierarchical spectral clustering. *IEEE Trans. Acoust. Speech Signal Process.* **37** 1923–1930.

[20] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2018). Community detection in degree-corrected block models. *Ann. Statist.* **46** 2153–2185. MR3845014 https://doi.org/10.1214/17-AOS1615

[21] GINÉ, E. and KOLTCHINSKII, V. (2006). Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results. In *High Dimensional Probability. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **51** 238–259. IMS, Beachwood, OH. MR2387773 https://doi.org/10.1214/074921706000000888

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

[23] GIRAUD, C. and VERZELEN, N. (2018). Partial recovery bounds for clustering with the relaxed K-means. *Math. Stat. Learn.* **1** 317–374. MR4059724

[24] GUATTERY, S. and MILLER, G. L. (1998). On the quality of spectral separators. *SIAM J. Matrix Anal. Appl.* **19** 701–719. MR1611179 https://doi.org/10.1137/S0895479896312262

[25] HALL, K. M. (1970). An r-dimensional quadratic placement algorithm. *Manage. Sci.* **17** 219–229.

[26] HAN, X., TONG, X. and FAN, Y. (2020). Eigen selection in spectral clustering: A theory guided practice. Preprint.

[27] HEIN, M. (2006). Uniform convergence of adaptive graph-based regularization. In *Learning Theory. Lecture Notes in Computer Science* **4005** 50–64. Springer, Berlin. MR2277918 https://doi.org/10.1007/11776420_7

[28] HEIN, M., AUDIBERT, J.-Y. and VON LUXBURG, U. (2005). From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians. In *Learning Theory. Lecture Notes in Computer Science* **3559** 470–485. Springer, Berlin. MR2203281 https://doi.org/10.1007/11503415_32

[29] HENDRICKSON, B. and LELAND, R. (1995). An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM J. Sci. Comput.* **16** 452–469. MR1317066 https://doi.org/10.1137/0916028

[30] INABA, M., KATOH, N. and IMAI, H. (1994). Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering. In *Proceedings of* 10*th ACM Symposium on Computational Geometry* 332–339.

[31] JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

[32] JOHNSTONE, I. M. and PAUL, D. (2018). PCA in high dimensions: An orientation. *Proc. IEEE Inst. Electr. Electron. Eng.* **106** 1277–1292. https://doi.org/10.1109/JPROC.2018.2846730

[33] KANNAN, R. and VEMPALA, S. (2009). Spectral algorithms. *Found. Trends Theor. Comput. Sci.* **4** 157–288. MR2558901 https://doi.org/10.1561/0400000025

[34] KANNAN, R., VEMPALA, S. and VETTA, A. (2004). On clusterings: Good, bad and spectral. *J. ACM* **51** 497–515. MR2145863 https://doi.org/10.1145/990308.990313

[35] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1976–2013. MR3573302 https://doi.org/10.1214/15-AIHP705

[36] KOLTCHINSKII, V. and XIA, D. (2016). Perturbation of linear forms of singular vectors under Gaussian noise. In *High Dimensional Probability VII. Progress in Probability* **71** 397–423. Springer, Cham. MR3565274 https://doi.org/10.1007/978-3-319-40519-3_18

[37] KUMAR, A. and KANNAN, R. (2010). Clustering with spectral norm and the k-means algorithm. In 2010 *IEEE* 51*st Annual Symposium on Foundations of Computer Science—FOCS* 2010 299–308. IEEE Computer Soc., Los Alamitos, CA. MR3025203

[38] KUMAR, A., SABHARWAL, Y. and SEN, S. (2004). A simple linear time $(1+\epsilon)$-approximation algorithm for k-means clustering in any dimensions. In 45*th Annual IEEE Symposium on Foundations of Computer Science* 454–462.

[39] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 https://doi.org/10.1214/aos/1015957395

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

[41] LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28** 129–137. MR0651807 https://doi.org/10.1109/TIT.1982.1056489

[42] LÖFFLER, M., ZHANG, A. Y. and ZHOU, H. H. (2021). Supplement to "Optimality of spectral clustering in the Gaussian mixture model." https://doi.org/10.1214/20-AOS2044SUPP

[43] LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of Lloyd's algorithm and its variants. Preprint.

[44] MAHAJAN, M., NIMBHORKAR, P. and VARADARAJAN, K. (2009). The planar k-means problem is NP-hard. In *WALCOM—Algorithms and Computation*. *Lecture Notes in Computer Science* **5431** 274–285. Springer, Berlin. MR2540919 https://doi.org/10.1007/978-3-642-00202-1_24

[45] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science* (*Las Vegas*, *NV*, 2001) 529–537. IEEE Computer Soc., Los Alamitos, CA. MR1948742

[46] MEILA, M. and SHI, J. (2001). Learning segmentation by random walks. In *Advances in Neural Information Processing Systems* 873–879.

[47] MONTI, S., TAMAYO, P., MESIROV, J. and GOLUB, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* **52** 91–118.

[48] NDAOUD, M. (2019). Sharp optimal recovery in the two component Gaussian mixture model. Preprint.

[49] NG, A. Y., JORDAN, M. I. and WEISS, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems* 849–856.

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

[51] PAN, S. J., NI, X., SUN, J., YANG, Q. and CHEN, Z. (2010). Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th International Conference on World Wide Web* 751–760. ACM, New York.

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

[53] PENG, J. and WEI, Y. (2007). Approximating k-means-type clustering via semidefinite programming. *SIAM J. Optim.* **18** 186–205. MR2299680 https://doi.org/10.1137/050641983

[54] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems* 3120–3128.

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

[56] ROYER, M. (2017). Adaptive clustering through semidefinite programming. *Adv. Neural Inf. Process. Syst.* 1795–1803.

[57] SARKAR, P. and BICKEL, P. J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.* **43** 962–990. MR3346694 https://doi.org/10.1214/14-AOS1285

[58] SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. Departmental Papers (CIS) 107.

[59] SIMON, H. D. (1991). Partitioning of unstructured problems for parallel processing. *Comput. Syst. Eng.* **2** 135–148.

[60] SPIELMAN, D. A. and TENG, S.-H. (1996). Spectral partitioning works: Planar graphs and finite element meshes. In *37th Annual Symposium on Foundations of Computer Science* (*Burlington*, *VT*, 1996) 96–105. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1450607 https://doi.org/10.1109/SFCS.1996.548468

[61] SRIVASTAVA, P. R., PURNAMRITA, S. and HANASUSANTO, G. A. (2020). A robust spectral clustering algorithm for sub-Gaussian mixture models with outliers. Preprint.

[62] TIBSHIRANI, R., WALTHER, G. and HASTIE, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503 https://doi.org/10.1111/1467-9868.00293

[63] VAN DRIESSCHE, R. and ROOSE, D. (1995). An improved spectral bisection algorithm and its application to dynamic load balancing. *Parallel Comput.* **21** 29–48. MR1314376 https://doi.org/10.1016/0167-8191(94)00059-J

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

[65] VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416. MR2409803 https://doi.org/10.1007/s11222-007-9033-z

[66] VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. MR2396807 https://doi.org/10.1214/009053607000000640

[67] VU, V. (2018). A simple SVD algorithm for finding hidden partitions. *Combin. Probab. Comput.* **27** 124–140. MR3734334 https://doi.org/10.1017/S0963548317000463

[68] WANG, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika* **97** 893–904. MR2746159 https://doi.org/10.1093/biomet/asq061

[69] YU, S. and SHI, J. (2003). Multiclass spectral clustering. In *Proceedings Ninth IEEE International Conference on Computer Vision* 313–319.

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

[71] ZHOU, Z. and AMINI, A. A. (2019). Analysis of spectral clustering algorithms for community detection: The general bipartite setting. *J. Mach. Learn. Res.* **20** Paper No. 47, 47. MR3948087

The Annals of Statistics
2021, Vol. 49, No. 5, 2531–2551
https://doi.org/10.1214/20-AOS2045

ADAPTIVE ESTIMATION OF MULTIVARIATE PIECEWISE POLYNOMIALS AND BOUNDED VARIATION FUNCTIONS BY OPTIMAL DECISION TREES

BY SABYASACHI CHATTERJEE[1,??] AND SUBHAJIT GOSWAMI[2,??]

[1]*Department of Statistics, University of Illinois at Urbana–Champaign,* [??]*sc1706@illinois.edu*
[2]*Tata Institute of Fundamental Research,* [??]*goswami@math.tifr.res.in*

Proposed by Donoho (*Ann. Statist.* **25** (1997) 1870–1911), Dyadic CART is a nonparametric regression method which computes a globally optimal dyadic decision tree and fits piecewise constant functions in two dimensions. In this article, we define and study Dyadic CART and a closely related estimator, namely Optimal Regression Tree (ORT), in the context of estimating piecewise smooth functions in general dimensions in the fixed design setup. More precisely, these optimal decision tree estimators fit piecewise polynomials of any given degree. Like Dyadic CART in two dimensions, we reason that these estimators can also be computed in polynomial time in the sample size N via dynamic programming. We prove oracle inequalities for the finite sample risk of Dyadic CART and ORT, which imply tight risk bounds for several function classes of interest. First, they imply that the finite sample risk of ORT of *order* $r \geq 0$ is always bounded by $Ck\frac{\log N}{N}$ whenever the regression function is piecewise polynomial of degree r on some reasonably regular axis aligned rectangular partition of the domain with at most k rectangles. Beyond the univariate case, such guarantees are scarcely available in the literature for computationally efficient estimators. Second, our oracle inequalities uncover minimax rate optimality and adaptivity of the Dyadic CART estimator for function spaces with bounded variation. We consider two function spaces of recent interest where multivariate total variation denoising and univariate trend filtering are the state of the art methods. We show that Dyadic CART enjoys certain advantages over these estimators while still maintaining all their known guarantees.

REFERENCES

BERMAN, P., DASGUPTA, B. and MUTHUKRISHNAN, S. (2002). Exact size of binary space partitionings and improved rectangle tiling algorithms. *SIAM J. Discrete Math.* **15** 252–267. MR1920185 https://doi.org/10.1137/S0895480101384347

BERTIN, K. and LECUÉ, G. (2008). Selection of variables and dimension reduction in high-dimensional nonparametric regression. *Electron. J. Stat.* **2** 1224–1241. MR2461900 https://doi.org/10.1214/08-EJS327

BERTSIMAS, D. and DUNN, J. (2017). Optimal classification trees. *Mach. Learn.* **106** 1039–1082. MR3665788 https://doi.org/10.1007/s10994-017-5633-9

BLANCHARD, G., SCHÄFER, C., ROZENHOLC, Y. and MÜLLER, K.-R. (2007). Optimal dyadic decision trees. *Mach. Learn.* **66** 209–241.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees. Wadsworth Statistics/Probability Series*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

CHATTERJEE, S. and GOSWAMI, S. (2019). New risk bounds for 2D total variation denoising. Preprint. Available at arXiv:1902.01215.

CHATTERJEE, S. and GOSWAMI, S. (2021). Supplement to "Adaptive estimation of multivariate piecewise polynomials and bounded variation functions by optimal decision trees." https://doi.org/10.1214/20-AOS2045SUPP

MSC2020 subject classifications. 62G05.

Key words and phrases. Optimal decision trees, Dyadic CART, piecewise polynomial fitting, oracle risk bounds, bounded variation function estimation.

CHAUDHURI, P., HUANG, M.-C., LOH, W.-Y. and YAO, R. (1994). Piecewise-polynomial regression trees. *Statist*. *Sinica* **4** 143–167.

DALALYAN, A. S., HEBIRI, M. and LEDERER, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. MR3556784 https://doi.org/10.3150/15-BEJ756

DE BERG, M. (1995). Linear size binary space partitions for fat objects. In *Algorithms—ESA'95* (*Corfu*). *Lecture Notes in Computer Science* **979** 252–263. Springer, Berlin. MR1460747 https://doi.org/10.1007/3-540-60313-1_148

DENG, H. and ZHANG, C.-H. (2020). Isotonic regression in multi-dimensional spaces and graphs. *Ann*. *Statist*. **48** 3672–3698. MR4185824 https://doi.org/10.1214/20-AOS1947

DONOHO, D. L. (1997). CART and best-ortho-basis: A connection. *Ann*. *Statist*. **25** 1870–1911. MR1474073 https://doi.org/10.1214/aos/1069362377

DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann*. *Statist*. **26** 879–921. MR1635414 https://doi.org/10.1214/aos/1024691081

GEY, S. and NEDELEC, E. (2005). Model selection for CART regression trees. *IEEE Trans*. *Inf. Theory* **51** 658–670. MR2236074 https://doi.org/10.1109/TIT.2004.840903

GUNTUBOYINA, A., LIEU, D., CHATTERJEE, S. and SEN, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann*. *Statist*. **48** 205–229. MR4065159 https://doi.org/10.1214/18-AOS1799

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*. *Springer Series in Statistics*. Springer, New York.

HERSHBERGER, J., SURI, S. and TÓTH, C. D. (2005). Binary space partitions of orthogonal subdivisions. *SIAM J*. *Comput*. **34** 1380–1397. MR2165746 https://doi.org/10.1137/S0097539704445706

HÜTTER, J.-C. and RIGOLLET, P. (2016). Optimal rates for total variation denoising. In *Conference on Learning Theory* 1115–1146.

HYAFIL, L. and RIVEST, R. L. (1976). Constructing optimal binary decision trees is NP-complete. *Inform*. *Process*. *Lett*. **5** 15–17. MR0413598 https://doi.org/10.1016/0020-0190(76)90095-8

ISHWARAN, H. (2015). The effect of splitting on random forests. *Mach*. *Learn*. **99** 75–118. MR3324236 https://doi.org/10.1007/s10994-014-5451-2

KIM, S.-J., KOH, K., BOYD, S. and GORINEVSKY, D. (2009). l_1 trend filtering. *SIAM Rev.* **51** 339–360. MR2505584 https://doi.org/10.1137/070690274

LECUÉ, G. (2008). Classification with minimax fast rates for classes of Bayes rules with sparse representation. *Electron*. *J*. *Stat*. **2** 741–773. MR2430253 https://doi.org/10.1214/07-EJS015

LOH, W.-Y. (2014). Fifty years of classification and regression trees. *Int*. *Stat*. *Rev*. **82** 329–348. MR3280974 https://doi.org/10.1111/insr.12016

MAMMEN, E. and VAN DE GEER, S. (1997). Locally adaptive regression splines. *Ann*. *Statist*. **25** 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635

NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math*. **1738** 85–277. Springer, Berlin. MR1775640

NOWAK, R., MITRA, U. and WILLETT, R. (2004). Estimating inhomogeneous fields using wireless sensor networks. *IEEE J*. *Sel*. *Areas Commun*. **22** 999–1006.

ORTELLI, F. and VAN DE GEER, S. (2018). On the total variation regularized estimator over a class of tree graphs. *Electron*. *J*. *Stat*. **12** 4517–4570. MR3892703 https://doi.org/10.1214/18-ejs1519

RUDIN, L. I., OSHER, S. and FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms **60** 259–268. MR3363401 https://doi.org/10.1016/0167-2789(92)90242-F

SADHANALA, V., WANG, Y.-X. and TIBSHIRANI, R. J. (2016). Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In *Advances in Neural Information Processing Systems* 3513–3521.

SCORNET, E., BIAU, G. and VERT, J.-P. (2015). Consistency of random forests. *Ann*. *Statist*. **43** 1716–1741. MR3357876 https://doi.org/10.1214/15-AOS1321

SCOTT, C. and NOWAK, R. D. (2006). Minimax-optimal classification with dyadic decision trees. *IEEE Trans*. *Inf. Theory* **52** 1335–1353. MR2241192 https://doi.org/10.1109/TIT.2006.871056

STEIDL, G., DIDAS, S. and NEUMANN, J. (2006). Splines in higher order TV regularization. *Int*. *J*. *Comput*. *Vis*. **70** 241–255.

TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann*. *Statist*. **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

TIBSHIRANI, R. (2015). Nonparametric regression (and classification).

TÓTH, C. D. (2005). Binary space partitions: Recent developments. In *Combinatorial and Computational Geometry*. *Math*. *Sci*. *Res*. *Inst*. *Publ*. **52** 529–556. Cambridge Univ. Press, Cambridge. MR2178338

VAN DE GEER, S. and ORTELLI, F. (2019). Prediction bounds for (higher order) total variation regularized least squares. Preprint. Available at arXiv:1904.10871.

WAGER, S. and WALTHER, G. (2015). Adaptive concentration of regression trees, with application to random forests. Preprint. Available at arXiv:1503.06388.

WILLETT, R. M. and NOWAK, R. D. (2007). Multiscale Poisson intensity and density estimation. *IEEE Trans. Inf. Theory* **53** 3171–3187. MR2417680 https://doi.org/10.1109/TIT.2007.903139

The Annals of Statistics
2021, Vol. 49, No. 5, 2552–2576
https://doi.org/10.1214/20-AOS2046

RANK-BASED ESTIMATION UNDER ASYMPTOTIC DEPENDENCE AND INDEPENDENCE, WITH APPLICATIONS TO SPATIAL EXTREMES

BY MICHAËL LALANCETTE[1,*], SEBASTIAN ENGELKE[2,??] AND STANISLAV VOLGUSHEV[1,†]

[1]*Department of Statistical Sciences, University of Toronto,* [*]*lalancette@utstat.toronto.edu,* [†]*stanislav.volgushev@utoronto.ca*
[2]*Research Center for Statistics, University of Geneva,* [??]*sebastian.engelke@unige.ch*

Multivariate extreme value theory is concerned with modeling the joint tail behavior of several random variables. Existing work mostly focuses on asymptotic dependence, where the probability of observing a large value in one of the variables is of the same order as observing a large value in all variables simultaneously. However, there is growing evidence that asymptotic independence is equally important in real world applications. Available statistical methodology in the latter setting is scarce and not well understood theoretically. We revisit nonparametric estimation and introduce rank-based M-estimators for parametric models that simultaneously work under asymptotic dependence and asymptotic independence, without requiring prior knowledge on which of the two regimes applies. Asymptotic normality of the proposed estimators is established under weak regularity conditions. We further show how bivariate estimators can be leveraged to obtain parametric estimators in spatial tail models, and again provide a thorough theoretical justification for our approach.

REFERENCES

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. MR3456363 https://doi.org/10.1214/15-AOAS863

BROWN, B. M. and RESNICK, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probab.* **14** 732–739. MR0517438 https://doi.org/10.2307/3213346

BÜCHER, A. and DETTE, H. (2013). Multiplier bootstrap of tail copulas with applications. *Bernoulli* **19** 1655–1687. MR3129029 https://doi.org/10.3150/12-BEJ425

BÜCHER, A., SEGERS, J. and VOLGUSHEV, S. (2014). When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *Ann. Statist.* **42** 1598–1634. MR3262462 https://doi.org/10.1214/14-AOS1237

BÜCHER, A., VOLGUSHEV, S. and ZOU, N. (2019). On second order conditions in the multivariate block maxima and peak over threshold method. *J. Multivariate Anal.* **173** 604–619. MR3959186 https://doi.org/10.1016/j.jmva.2019.04.011

COLES, S., HEFFERNAN, J. and TAWN, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.

CSÖRGŐ, M. and HORVÁTH, L. (1987). Approximation of intermediate quantile processes. *J. Multivariate Anal.* **21** 250–262. MR0884099 https://doi.org/10.1016/0047-259X(87)90004-2

DAVISON, A. C. and GHOLAMREZAEE, M. M. (2012). Geostatistics of extremes. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **468** 581–608. MR2874052 https://doi.org/10.1098/rspa.2011.0412

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980 https://doi.org/10.1214/11-STS376

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DE HAAN, L. and LIN, T. (2001). On convergence toward an extreme value distribution in $C[0, 1]$. *Ann. Probab.* **29** 467–483. MR1825160 https://doi.org/10.1214/aop/1008956340

MSC2020 subject classifications. Primary 62G32; secondary 62F12, 62G20, 62G30.
Key words and phrases. Multivariate extremes, asymptotic independence, inverted max-stable distribution, spatial process, M-estimation.

DE HAAN, L., NEVES, C. and PENG, L. (2008). Parametric tail copula estimation and model testing. *J. Multivariate Anal.* **99** 1260–1275. MR2419346 https://doi.org/10.1016/j.jmva.2007.08.003

DOMBRY, C., ENGELKE, S. and OESTING, M. (2017). Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. Available from https://arxiv.org/abs/1612.05178.

DRAISMA, G., DREES, H., FERREIRA, A. and DE HAAN, L. (2004). Bivariate tail estimation: Dependence in asymptotic independence. *Bernoulli* **10** 251–280. MR2046774 https://doi.org/10.3150/bj/1082380219

DREES, H. and HUANG, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64** 25–47. MR1619974 https://doi.org/10.1006/jmva.1997.1708

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2008). A method of moments estimator of tail dependence. *Bernoulli* **14** 1003–1026. MR2543584 https://doi.org/10.3150/08-BEJ130

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. MR3015043 https://doi.org/10.1214/12-AOS1023

EINMAHL, J. H. J. and SEGERS, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.* **37** 2953–2989. MR2541452 https://doi.org/10.1214/08-AOS677

EINMAHL, J. H. J., KIRILIOUK, A., KRAJINA, A. and SEGERS, J. (2016). An M-estimator of spatial tail dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 275–298. MR3453656 https://doi.org/10.1111/rssb.12114

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. MR4136498

ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8**. https://doi.org/10.1146/annurev-statistics-040620-041554

ENGELKE, S., OPITZ, T. and WADSWORTH, J. (2019). Extremal dependence of random scale constructions. *Extremes* **22** 623–666. MR4031852 https://doi.org/10.1007/s10687-019-00353-3

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 239–265. MR3299407 https://doi.org/10.1111/rssb.12074

GUILLOTTE, S., PERRON, F. and SEGERS, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 377–406. MR2815781 https://doi.org/10.1111/j.1467-9868.2010.00770.x

GUMBEL, É. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Stat. Univ. Paris* **9** 171–173. MR0119279

HEFFERNAN, J. E. and RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.* **17** 537–571. MR2308335 https://doi.org/10.1214/105051606000000835

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 497–546. With discussions and reply by the authors. MR2088289 https://doi.org/10.1111/j.1467-9868.2004.02050.x

HUANG, X. (1992). Statistics of bivariate extreme value theory. Ph.D. thesis, Erasmus Univ. Rotterdam.

HUSER, R., OPITZ, T. and THIBAUD, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spat. Stat.* **21** 166–186. MR3692183 https://doi.org/10.1016/j.spasta.2017.06.004

HUSER, R. and WADSWORTH, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *J. Amer. Statist. Assoc.* **114** 434–444. MR3941266 https://doi.org/10.1080/01621459.2017.1411813

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699 https://doi.org/10.1016/0167-7152(89)90106-5

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440 https://doi.org/10.1214/09-AOP455

KEEF, C., TAWN, J. and SVENSSON, C. (2009). Spatial risk assessment for extreme river flows. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 601–618. MR2750258 https://doi.org/10.1111/j.1467-9876.2009.00672.x

LALANCETTE, M., ENGELKE, S. and VOLGUSHEV, S. (2021). Supplement to "Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes." https://doi.org/10.1214/20-AOS2046SUPP

LE, P. D., DAVISON, A. C., ENGELKE, S., LEONARD, M. and WESTRA, S. (2018). Dependence properties of spatial rainfall extremes and areal reduction factors. *J. Hydrol.* **565** 711–719.

LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. MR1399163 https://doi.org/10.1093/biomet/83.1.169

LEDFORD, A. W. and TAWN, J. A. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B* **59** 475–499. MR1440592 https://doi.org/10.1111/1467-9868.00080

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202 https://doi.org/10.1198/jasa.2009.tm08577

PENG, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Probab. Lett.* **43** 399–409. MR1707950 https://doi.org/10.1016/S0167-7152(98)00280-6

PENG, L. and QI, Y. (2008). Bootstrap approximation of tail dependence function. *J. Multivariate Anal.* **99** 1807–1824. MR2444820 https://doi.org/10.1016/j.jmva.2008.01.018

POON, S.-H., ROCKINGER, M. and TAWN, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Rev. Financ. Stud.* **17** 581–610.

RAMOS, A. and LEDFORD, A. (2009). A new class of models for bivariate joint tails. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 219–241. MR2655531 https://doi.org/10.1111/j.1467-9868.2008.00684.x

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes. Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

RESNICK, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5** 303–336. MR2002121 https://doi.org/10.1023/A:1025148622954

ROCKAFELLAR, R. T. (1970). *Convex Analysis. Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, N.J. MR0274683

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786 https://doi.org/10.1023/A:1020977924878

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.

TAWN, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75** 397–415. MR0967580 https://doi.org/10.1093/biomet/75.3.397

WADSWORTH, J. L. and TAWN, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika* **99** 253–272. MR2931252 https://doi.org/10.1093/biomet/asr080

WADSWORTH, J. L., TAWN, J. A., DAVISON, A. C. and ELTON, D. M. (2017). Modelling across extremal dependence classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 149–175. MR3597968 https://doi.org/10.1111/rssb.12157

WELLER, G. B. and COOLEY, D. (2014). A sum characterization of hidden regular variation with likelihood inference via expectation–maximization. *Biometrika* **101** 17–36. MR3180655 https://doi.org/10.1093/biomet/ast046

WESTRA, S. and SISSON, S. A. (2011). Detection of non-stationarity in precipitation extremes using a max-stable process model. *J. Hydrol.* **406** 119–128.

ZHOU, C. (2010). Dependence structure of risk factors and diversification effects. *Insurance Math. Econom.* **46** 531–540. MR2642530 https://doi.org/10.1016/j.insmatheco.2010.01.010

ZSCHEISCHLER, J. and SENEVIRATNE, S. I. (2017). Dependence of drivers affects risks associated with compound events. *Sci. Adv.* **3** e1700263. https://doi.org/10.1126/sciadv.1700263

The Annals of Statistics
2021, Vol. 49, No. 5, 2577–2610
https://doi.org/10.1214/20-AOS2047

ESTIMATION OF SMOOTH FUNCTIONALS IN NORMAL MODELS: BIAS REDUCTION AND ASYMPTOTIC EFFICIENCY

BY VLADIMIR KOLTCHINSKII* AND MAYYA ZHILOVA†

School of Mathematics, Georgia Institute of Technology, *vlad@math.gatech.edu, †mzhilova@math.gatech.edu

Let $X_1, \dots, X_n$ be i.i.d. random variables sampled from a normal distribution $N(\mu, \Sigma)$ in $\mathbb{R}^d$ with unknown parameter $\theta = (\mu, \Sigma) \in \Theta := \mathbb{R}^d \times \mathcal{C}_+^d$, where $\mathcal{C}_+^d$ is the cone of positively definite covariance operators in $\mathbb{R}^d$. Given a smooth functional $f : \Theta \mapsto \mathbb{R}^1$, the goal is to estimate $f(\theta)$ based on $X_1, \dots, X_n$. Let

$$\Theta(a; d) := \mathbb{R}^d \times \{\Sigma \in \mathcal{C}_+^d : \sigma(\Sigma) \subset [1/a, a]\}, \quad a \geq 1,$$

where $\sigma(\Sigma)$ is the spectrum of covariance Σ. Let $\hat{\theta} := (\hat{\mu}, \hat{\Sigma})$, where $\hat{\mu}$ is the sample mean and $\hat{\Sigma}$ is the sample covariance, based on the observations $X_1, \dots, X_n$. For an arbitrary functional $f \in C^s(\Theta)$, $s = k + 1 + \rho$, $k \geq 0$, $\rho \in (0, 1]$, we define a functional $f_k : \Theta \mapsto \mathbb{R}$ such that

$$\sup_{\theta \in \Theta(a;d)} \|f_k(\hat{\theta}) - f(\theta)\|_{L_2(\mathbb{P}_\theta)} \lesssim_{s,\beta} \|f\|_{C^s(\Theta)} \left[\left(\frac{a}{\sqrt{n}} \vee a^{\beta s} \left(\sqrt{\frac{d}{n}} \right)^s \right) \wedge 1 \right],$$

where $\beta = 1$ for $k = 0$ and $\beta > s - 1$ is arbitrary for $k \geq 1$. This error rate is minimax optimal and similar bounds hold for more general loss functions. If $d = d_n \leq n^\alpha$ for some $\alpha \in (0, 1)$ and $s \geq \frac{1}{1-\alpha}$, the rate becomes $O(n^{-1/2})$. Moreover, for $s > \frac{1}{1-\alpha}$, the estimator $f_k(\hat{\theta})$ is shown to be asymptotically efficient. The crucial part of the construction of estimator $f_k(\hat{\theta})$ is a bias reduction method studied in the paper for more general statistical models than normal.

REFERENCES

[1] ADAMCZAK, R. and WOLFF, P. (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** 531–586. MR3383337 https://doi.org/10.1007/s00440-014-0579-3

[2] ALEKSANDROV, A. B. and PELLER, V. V. (2016). Operator Lipschitz functions. *Uspekhi Mat. Nauk* **71** 3–106. MR3588921 https://doi.org/10.4213/rm9729

[3] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

[4] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

[5] BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29. MR1331653 https://doi.org/10.1214/aos/1176324452

[6] CAI, T. T. and LOW, M. G. (2005). On adaptive estimation of linear functionals. *Ann. Statist.* **33** 2311–2343. MR2211088 https://doi.org/10.1214/009053605000000633

[7] CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. MR2253108 https://doi.org/10.1214/009053605000000147

[8] COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.* **45** 923–958. MR3662444 https://doi.org/10.1214/15-AOS1432

MSC2020 subject classifications. Primary 62H12; secondary 62G20, 62H25, 60B20.
Key words and phrases. Efficiency, smooth functionals, bias reduction, bootstrap chain, random homotopy, concentration.

[9] DONOHO, D. and LIU, R. (1987). On minimax estimation of linear functionals Technical Report N 105, Department of Statistics, UC Berkeley.

[10] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667, 668–701. MR1105839 https://doi.org/10.1214/aos/1176348114

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

[12] FAÀ DI BRUNO, F. (1855). Sullo sviluppo delle funzioni. *Annali di Scienze Matematiche e Fisiche* **6** 479–480.

[13] GILL, R. D. and LEVIT, B. Y. (1995). Applications of the Van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli* **1** 59–79. MR1354456 https://doi.org/10.2307/3318681

[14] GIRKO, V. L. (1987). An introduction to general statistical analysis. *Theory Probab. Appl.* **32** 229–242.

[15] GIRKO, V. L. (1995). *Statistical Analysis of Observations of Increasing Dimension*. *Theory and Decision Library. Series B*: *Mathematical and Statistical Methods* **28**. Kluwer Academic, Dordrecht. MR1473719 https://doi.org/10.1007/978-94-015-8567-5

[16] HAN, Y., JIAO, J. and MUKHERJEE, R. (2020). On estimation of L_r-norms in Gaussian white noise models. *Probab. Theory Related Fields* **177** 1243–1294. MR4126939 https://doi.org/10.1007/s00440-020-00982-x

[17] HARDY, M. (2006). Combinatorics of partial derivatives. *Electron. J. Combin.* **13** Research Paper 1. MR2200529

[18] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. Springer, New York. MR0620321

[19] IBRAGIMOV, I. A., NEMIROVSKIĬ, A. S. and KHAS'MINSKIĬ, R. Z. (1987). Some problems of nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.

[20] JIAO, J. and HAN, Y. (2020). Bias correction with jackknife, bootstrap, and Taylor series. *IEEE Trans. Inf. Theory* **66** 4392–4418. MR4130622 https://doi.org/10.1109/TIT.2020.2969439

[21] JOST, J., KELL, M. and RODRIGUES, C. S. (2015). Representation of Markov chains by random maps: Existence and regularity conditions. *Calc. Var. Partial Differential Equations* **54** 2637–2655. MR3412387 https://doi.org/10.1007/s00526-015-0878-2

[22] KLEMELÄ, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** 539–564. MR2214904 https://doi.org/10.1007/s00440-005-0447-2

[23] KOLTCHINSKII, V. (2020). Estimation of smooth functionals in high-dimensional models: bootstrap chains and Gaussian approximation. Available at arXiv:2011.03789.

[24] KOLTCHINSKII, V. (2018). Asymptotic efficiency in high-dimensional covariance estimation. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro* 2018. *Vol. IV. Invited Lectures* 2921–2943. World Sci. Publ., Hackensack, NJ. MR3966516

[25] KOLTCHINSKII, V. (2021). Asymptotically efficient estimation of smooth functionals of covariance operators. *J. Eur. Math. Soc.* (*JEMS*). **23** 765–843. MR4210724 https://doi.org/10.4171/JEMS/1023

[26] KOLTCHINSKII, V., LÖFFLER, M. and NICKL, R. (2020). Efficient estimation of linear functionals of principal components. *Ann. Statist.* **48** 464–490. MR4065170 https://doi.org/10.1214/19-AOS1816

[27] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1976–2013. MR3573302 https://doi.org/10.1214/15-AIHP705

[28] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

[29] KOLTCHINSKII, V. and LOUNICI, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Statist.* **45** 121–157. MR3611488 https://doi.org/10.1214/16-AOS1437

[30] KOLTCHINSKII, V. and XIA, D. (2016). Perturbation of linear forms of singular vectors under Gaussian noise. In *High Dimensional Probability VII*. *Progress in Probability* **71** 397–423. Springer, Cham. MR3565274 https://doi.org/10.1007/978-3-319-40519-3_18

[31] KOLTCHINSKII, V. and ZHILOVA, M. (2021). Efficient estimation of smooth functionals in Gaussian shift models. *Ann. Inst. Henri Poincaré Probab. Stat.* To appear. Available at arXiv:1810.02767.

[32] KOLTCHINSKII, V. and ZHILOVA, M. (2021). Estimation of smooth functionals of location parameter in Gaussian and Poincaré random shift models. *Sankhyā*. To appear.

[33] KOLTCHINSKII, V. and ZHILOVA, M. (2021). Supplement to "Estimation of smooth functionals in normal models: Bias reduction and asymptotic efficiency." https://doi.org/10.1214/20-AOS2047SUPP.

[34] LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981 https://doi.org/10.1214/aos/1032894458

[35] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867 https://doi.org/10.1007/s004409970006

[36] LEVIT, B. Y. (1975). The efficiency of a certain class of nonparametric estimators. *Theory Probab. Appl.* **20** 723–740.

[37] LEVIT, B. Y. (1978). Asymptotically efficient estimation of nonlinear functionals. *Problemy Peredachi Informatsii* **14** 65–72. MR0533450

[38] MUKHERJEE, R., NEWEY, W. and ROBINS, J. (2017). Semiparametric efficient empirical higher order influence function estimators. Available at arXiv:1705.07577.

[39] NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math*. **1738** 85–277. Springer, Berlin. MR1775640

[40] NEMIROVSKIĬ, A. S. (1990). Necessary conditions for efficient estimation of functionals of a nonparametric signal observed in white noise. *Theory Probab. Appl*. **35** 94–103.

[41] PISIER, G. (1986). Probabilistic methods in the geometry of Banach spaces. In *Probability and Analysis* (*Varenna*, 1985). *Lecture Notes in Math*. **1206** 167–241. Springer, Berlin. MR0864714 https://doi.org/10.1007/BFb0076302

[42] ROBINS, J., LI, L., TCHETGEN, E. and VAN DER VAART, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics*: *Essays in Honor of David A*. *Freedman*. *Inst*. *Math*. *Stat*. (*IMS*) *Collect*. **2** 335–421. IMS, Beachwood, OH. MR2459958 https://doi.org/10.1214/193940307000000527

[43] ROBINS, J. M., LI, L., TCHETGEN, E. T. and VAN DER VAART, A. (2016). Asymptotic normality of quadratic estimators. *Stochastic Process*. *Appl*. **126** 3733–3759. MR3565475 https://doi.org/10.1016/j.spa.2016.04.005

[44] TRIEBEL, H. (1983). *Theory of Function Spaces*. *Monographs in Mathematics* **78**. Birkhäuser, Basel. MR0781540 https://doi.org/10.1007/978-3-0346-0416-1

[45] VAN DER VAART, A. (2014). Higher order tangent spaces and influence functions. *Statist*. *Sci*. **29** 679–686. MR3300365 https://doi.org/10.1214/14-STS478

[46] VILLANI, C. (2009). *Optimal Transport*. *Old and New*. Springer, Berlin. MR2459454

The Annals of Statistics
2021, Vol. 49, No. 5, 2611–2641
https://doi.org/10.1214/21-AOS2048

ADDITIVE REGRESSION FOR NON-EUCLIDEAN RESPONSES AND PREDICTORS

BY JEONG MIN JEON[1,*], BYEONG U. PARK[2,??] AND INGRID VAN KEILEGOM[1,†]

[1]*ORSTAT, KU Leuven,* [*]*jeongmin.jeon@kuleuven.be,* [†]*ingrid.vankeilegom@kuleuven.be*
[2]*Department of Statistics, Seoul National University,* [??]*bupark@stats.snu.ac.kr*

Additive regression is studied in a very general setting where both the response and predictors are allowed to be non-Euclidean. The response takes values in a general separable Hilbert space, whereas the predictors take values in general semimetric spaces, which covers a very wide range of nonstandard response variables and predictors. A general framework of estimating additive models is presented for semimetric space-valued predictors. In particular, full details of implementation and the corresponding theory are given for predictors taking values in Hilbert spaces and/or Riemannian manifolds. The existence of the estimators, convergence of a backfitting algorithm, rates of convergence and asymptotic distributions of the estimators are discussed. The finite sample performance of the estimators is investigated by means of two simulation studies. Finally, three data sets covering several types of non-Euclidean data are analyzed to illustrate the usefulness of the proposed general approach.

REFERENCES

[1] CHACÓN, J. E., MATEU-FIGUERAS, G. and MARTÍN-FERNÁNDEZ LAUDER, J. A. (2014). Gaussian kernels for density estimation with compositional data. *Comput*. *Geosci*. **37** 702–711.
[2] CHAKRABORTY, R. and VEMURI, B. C. (2019). Statistics on the Stiefel manifold: Theory and applications. *Ann*. *Statist*. **47** 415–438. MR3909937 https://doi.org/10.1214/18-AOS1692
[3] CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J*. *Amer*. *Statist*. *Assoc*. **108** 1421–1434. MR3174718 https://doi.org/10.1080/01621459.2013.827984
[4] CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169
[5] 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
[6] DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2014). Nonparametric regression for spherical data. *J*. *Amer*. *Statist*. *Assoc*. **109** 748–763. MR3223747 https://doi.org/10.1080/01621459.2013.866567
[7] DI MARZIO, M., PANZERA, A. and VENIERI, C. (2015). Non-parametric regression for compositional data. *Stat*. *Model*. **15** 113–133. MR3325749 https://doi.org/10.1177/1471082X14535522
[8] DRYDEN, I. L. and MARDIA, K. V. (2016). *Statistical Shape Analysis with Applications in R*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3559734 https://doi.org/10.1002/9781119072492
[9] EGOZCUE, J. J., PAWLOWSKY-GLAHN, V., MATEU-FIGUERAS, G. and BARCELÓ-VIDAL, C. (2003). Isometric logratio transformations for compositional data analysis. *Math*. *Geol*. **35** 279–300. MR1986165 https://doi.org/10.1023/A:1023818214614
[10] FAN, Y., JAMES, G. M. and RADCHENKO, P. (2015). Functional additive regression. *Ann*. *Statist*. **43** 2296–2325. MR3396986 https://doi.org/10.1214/15-AOS1346
[11] FEBRERO-BANDE, M. and GONZÁLEZ-MANTEIGA, W. (2013). Generalized additive models for functional data. *TEST* **22** 278–292. MR3062258 https://doi.org/10.1007/s11749-012-0308-0
[12] GONZALEZ-MANTEIGA, W., HENRY, G. and RODRIGUEZ, D. (2012). Partly linear models on Riemannian manifolds. *J*. *Appl*. *Stat*. **39** 1797–1809. MR2935559 https://doi.org/10.1080/02664763.2012.683169

MSC2020 subject classifications. Primary 62G08; secondary 62G20.
Key words and phrases. Additive models, smooth backfitting, Hilbert spaces, Riemannian manifolds, non-Euclidean data, functional data, density-valued data, compositional data, directional data, shape data.

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

[14] HAN, K., MÜLLER, H.-G. and PARK, B. U. (2020). Additive functional regression for densities as responses. *J. Amer. Statist. Assoc.* **115** 997–1010. MR4107695 https://doi.org/10.1080/01621459.2019.1604365

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

[16] HENRY, G. and RODRIGUEZ, D. (2009). Kernel density estimation on Riemannian manifolds: Asymptotic results. *J. Math. Imaging Vision* **34** 235–239. MR2515446 https://doi.org/10.1007/s10851-009-0145-2

[17] 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

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

[19] JEON, J. M., PARK, B. U. and VAN KEILEGOM, I. (2021). Supplement to "Additive regression for non-Euclidean responses and predictors." https://doi.org/10.1214/21-AOS2048SUPP

[20] KENT, J. T. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **56** 285–299. MR1281934

[21] KIM, J. S., STAICU, A.-M., MAITY, A., CARROLL, R. J. and RUPPERT, D. (2018). Additive function-on-function regression. *J. Comput. Graph. Statist.* **27** 234–244. MR3788315 https://doi.org/10.1080/10618600.2017.1356730

[22] LANG, R. (1986). A note on the measurability of convex sets. *Arch. Math.* (*Basel*) **47** 90–92. MR0855142 https://doi.org/10.1007/BF01202504

[23] LE, H. L. and KENDALL, D. G. (1993). The Riemannian structure of Euclidean shape spaces: A novel environment for statistics. *Ann. Statist.* **21** 1225–1271. MR1241266 https://doi.org/10.1214/aos/1176349259

[24] LEE, J. M. (2018). *Introduction to Riemannian Manifolds*. *Graduate Texts in Mathematics* **176**. Springer, Cham. MR3887684

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

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

[27] LIN, Z. and YAO, F. (2019). Intrinsic Riemannian functional data analysis. *Ann. Statist.* **47** 3533–3577. MR4025751 https://doi.org/10.1214/18-AOS1787

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

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

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

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

[32] OPSOMER, J. D. and RUPPERT, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25** 186–211. MR1429922 https://doi.org/10.1214/aos/1034276626

[33] PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Probab. Lett.* **73** 297–304. MR2179289 https://doi.org/10.1016/j.spl.2005.04.004

[34] PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. MR2214065 https://doi.org/10.1080/10485250500504828

[35] PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

[36] REIMHERR, M., SRIPERUMBUDUR, B. and TAOUFIK, B. (2018). Optimal prediction for additive function-on-function regression. *Electron. J. Stat.* **12** 4571–4601. MR3893421 https://doi.org/10.1214/18-EJS1505

[37] ROSENTHAL, M., WU, W., KLASSEN, E. and SRIVASTAVA, A. (2014). Spherical regression models using projective linear transformations. *J. Amer. Statist. Assoc.* **109** 1615–1624. MR3293614 https://doi.org/10.1080/01621459.2014.892881

[38] Sakai, T. (1996). *Riemannian Geometry*. *Translations of Mathematical Monographs* **149**. Amer. Math. Soc., Providence, RI. Translated from the 1992 Japanese original by the author. MR1390760 https://doi.org/10.1090/mmono/149

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

[40] Simó, A., Ibáñez, M. V., Epifanio, I. and Gimeno, V. (2020). Generalized partially linear models on Riemannian manifolds. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 641–661. MR4098966

[41] Tu, L. W. (2017). *Differential Geometry*. Springer, Berlin.

[42] Wood, S. N. (2017). *Generalized Additive Models*: *An Introduction with R*. Chapman and Hall/CRC.

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

[44] Yuan, Y., Zhu, H., Lin, W. and Marron, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x

The Annals of Statistics
2021, Vol. 49, No. 5, 2642–2671
https://doi.org/10.1214/21-AOS2049

SET STRUCTURED GLOBAL EMPIRICAL RISK MINIMIZERS ARE RATE OPTIMAL IN GENERAL DIMENSIONS

BY QIYANG HAN??

Department of Statistics, Rutgers University, ??*qh85@stat.rutgers.edu*

Entropy integrals are widely used as a powerful empirical process tool to obtain upper bounds for the rates of convergence of global empirical risk minimizers (ERMs), in standard settings such as density estimation and regression. The upper bound for the convergence rates thus obtained typically matches the minimax lower bound when the entropy integral converges, but admits a strict gap compared to the lower bound when it diverges. Birgé and Massart (*Probab. Theory Related Fields* **97** (1993) 113–150) provided a striking example showing that such a gap is real with the entropy structure alone: for a variant of the natural Hölder class with low regularity, the global ERM actually converges at the rate predicted by the entropy integral that substantially deviates from the lower bound. The counter-example has spawned a long-standing negative position on the use of global ERMs in the regime where the entropy integral diverges, as they are heuristically believed to converge at a suboptimal rate in a variety of models.

The present paper demonstrates that this gap can be closed if the models admit certain degree of "set structures" in addition to the entropy structure. In other words, the global ERMs in such set structured models will indeed be rate-optimal, matching the lower bound even when the entropy integral diverges. The models with set structures we investigate include (i) image and edge estimation, (ii) binary classification, (iii) multiple isotonic regression, (iv) s-concave density estimation, all in general dimensions when the entropy integral diverges. Here, set structures are interpreted broadly in the sense that the complexity of the underlying models can be essentially captured by the size of the empirical process over certain class of measurable sets, for which matching upper and lower bounds are obtained to facilitate the derivation of sharp convergence rates for the associated global ERMs.

REFERENCES

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

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

[3] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. MR0722129 https://doi.org/10.1007/BF00532480

[4] BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719 https://doi.org/10.1007/BF01199316

[5] BRUNEL, V.-E. (2013). Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron. J. Stat.* **7** 1301–1327. MR3063609 https://doi.org/10.1214/13-EJS804

[6] CARPENTER, T., DIAKONIKOLAS, I., SIDIROPOULOS, A. and STEWART, A. (2018). Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. In *Conference on Learning Theory* 1–29.

[7] CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.* **42** 2340–2381. MR3269982 https://doi.org/10.1214/14-AOS1254

[8] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788 https://doi.org/10.3150/16-BEJ865

MSC2020 subject classifications. Primary 60E15; secondary 62G05.

Key words and phrases. Empirical risk minimization, empirical process, classification, nonparametric regression, density estimation, non-Donsker.

[9] DAGAN, Y. and KUR, G. (2019). The log-concave maximum likelihood estimator is optimal in high dimensions. Preprint. Available at arXiv:1903.05315.

[10] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition*. *Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

[11] DOSS, C. R. and WELLNER, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann*. *Statist*. **44** 954–981. MR3485950 https://doi.org/10.1214/15-AOS1394

[12] DUDLEY, R. M. (1982). Empirical and Poisson processes on classes of sets or functions too large for central limit theorems. *Z*. *Wahrsch*. *Verw*. *Gebiete* **61** 355–368. MR0679680 https://doi.org/10.1007/BF00539835

[13] DUDLEY, R. M. (2014). *Uniform Central Limit Theorems*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **142**. Cambridge Univ. Press, New York. MR3445285

[14] GAO, F. and WELLNER, J. A. (2007). Entropy estimate for high-dimensional monotonic functions. *J*. *Multivariate Anal*. **98** 1751–1764. MR2392431 https://doi.org/10.1016/j.jmva.2006.09.003

[15] GINÉ, E. and KOLTCHINSKII, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann*. *Probab*. **34** 1143–1216. MR2243881 https://doi.org/10.1214/009117906000000070

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

[17] GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York. MR0599175

[18] GUNTUBOYINA, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann*. *Statist*. **40** 385–411. MR3014311 https://doi.org/10.1214/11-AOS959

[19] HAN, Q. (2021). Supplement to "Set structured global empirical risk minimizers are rate optimal in general dimensions." https://doi.org/10.1214/21-AOS2049SUPP.

[20] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2019). Isotonic regression in general dimensions. *Ann*. *Statist*. **47** 2440–2471. MR3988762 https://doi.org/10.1214/18-AOS1753

[21] HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann*. *Statist*. **44** 1332–1359. MR3485962 https://doi.org/10.1214/15-AOS1408

[22] HAN, Q. and WELLNER, J. A. (2019). Convergence rates of least squares regression estimators with heavy-tailed errors. *Ann*. *Statist*. **47** 2286–2319. MR3953452 https://doi.org/10.1214/18-AOS1748

[23] KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann*. *Statist*. **44** 2756–2779. MR3576560 https://doi.org/10.1214/16-AOS1480

[24] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann*. *Statist*. **34** 2593–2656. MR2329442 https://doi.org/10.1214/009053606000001019

[25] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1992). Asymptotically minimax image reconstruction problems. In *Topics in Nonparametric Estimation*. *Adv*. *Soviet Math*. **12** 45–86. Amer. Math. Soc., Providence, RI. MR1191691 https://doi.org/10.1137/1136015

[26] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics* **82**. Springer, New York. MR1226450 https://doi.org/10.1007/978-1-4612-2712-0

[27] KUR, G., GAO, F., GUNTUBOYINA, A. and SEN, B. (2020). Convex regression in multidimensions: Suboptimality of least squares estimators. Preprint. Available at arXiv:2006.02044.

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

[29] LECUÉ, G. (2007). Simultaneous adaptation to the margin and to complexity in classification. *Ann*. *Statist*. **35** 1698–1721. MR2351102 https://doi.org/10.1214/009053607000000055

[30] MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann*. *Statist*. **23** 502–524. MR1332579 https://doi.org/10.1214/aos/1176324533

[31] MAMMEN, E. and TSYBAKOV, A. B. (1999). Smooth discrimination analysis. *Ann*. *Statist*. **27** 1808–1829. MR1765618 https://doi.org/10.1214/aos/1017939240

[32] MASSART, P. and NÉDÉLEC, É. (2006). Risk bounds for statistical learning. *Ann*. *Statist*. **34** 2326–2366. MR2291502 https://doi.org/10.1214/009053606000000786

[33] OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann*. *Probab*. **15** 897–919. MR0893905

[34] POLLARD, D. (2002). Maximal inequalities via bracketing with adaptive truncation *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **38** 1039–1052. MR1955351 https://doi.org/10.1016/S0246-0203(02)01138-X

[35] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

[36] SAUMARD, A. (2012). Optimal upper and lower bounds for the true and empirical excess risks in heteroscedastic least-squares regression. *Electron. J. Stat.* **6** 579–655. MR2988421 https://doi.org/10.1214/12-EJS679

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

[38] TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002 https://doi.org/10.1214/aos/1079120131

[39] VAN DE GEER, S. (1987). A new approach to least-squares estimation, with applications. *Ann. Statist.* **15** 587–602. MR0888427 https://doi.org/10.1214/aos/1176350362

[40] VAN DE GEER, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907–924. MR1056343 https://doi.org/10.1214/aos/1176347632

[41] VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44. MR1212164 https://doi.org/10.1214/aos/1176349013

[42] VAN DE GEER, S. (1995). The method of sieves and minimum contrast estimators. *Math. Methods Statist.* **4** 20–38. MR1324688

[43] VAN DE GEER, S. and WAINWRIGHT, M. J. (2017). On concentration for (regularized) empirical risk minimization. *Sankhya A* **79** 159–200. MR3707417 https://doi.org/10.1007/s13171-017-0111-9

[44] VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

[45] VAN DER VAART, A. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24** 862–878. MR1394993 https://doi.org/10.1214/aos/1032894470

[46] VAN DER VAART, A. (1996). New Donsker classes. *Ann. Probab.* **24** 2128–2140. MR1415244 https://doi.org/10.1214/aop/1041903221

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

[48] WONG, W. H. and SHEN, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362. MR1332570 https://doi.org/10.1214/aos/1176324524

[49] YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500 https://doi.org/10.1214/aos/1017939142

The Annals of Statistics
2021, Vol. 49, No. 5, 2672–2696
https://doi.org/10.1214/21-AOS2050

EMPIRICAL TAIL COPULAS FOR FUNCTIONAL DATA

BY JOHN H. J. EINMAHL[1,??] AND JOHAN SEGERS[2,??]

[1]*Department of Econometrics and OR and CentER, Tilburg University,* [??]*j.h.j.einmahl@uvt.nl*
[2]*LIDAM/ISBA, UCLouvain,* [??]*johan.segers@uclouvain.be*

For multivariate distributions in the domain of attraction of a max-stable distribution, the tail copula and the stable tail dependence function are equivalent ways to capture the dependence in the upper tail. The empirical versions of these functions are rank-based estimators whose inflated estimation errors are known to converge weakly to a Gaussian process that is similar in structure to the weak limit of the empirical copula process. We extend this multivariate result to continuous functional data by establishing the asymptotic normality of the estimators of the tail copula, uniformly over all finite subsets of at most D points (D fixed). An application for testing tail copula stationarity is presented. The main tool for deriving the result is the uniform asymptotic normality of all the D-variate tail empirical processes. The proof of the main result is nonstandard.

REFERENCES

AZZALINI, A. and GENZ, A. (2020). The R package `mnormt`: The multivariate normal and t distributions (version 2.0.2).

BÜCHER, A. and DETTE, H. (2013). Multiplier bootstrap of tail copulas with applications. *Bernoulli* **19** 1655–1687. MR3129029 https://doi.org/10.3150/12-BEJ425

CHIAPINO, M., SABOURIN, A. and SEGERS, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes* **22** 193–222. MR3949044 https://doi.org/10.1007/s10687-018-0339-3

COLES, S. G. (1993). Regional modelling of extreme storms via max-stable processes. *J. Roy. Statist. Soc. Ser. B* **55** 797–816. MR1229882

DE HAAN, L. and LIN, T. (2001). On convergence toward an extreme value distribution in $C[0,1]$. *Ann. Probab.* **29** 467–483. MR1825160 https://doi.org/10.1214/aop/1008956340

DE HAAN, L. and LIN, T. (2003). Weak consistency of extreme value estimators in $C[0,1]$. *Ann. Statist.* **31** 1996–2012. MR2036397 https://doi.org/10.1214/aos/1074290334

DOMBRY, C. (2017). Personal communication.

DOMBRY, C. and RIBATET, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface* **8** 9–17. MR3320385 https://doi.org/10.4310/SII.2015.v8.n1.a2

DOMBRY, C., RIBATET, M. and STOEV, S. (2018). Probabilities of concurrent extremes. *J. Amer. Statist. Assoc.* **113** 1565–1582. MR3902230 https://doi.org/10.1080/01621459.2017.1356318

DREES, H. and HUANG, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64** 25–47. MR1619974 https://doi.org/10.1006/jmva.1997.1708

EINMAHL, J. H. J., GANTNER, M. and SAWITZKI, G. (2010). Asymptotics of the shorth plot. *J. Statist. Plann. Inference* **140** 3003–3012. MR2659831 https://doi.org/10.1016/j.jspi.2010.03.041

EINMAHL, J. H. J., KIRILIOUK, A. and SEGERS, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes* **21** 205–233. MR3800297 https://doi.org/10.1007/s10687-017-0303-7

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. MR3015043 https://doi.org/10.1214/12-AOS1023

EINMAHL, J. H. J. and LIN, T. (2006). Asymptotic normality of extreme value estimators on $C[0,1]$. *Ann. Statist.* **34** 469–492. MR2275250 https://doi.org/10.1214/009053605000000831

EINMAHL, J. H. J. and SEGERS, J. (2021). Supplement to "Empirical tail copulas for functional data." https://doi.org/10.1214/21-AOS2050SUPP.

MSC2020 subject classifications. Primary 62M99, 62G05, 62G20, 62G30, 62G32; secondary 60F17, 60G70.
Key words and phrases. Extreme value statistics, functional data, tail empirical process, tail dependence, tail copula estimation, uniform asymptotic normality.

FALK, M. (2019). *Multivariate Extreme Value Theory and D-Norms*. *Springer Series in Operations Research and Financial Engineering*. Springer, Cham. MR3890054 https://doi.org/10.1007/978-3-030-03819-9

FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717–1737. MR3263087 https://doi.org/10.3150/13-BEJ538

GINÉ, E., HAHN, M. G. and VATAN, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probab. Theory Related Fields* **87** 139–165. MR1080487 https://doi.org/10.1007/BF01198427

GOMES, M. I., DE HAAN, L. and PESTANA, D. (2004). Joint exceedances of the ARCH process. *J. Appl. Probab.* **41** 919–926. MR2074832 https://doi.org/10.1017/s0021900200020635

GUDENDORF, G. and SEGERS, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *J. Statist. Plann. Inference* **142** 3073–3085. MR2956794 https://doi.org/10.1016/j.jspi.2012.05.007

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699 https://doi.org/10.1016/0167-7152(89)90106-5

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440 https://doi.org/10.1214/09-AOP455

KOCH, E. (2017). Spatial risk measures and applications to max-stable processes. *Extremes* **20** 635–670. MR3679985 https://doi.org/10.1007/s10687-016-0274-0

MERCADIER, C. and ROUSTANT, O. (2019). The tail dependograph. *Extremes* **22** 343–372. MR3949049 https://doi.org/10.1007/s10687-019-00345-3

NIKOLOULOPOULOS, A. K., JOE, H. and LI, H. (2009). Extreme value properties of multivariate *t* copulas. *Extremes* **12** 129–148. MR2515644 https://doi.org/10.1007/s10687-008-0072-4

PENG, L. and QI, Y. (2008). Bootstrap approximation of tail dependence function. *J. Multivariate Anal.* **99** 1807–1824. MR2444820 https://doi.org/10.1016/j.jmva.2008.01.018

R CORE TEAM (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RESSEL, P. (2013). Homogeneous distributions—and a spectral representation of classical mean values and stable tail dependence functions. *J. Multivariate Anal.* **117** 246–256. MR3053546 https://doi.org/10.1016/j.jmva.2013.02.013

RIBATET, M. (2020). `SpatialExtremes`: Modelling spatial extremes. R package version 2.0-9.

RUYMGAART, F. H. (1973). *Asymptotic Theory of Rank Tests for Independence*. Mathematisch Centrum, Amsterdam Mathematical Centre Tracts, 43. MR0324816

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786 https://doi.org/10.1023/A:1020977924878

SCHMIDT, R. and STADTMÜLLER, U. (2006). Non-parametric estimation of tail dependence. *Scand. J. Stat.* **33** 307–335. MR2279645 https://doi.org/10.1111/j.1467-9469.2005.00483.x

SEGERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900 https://doi.org/10.3150/11-BEJ387

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.

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

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

The Annals of Statistics
2021, Vol. 49, No. 5, 2697–2720
https://doi.org/10.1214/21-AOS2051

INFERENCE FOR A TWO-STAGE ENRICHMENT DESIGN

BY ZHANTAO LIN[1,*], NANCY FLOURNOY[2,??] AND WILLIAM F. ROSENBERGER[1,†]

[1]*Department of Statistics, George Mason University,* [*]*zlin4@masonlive.gmu.edu,* [†]*wrosenbe@gmu.edu*
[2]*Department of Statistics, University of Missouri, Columbia,* [??]*flournoyn@missouri.edu*

Two-stage enrichment designs can be used to target the benefiting population in clinical trials based on patients' biomarkers. In the case of continuous biomarkers, we show that using a bivariate model that treats biomarkers as random variables more accurately identifies a treatment-benefiting enriched population than assuming biomarkers are fixed. Additionally, we show that under the bivariate model, the maximum likelihood estimators (MLEs) follow a randomly scaled mixture of normal distributions. Using random normings, we obtain asymptotically standard normal MLEs and construct hypothesis tests. Finally, in a simulation study, we demonstrate that our proposed design is more powerful than a single stage design when outcomes and biomarkers are correlated; the model-based estimators have smaller bias and mean square error (MSE) than weighted average estimators.

REFERENCES

ALDOUS, D. J. and EAGLESON, G. K. (1978). On mixing and stability of limit theorems. *Ann. Probab.* **6** 325–331. MR0517416 https://doi.org/10.1214/aop/1176995577

AUDEH, M. W., CARMICHAEL, J., PENSON, R. T., FRIEDLANDER, M., POWELL, B., BELL-MCGUINN, K. M., SCOTT, C., WEITZEL, J. N., OAKNIN, A. et al. (2010). Oral poly (ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: A proof-of-concept trial. *Lancet* **376** 245–251.

BARNDORFF-NIELSEN, O. E. and SØRENSEN, M. (1994). A review of some aspects of asymptotic likelihood theory for stochastic processes. *Int. Stat. Rev.* **62** 133–165.

CHAPMAN, P. B., HAUSCHILD, A., ROBERT, C., HAANEN, J. B., ASCIERTO, P., LARKIN, J., DUMMER, R., GARBE, C., TESTORI, A. et al. (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364** 2507–2516.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics. Princeton Mathematical Series, Vol.* 9. Princeton Univ. Press, Princeton, NJ. MR0016588

FERGUSON, T. S. (1996). *A Course in Large Sample Theory. Texts in Statistical Science Series*. CRC Press, London. MR1699953 https://doi.org/10.1007/978-1-4899-4549-5

FLOURNOY, N., MAY, C. and TOMMASI, C. (in press). The effects of adaptation on maximum likelihood inference for nonlinear models with normal errors. *J. Statist. Plann. Inference*. https://doi.org/10.1016/j.jspi.2021.02.002

FONG, P. C., BOSS, D. S., YAP, T. A., TUTT, A., WU, P., MERGUI-ROELVINK, M., MORTIMER, P., SWAISLAND, H., LAU, A. et al. (2009). Inhibition of poly (ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361** 123–134.

FRIEDE, T., PARSONS, N. and STALLARD, N. (2012). A conditional error function approach for subgroup selection in adaptive clinical trials. *Stat. Med.* **31** 4309–4320. MR3040083 https://doi.org/10.1002/sim.5541

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

HÄUSLER, E. and LUSCHGY, H. (2015). *Stable Convergence and Stable Limit Theorems. Probability Theory and Stochastic Modelling* **74**. Springer, Cham. MR3362567 https://doi.org/10.1007/978-3-319-18329-9

HOLGERSSON, H. E. T., NORDSTRÖM, L. and ÖNER, Ö. (2014). Dummy variables vs. category-wise models. *J. Appl. Stat.* **41** 233–241. MR3291232 https://doi.org/10.1080/02664763.2013.838665

JENNISON, C. and TURNBULL, B. W. (2007). Adaptive seamless designs: Selection and prospective testing of hypotheses. *J. Biopharm. Statist.* **17** 1135–1161. MR2414567 https://doi.org/10.1080/10543400701645215

MSC2020 subject classifications. Primary 62L05, 62K99, 62F12; secondary 62E20, 60F05, 60G52.
Key words and phrases. Adaptive designs, design of experiments, inference for stochastic processes, precision medicine, random biomarker, threshold determination.

LANE, A. and FLOURNOY, N. (2012). Two-stage adaptive optimal design with fixed first-stage sample size. *J. Probab. Stat.* **15** Art. ID 436239. MR2988297 https://doi.org/10.1155/2012/436239

LANE, A., YAO, P. and FLOURNOY, N. (2014). Information in a two-stage adaptive optimal design. *J. Statist. Plann. Inference* **144** 173–187. MR3131372 https://doi.org/10.1016/j.jspi.2013.07.010

LEHMANN, E. L. (1999). *Elements of Large-Sample Theory*. *Springer Texts in Statistics*. Springer, New York. MR1663158 https://doi.org/10.1007/b98855

LIN, Z., FLOURNOY, N. and ROSENBERGER, W. F. (2020). Random norming aids analysis of non-linear regression models with sequential informative dose selection. *J. Statist. Plann. Inference* **206** 29–42. MR4036693 https://doi.org/10.1016/j.jspi.2019.09.003

MAITOURNAM, A. and SIMON, R. (2005). On the efficiency of targeted clinical trials. *Stat. Med.* **24** 329–339. MR2134510 https://doi.org/10.1002/sim.1975

MANDREKAR, S. J. and SARGENT, D. J. (2009). Clinical trial designs for predictive biomarker validation: Theoretical considerations and practical challenges. *J. Clin. Oncol.* **27** 4027.

PRAKASA RAO, B. L. S. (2009). Conditional independence, conditional mixing and conditional association. *Ann. Inst. Statist. Math.* **61** 441–460. MR2505397 https://doi.org/10.1007/s10463-007-0152-2

RÉNYI, A. (1963). On stable sequences of events. *Sankhyā Ser. A* **25** 293–302. MR0170385

ROSENBLUM, M., FANG, E. X. and LIU, H. (2020). Optimal, two-stage, adaptive enrichment designs for randomized trials, using sparse linear programming. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 749–772. MR4112783 https://doi.org/10.1111/rssb.12366

ROSENBLUM, M. and VAN DER LAAN, M. J. (2011). Optimizing randomized trial designs to distinguish which subpopulations benefit from treatment. *Biometrika* **98** 845–860. MR2860328 https://doi.org/10.1093/biomet/asr055

ROSSET, S. and TIBSHIRANI, R. J. (2020). From Fixed-X to Random-X regression: Bias-variance decompositions, covariance penalties, and prediction error estimation. *J. Amer. Statist. Assoc.* **115** 138–151. MR4078450 https://doi.org/10.1080/01621459.2018.1424632

RUSSEK-COHEN, E. and SIMON, R. M. (1997). Evaluating treatments when a gender by treatment interaction may exist. *Stat. Med.* **16** 455–464.

SCHEPERS, J. (2016). On regression modelling with dummy variables versus separate regressions per group: Comment on Holgersson et al. *J. Appl. Stat.* **43** 674–681.

SIMON, N. (2015). Adaptive enrichment designs: Applications and challenges. *Clin. Invest.* **5** 383–391.

SIMON, R. and MAITOURNAM, A. (2004). Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin. Cancer Res.* **10** 6759–6763.

SIMON, N. and SIMON, R. (2013). Adaptive enrichment designs for clinical trials. *Biostatistics* **14** 613–625.

SIMON, R. and SIMON, N. (2017). Inference for multimarker adaptive enrichment trials. *Stat. Med.* **36** 4083–4093. MR3713651 https://doi.org/10.1002/sim.7422

SPENCER, A. V., HARBRON, C., MANDER, A., WASON, J. and PEERS, I. (2016). An adaptive design for updating the threshold value of a continuous biomarker. *Stat. Med.* **35** 4909–4923. MR3569903 https://doi.org/10.1002/sim.7042

STALLARD, N., HAMBORG, T., PARSONS, N. and FRIEDE, T. (2014). Adaptive designs for confirmatory clinical trials with subgroup selection. *J. Biopharm. Statist.* **24** 168–187. MR3196133 https://doi.org/10.1080/10543406.2013.857238

TARIMA, S. and FLOURNOY, N. (2019). Asymptotic properties of maximum likelihood estimators with sample size recalculation. *Statist. Papers* **60** 23–44. MR3969035 https://doi.org/10.1007/s00362-019-01095-x

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge.

WANG, S.-J., HUNG, H. M. J. and O'NEILL, R. T. (2009). Adaptive patient enrichment designs in therapeutic trials. *Biom. J.* **51** 358–374. MR2668689 https://doi.org/10.1002/bimj.200900003

WANG, S.-J., O'NEILL, R. T. and HUNG, H. J. (2007). Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm. Stat.* **6** 227–244.

ZHANG, Z., CHEN, R., SOON, G. and ZHANG, H. (2018). Treatment evaluation for a data-driven subgroup in adaptive enrichment designs of clinical trials. *Stat. Med.* **37** 1–11. MR3738057 https://doi.org/10.1002/sim.7497

The Annals of Statistics
2021, Vol. 49, No. 5, 2721–2754
https://doi.org/10.1214/21-AOS2052

EXISTENCE AND UNIQUENESS OF THE KRONECKER COVARIANCE MLE

BY MATHIAS DRTON[1,??], SATOSHI KURIKI[2,??] AND PETER HOFF[3,??]

[1]*Department of Mathematics, Technical University of Munich,* [??]*mathias.drton@tum.de*
[2]*The Institute of Statistical Mathematics,* [??]*kuriki@ism.ac.jp*
[3]*Department of Statistical Science, Duke University,* [??]*peter.hoff@duke.edu*

In matrix-valued datasets the sampled matrices often exhibit correlations among both their rows and their columns. A useful and parsimonious model of such dependence is the matrix normal model, in which the covariances among the elements of a random matrix are parameterized in terms of the Kronecker product of two covariance matrices, one representing row covariances and one representing column covariance. An appealing feature of such a matrix normal model is that the Kronecker covariance structure allows for standard likelihood inference even when only a very small number of data matrices is available. For instance, in some cases a likelihood ratio test of dependence may be performed with a sample size of one. However, more generally the sample size required to ensure boundedness of the matrix normal likelihood or the existence of a unique maximizer depends in a complicated way on the matrix dimensions. This motivates the study of how large a sample size is needed to ensure that maximum likelihood estimators exist, and exist uniquely with probability one. Our main result gives precise sample size thresholds in the paradigm where the number of rows and the number of columns of the data matrices differ by at most a factor of two. Our proof uses invariance properties that allow us to consider data matrices in canonical form, as obtained from the Kronecker canonical form for matrix pencils.

REFERENCES

[1] ALLEN, G. I. and TIBSHIRANI, R. (2010). Transposable regularized covariance models with an application to missing data imputation. *Ann. Appl. Stat.* **4** 764–790. MR2758420 https://doi.org/10.1214/09-AOAS314

[2] ALLEN, G. I. and TIBSHIRANI, R. (2012). Inference with transposable data: Modelling the effects of row and column correlations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 721–743. MR2965957 https://doi.org/10.1111/j.1467-9868.2011.01027.x

[3] AMÉNDOLA, C., KOHN, K., REICHENBACH, P. and SEIGAL, A. (2020). Invariant theory and scaling algorithms for maximum likelihood estimation.

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

[5] CHEN, X. and LIU, W. (2019). Graph estimation for matrix-variate Gaussian data. *Statist. Sinica* **29** 479–504. MR3889377

[6] DAWID, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68** 265–274. MR0614963 https://doi.org/10.1093/biomet/68.1.265

[7] DERKSEN, H. and MAKAM, V. (2020). Maximum likelihood estimation for matrix normal models via quiver representations.

[8] DRTON, M., FOX, C., KÄUFL, A. and POULIOT, G. (2019). The maximum likelihood threshold of a path diagram. *Ann. Statist.* **47** 1536–1553. MR3911121 https://doi.org/10.1214/18-AOS1724

[9] DUTILLEUL, P. (1999). The mle algorithm for the matrix normal distribution. *J. Stat. Comput. Simul.* **64** 105–123. https://doi.org/10.1080/00949659908811970

[10] EDELMAN, A., ELMROTH, E. and KÅGSTRÖM, B. (1997). A geometric approach to perturbation theory of matrices and matrix pencils. I. Versal deformations. *SIAM J. Matrix Anal. Appl.* **18** 653–692. MR1453545 https://doi.org/10.1137/S0895479895284634

MSC2020 subject classifications. Primary 62H12; secondary 62R01.

Key words and phrases. Gaussian distribution, Kronecker canonical form, matrix normal model, maximum likelihood estimation, separable covariance.

[11] EFRON, B. (2009). Are a set of microarrays independent of each other? *Ann. Appl. Stat.* **3** 922–942. MR2750220 https://doi.org/10.1214/09-AOAS236

[12] FOSDICK, B. K. and HOFF, P. D. (2014). Separable factor analysis with applications to mortality data. *Ann. Appl. Stat.* **8** 120–147. MR3191985 https://doi.org/10.1214/13-AOAS694

[13] GLANZ, H. and CARVALHO, L. (2018). An expectation-maximization algorithm for the matrix normal distribution with an application in remote sensing. *J. Multivariate Anal.* **167** 31–48. MR3830632 https://doi.org/10.1016/j.jmva.2018.03.010

[14] GREENEWALD, K. and HERO, A. O. III (2015). Robust Kronecker product PCA for spatio-temporal covariance estimation. *IEEE Trans. Signal Process.* **63** 6368–6378. MR3421045 https://doi.org/10.1109/TSP.2015.2472364

[15] GROSS, E. and SULLIVANT, S. (2018). The maximum likelihood threshold of a graph. *Bernoulli* **24** 386–407. MR3706762 https://doi.org/10.3150/16-BEJ881

[16] KOCH, D., LELE, S. and LEWIS, M. A. (2020). Computationally simple anisotropic lattice covariograms. *Environ. Ecol. Stat.* **27** 665–688.

[17] LANDSBERG, J. M. (2012). *Tensors: Geometry and Applications. Graduate Studies in Mathematics* **128**. Amer. Math. Soc., Providence, RI. MR2865915 https://doi.org/10.1090/gsm/128

[18] MAKEIG, S., KOTHE, C., MULLEN, T., BIGDELY-SHAMLO, N., ZHANG, Z. and KREUTZ-DELGADO, K. (2012). Evolving signal processing for brain–computer interfaces. *Proc. IEEE* **100** 1567–1584.

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

[20] MUROTA, K. (2000). *Matrices and Matroids for Systems Analysis. Algorithms and Combinatorics* **20**. Springer, Berlin. MR1739147

[21] OHLSON, M., RAUF AHMAD, M. and VON ROSEN, D. (2013). The multilinear normal distribution: Introduction and some basic properties. *J. Multivariate Anal.* **113** 37–47. MR2984354 https://doi.org/10.1016/j.jmva.2011.05.015

[22] RAPCSÁK, T. (1997). *Smooth Nonlinear Optimization in* $\mathbf{R}^N$. *Nonconvex Optimization and Its Applications* **19**. Kluwer Academic, Dordrecht. MR1480415 https://doi.org/10.1007/978-1-4615-6357-0

[23] SOLOVEYCHIK, I. and TRUSHIN, D. (2016). Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness. *J. Multivariate Anal.* **149** 92–113. MR3507317 https://doi.org/10.1016/j.jmva.2016.04.001

[24] TEN BERGE, J. M. F. and KIERS, H. A. L. (1999). Simplicity of core arrays in three-way principal component analysis and the typical rank of $p \times q \times 2$ arrays. *Linear Algebra Appl.* **294** 169–179. MR1693919 https://doi.org/10.1016/S0024-3795(99)00057-9

[25] VOLFOVSKY, A. and HOFF, P. D. (2015). Testing for nodal dependence in relational data matrices. *J. Amer. Statist. Assoc.* **110** 1037–1046. MR3420682 https://doi.org/10.1080/01621459.2014.965777

[26] WIESEL, A. (2012). Geodesic convexity and covariance estimation. *IEEE Trans. Signal Process.* **60** 6182–6189. MR3006411 https://doi.org/10.1109/TSP.2012.2218241

[27] YIN, J. and LI, H. (2012). Model selection and estimation in the matrix normal graphical model. *J. Multivariate Anal.* **107** 119–140. MR2890437 https://doi.org/10.1016/j.jmva.2012.01.005

[28] ZHOU, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Ann. Statist.* **42** 532–562. MR3210978 https://doi.org/10.1214/13-AOS1187

[29] ZHU, Y. and LI, L. (2018). Multiple matrix Gaussian graphs estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 927–950. MR3874304 https://doi.org/10.1111/rssb.12278

The Annals of Statistics
2021, Vol. 49, No. 5, 2755–2773
https://doi.org/10.1214/21-AOS2054

PREDICTION BOUNDS FOR HIGHER ORDER TOTAL VARIATION REGULARIZED LEAST SQUARES

BY FRANCESCO ORTELLI* AND SARA VAN DE GEER†

*Seminar for Statistics, ETH Zürich, *fortelli@ethz.ch, †vsara@ethz.ch*

We establish adaptive results for trend filtering: least squares estimation with a penalty on the total variation of $(k-1)$th order differences. Our approach is based on combining a general oracle inequality for the ℓ_1-penalized least squares estimator with "interpolating vectors" to upper bound the "effective sparsity." This allows one to show that the ℓ_1-penalty on the kth order differences leads to an estimator that can adapt to the number of jumps in the $(k-1)$th order differences of the underlying signal or an approximation thereof. We show the result for $k \in \{1, 2, 3, 4\}$ and indicate how it could be derived for general $k \in \mathbb{N}$.

REFERENCES

BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806. MR2860324 https://doi.org/10.1093/biomet/asr043

BOYER, C., DE CASTRO, Y. and SALMON, J. (2017). Adapting to unknown noise level in sparse deconvolution. *Inf. Inference* **6** 310–348. MR3764527 https://doi.org/10.1093/imaiai/iaw024

CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2014). Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.* **67** 906–956. MR3193963 https://doi.org/10.1002/cpa.21455

CANDÈS, E. J. and PLAN, Y. (2011). A probabilistic and RIPless theory of compressed sensing. *IEEE Trans. Inf. Theory* **57** 7235–7254. MR2883653 https://doi.org/10.1109/TIT.2011.2161794

CHATTERJEE, S. and GOSWAMI, S. (2019). Adaptive estimation of multivariate piecewise polynomials and bounded variation functions by optimal decision trees. Preprint. Available at arXiv:1911.11562.

DALALYAN, A. S., HEBIRI, M. and LEDERER, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. MR3556784 https://doi.org/10.3150/15-BEJ756

DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26** 879–921. MR1635414 https://doi.org/10.1214/aos/1024691081

ELAD, M., MILANFAR, P. and RUBINSTEIN, R. (2007). Analysis versus synthesis in signal priors. *Inverse Probl.* **23** 947–968. MR2329926 https://doi.org/10.1088/0266-5611/23/3/007

GUNTUBOYINA, A., LIEU, D., CHATTERJEE, S. and SEN, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist.* **48** 205–229. MR4065159 https://doi.org/10.1214/18-AOS1799

KIM, S.-J., KOH, K., BOYD, S. and GORINEVSKY, D. (2009). l_1 trend filtering. *SIAM Rev.* **51** 339–360. MR2505584 https://doi.org/10.1137/070690274

LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 https://doi.org/10.1214/aos/1015957395

MAMMEN, E. and VAN DE GEER, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635

ORTELLI, F. and VAN DE GEER, S. (2020a). Adaptive rates for total variation image denoising. *J. Mach. Learn. Res.* **21** 1–38.

ORTELLI, F. and VAN DE GEER, S. (2020b). Oracle inequalities for square root analysis estimators with application to total variation penalties. *Inf. Inference* iaaa002.

ORTELLI, F. and VAN DE GEER, S. (2021). Supplement to "Prediction bounds for higher order total variation regularized least squares." https://doi.org/10.1214/21-AOS2054SUPP

PADILLA, O. H. M. and CHATTERJEE, S. (2020). Adaptive quantile trend filtering. Preprint. Available at arXiv:2007.07472.

MSC2020 subject classifications. Primary 62J05; secondary 62J99.

Key words and phrases. Oracle inequality, projection, compatibility, lasso, analysis, total variation regularization, minimax, Moore–Penrose pseudo inverse.

QIAN, J. and JIA, J. (2016). On stepwise pattern recovery of the fused Lasso. *Comput. Statist. Data Anal.* **94** 221–237. MR3412821 https://doi.org/10.1016/j.csda.2015.08.013

RUDIN, L. I., OSHER, S. and FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D* **60** 259–268. MR3363401 https://doi.org/10.1016/0167-2789(92)90242-F

SADHANALA, V. and TIBSHIRANI, R. J. (2019). Additive models with trend filtering. *Ann. Statist.* **47** 3032–3068. MR4025734 https://doi.org/10.1214/19-AOS1833

SADHANALA, V., WANG, Y.-X., SHARPNACK, J. and TIBSHIRANI, R. (2017). Higher-order total variation classes on grids: Minimax theory and trend filtering methods. In *Advances in Neural Information Processing Systems* 5800–5810.

STEIDL, G., DIDAS, S. and NEUMANN, J. (2006). Splines in higher order TV regularization. *Int. J. Comput. Vis.* **70** 241–255.

TANG, G., BHASKAR, B. N. and RECHT, B. (2015). Near minimax line spectral estimation. *IEEE Trans. Inf. Theory* **61** 499–512. MR3299978 https://doi.org/10.1109/TIT.2014.2368122

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

TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

TIBSHIRANI, R. (2020). Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. Preprint. Available at arXiv:2003.03886.

VAN DE GEER, S. (2020). Logistic regression with total variation regularization. *Trans. A. Razmadze Math. Inst.* **174** 217–233. MR4150558

WANG, Y.-X., SMOLA, A. and TIBSHIRANI, R. (2014). The falling factorial basis and its statistical applications. In *International Conference on Machine Learning* 730–738.

WANG, Y.-X., SHARPNACK, J., SMOLA, A. J. and TIBSHIRANI, R. J. (2016). Trend filtering on graphs. *J. Mach. Learn. Res.* **17** Paper No. 105, 41. MR3543511

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

The Annals of Statistics
2021, Vol. 49, No. 5, 2774–2802
https://doi.org/10.1214/21-AOS2055

RECONCILING THE GAUSSIAN AND WHITTLE LIKELIHOOD WITH AN APPLICATION TO ESTIMATION IN THE FREQUENCY DOMAIN

BY SUHASINI SUBBA RAO[1,??] AND JUNHO YANG[2,??]

[1]*Department of Statistics, Texas A&M University,* [??]*suhasini@stat.tamu.edu*
[2]*Institute of Statistical Science, Academia Sinica,* [??]*junhoyang@stat.sinica.edu.tw*

In time series analysis there is an apparent dichotomy between time and frequency domain methods. The aim of this paper is to draw connections between frequency and time domain methods. Our focus will be on reconciling the Gaussian likelihood and the Whittle likelihood. We derive an exact, interpretable, bound between the Gaussian and Whittle likelihood of a second order stationary time series. The derivation is based on obtaining the transformation which is biorthogonal to the discrete Fourier transform of the time series. Such a transformation yields a new decomposition for the inverse of a Toeplitz matrix and enables the representation of the Gaussian likelihood within the frequency domain. We show that the difference between the Gaussian and Whittle likelihood is due to the omission of the best linear predictions outside the domain of observation in the periodogram associated with the Whittle likelihood. Based on this result, we obtain an approximation for the difference between the Gaussian and Whittle likelihoods in terms of the best fitting, finite order autoregressive parameters. These approximations are used to define two new frequency domain quasi-likelihood criteria. We show that these new criteria can yield a better approximation of the spectral divergence criterion, as compared to both the Gaussian and Whittle likelihoods. In simulations, we show that the proposed estimators have satisfactory finite sample properties.

REFERENCES

ABADIR, K. M., DISTASO, W. and GIRAITIS, L. (2007). Nonstationarity-extended local Whittle estimation. *J. Econometrics* **141** 1353–1384. MR2413505 https://doi.org/10.1016/j.jeconom.2007.01.020

BARTLETT, M. S. (1953). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika* **40** 306–317. MR0059519 https://doi.org/10.2307/2333349

BAXTER, G. (1962). An asymptotic result for the finite predictor. *Math. Scand.* **10** 137–144. MR0149584 https://doi.org/10.7146/math.scand.a-10520

BHANSALI, R. J. (1982). The evaluation of certain quadratic forms occurring in autoregressive model fitting. *Ann. Statist.* **10** 121–131. MR0642724

BHANSALI, R. J. (1996). Asymptotically efficient autoregressive model selection for multistep prediction. *Ann. Inst. Statist. Math.* **48** 577–602. MR1424784 https://doi.org/10.1007/BF00050857

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

CHENG, R. and POURAHMADI, M. (1993). Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probab. Theory Related Fields* **95** 115–124. MR1207310 https://doi.org/10.1007/BF01197341

CHOUDHURI, N., GHOSAL, S. and ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99** 1050–1059. MR2109494 https://doi.org/10.1198/016214504000000557

COURSOL, J. and DACUNHA-CASTELLE, D. (1982). Remarques sur l'approximation de la vraisemblance d'un processus gaussien stationnaire. *Teor. Veroyatn. Primen.* **27** 155–160. MR0645138

COX, D. R. and SNELL, E. J. (1968). A general definition of residuals. *J. Roy. Statist. Soc. Ser. B* **30** 248–275. MR0237052

MSC2020 subject classifications. Primary 62M10, 62M15; secondary 62F10.
Key words and phrases. Biorthogonal transforms, discrete Fourier transform, periodogram, quasi-likelihoods, second order stationary time series, Toeplitz inverse.

DAHLHAUS, R. (1983). Spectral analysis with tapered data. *J. Time Series Anal.* **4** 163–175. MR0732895 https://doi.org/10.1111/j.1467-9892.1983.tb00366.x

DAHLHAUS, R. (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Ann. Statist.* **16** 808–841. MR0947580 https://doi.org/10.1214/aos/1176350838

DAHLHAUS, R. (1990). Nonparametric high resolution spectral estimation. *Probab. Theory Related Fields* **85** 147–180. MR1050742 https://doi.org/10.1007/BF01277980

DAHLHAUS, R. and KÜNSCH, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74** 877–882. MR0919857 https://doi.org/10.1093/biomet/74.4.877

DAS, S., SUBBA RAO, S. and YANG, J. (2021). Spectral methods for small sample time series: A complete periodogram approach. *J. Time Series Anal.*

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532. MR0840512 https://doi.org/10.1214/aos/1176349936

GALBRAITH, R. F. and GALBRAITH, J. I. (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. *J. Appl. Probab.* **11** 63–71. MR0365959 https://doi.org/10.1017/s0021900200036408

GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London. MR2977317 https://doi.org/10.1142/p591

GIRAITIS, L. and ROBINSON, P. M. (2001). Whittle estimation of arch models. *Econometric Theory* **17** 608–631. MR1841822 https://doi.org/10.1017/S0266466601173056

HURVICH, C. M. and CHEN, W. W. (2000). An efficient taper for potentially overdifferenced long-memory time series. *J. Time Series Anal.* **21** 155–180. MR1771841 https://doi.org/10.1111/1467-9892.00179

ING, C.-K. and WEI, C.-Z. (2005). Order selection for same-realization predictions in autoregressive processes. *Ann. Statist.* **33** 2423–2474. MR2211091 https://doi.org/10.1214/009053605000000525

INOUE, A. and KASAHARA, Y. (2006). Explicit representation of finite predictor coefficients and its applications. *Ann. Statist.* **34** 973–993. MR2283400 https://doi.org/10.1214/009053606000000209

INOUE, A., KASAHARA, Y. and POURAHMADI, M. (2018). Baxter's inequality for finite predictor coefficients of multivariate long-memory stationary processes. *Bernoulli* **24** 1202–1232. MR3706792 https://doi.org/10.3150/16-BEJ897

KASAHARA, Y., POURAHMADI, M. and INOUE, A. (2009). Duals of random vectors and processes with applications to prediction problems with missing values. *Statist. Probab. Lett.* **79** 1637–1646. MR2547424 https://doi.org/10.1016/j.spl.2009.04.005

KIRCH, C., EDWARDS, M. C., MEIER, A. and MEYER, R. (2019). Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis. *Bayesian Anal.* **14** 1037–1073. MR4044845 https://doi.org/10.1214/18-BA1126

KLEY, T., PREUSS, P. and FRYZLEWICZ, P. (2019). Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electron. J. Stat.* **13** 3710–3774. MR4013749 https://doi.org/10.1214/19-ejs1606

KREISS, J.-P., PAPARODITIS, E. and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130. MR2893863 https://doi.org/10.1214/11-AOS900

LAHIRI, S. N. (2003). A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence *Ann. Statist.* **31** 613–641. MR1983544 https://doi.org/10.1214/aos/1051027883

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

PARZEN, E. (1983). Autoregressive spectral estimation. In *Time Series in the Frequency Domain*. *Handbook of Statist.* **3** 221–247. North-Holland, Amsterdam. MR0749788 https://doi.org/10.1016/S0169-7161(83)03013-8

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. *Vol.* 1. *Probability and Mathematical Statistics*. Academic Press, London–New York. MR0628735

ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301 https://doi.org/10.1214/aos/1176324317

SHAMAN, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Ann. Statist.* **3** 532–538. MR0362784

SHAMAN, P. (1976). Approximations for stationary covariance matrices and their inverses with application to ARMA models. *Ann. Statist.* **4** 292–301. MR0398033

SHAO, X. and WU, W. B. (2007). Local Whittle estimation of fractional integration for nonlinear processes. *Econometric Theory* **23** 899–929. MR2396737 https://doi.org/10.1017/S0266466607070387

SIDDIQUI, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *Ann. Math. Stat.* **29** 585–588. MR0095568 https://doi.org/10.1214/aoms/1177706636

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

Subba Rao, S. and Yang, J. (2021). Supplement to "Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain." https://doi.org/10.1214/21-AOS2055SUPP

Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M. and Early, J. J. (2019). The debiased Whittle likelihood. *Biometrika* **106** 251–266. MR3949302 https://doi.org/10.1093/biomet/asy071

Szegö, G. (1921). Über die Randwerte einer analytischen Funktion. *Math. Ann.* **84** 232–244. MR1512033 https://doi.org/10.1007/BF01459407

van Delft, A. and Eichler, M. (2020). A note on Herglotz's theorem for time series on function spaces. *Stochastic Process. Appl.* **130** 3687–3710. MR4092417 https://doi.org/10.1016/j.spa.2019.10.006

Walker, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Aust. Math. Soc.* **4** 363–384. MR0171345

Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Thesis, Uppsala University. MR0040634

Whittle, P. (1953). The analysis of multiple stationary time series. *J. Roy. Statist. Soc. Ser. B* **15** 125–139. MR0056902

Zhang, H.-C. (1992). Reduction of the asymptotic bias of autoregressive and spectral estimators by tapering. *J. Time Series Anal.* **13** 451–469. MR1183157 https://doi.org/10.1111/j.1467-9892.1992.tb00120.x

The Annals of Statistics
2021, Vol. 49, No. 5, 2803–2824
https://doi.org/10.1214/21-AOS2057

INTEGRATIVE METHODS FOR POST-SELECTION INFERENCE UNDER CONVEX CONSTRAINTS

BY SNIGDHA PANIGRAHI[1,??], JONATHAN TAYLOR[2,??] AND ASAF WEINSTEIN[3,??]

[1]*Department of Statistics, University of Michigan,* [??]*psnigdha@umich.edu*
[2]*Department of Statistics, Stanford University,* [??]*jtaylo@stanford.edu*
[3]*Department of Statistics, Hebrew University of Jerusalem,* [??]*asaf.weinstein@mail.huji.ac.il*

Inference after model selection has been an active research topic in the past few years, with numerous works offering different approaches to addressing the perils of the reuse of data. In particular, major progress has been made recently on large and useful classes of problems by harnessing general theory of hypothesis testing in exponential families, but these methods have their limitations. Perhaps most immediate is the gap between theory and practice: implementing the exact theoretical prescription in realistic situations—for example, when new data arrives and inference needs to be adjusted accordingly—may be a prohibitive task. In this paper, we propose a Bayesian framework for carrying out inference after variable selection. Our framework is very flexible in the sense that it naturally accommodates different models for the data instead of requiring a case-by-case treatment. This flexibility is achieved by considering the full selective likelihood function where, crucially, we propose a novel and nontrivial approximation to the exact but intractable expression. The advantages of our methods in practical data analysis are demonstrated in an application to HIV drug-resistance data.

REFERENCES

[1] ARCONES, M. A. (2002). Moderate deviations for M-estimators. *TEST* **11** 465–500. MR1947607 https://doi.org/10.1007/BF02595717

[2] BI, N., MARKOVIC, J., XIA, L. and TAYLOR, J. (2020). Inferactive data analysis. *Scand. J. Stat.* **47** 212–249. MR4075236 https://doi.org/10.1111/sjos.12425

[3] EICHELSBACHER, P. and LÖWE, M. (2003). Moderate deviations for i.i.d. random variables. *ESAIM Probab. Stat.* **7** 209–218. MR1956079 https://doi.org/10.1051/ps:2003005

[4] FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. arXiv preprint, arXiv:1410.2597.

[5] LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

[6] MARKOVIC, J. and TAYLOR, J. (2016). Bootstrap inference after using multiple queries for model selection. arXiv preprint, arXiv:1612.07811.

[7] NEGAHBAN, S., YU, B., WAINWRIGHT, M. J. and RAVIKUMAR, P. K. (2009). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. In *Advances in Neural Information Processing Systems* 1348–1356.

[8] PANIGRAHI, S. and TAYLOR, J. (2019). Approximate selective inference via maximum likelihood. arXiv preprint, arXiv:1902.07884.

[9] PANIGRAHI, S., TAYLOR, J. and WEINSTEIN, A. (2021). Supplement to "Integrative methods for post-selection inference under convex constraints." https://doi.org/10.1214/21-AOS2057SUPP

[10] PANIGRAHI, S., ZHU, J. and SABATTI, C. (2021). Selection-adjusted inference: An application to confidence intervals for cis-eQTL effect sizes. *Biostatistics* **22** 181–197. MR4207150 https://doi.org/10.1093/biostatistics/kxz024

[11] PANIGRAHI, S., MOHAMMED, S., RAO, A. and BALADANDAYUTHAPANI, V. (2020). Integrative bayesian models using post-selective inference: A case study in radiogenomics. arXiv preprint, arXiv:2004.12012.

MSC2020 subject classifications. Primary 62F30; secondary 62F10, 62G15.
Key words and phrases. Adaptive data analysis, Bayesian inference, carving, conditional inference, convex constraints, selective inference.

[12] REID, S., TAYLOR, J. and TIBSHIRANI, R. (2017). Post-selection point and interval estimation of signal sizes in Gaussian samples. *Canad. J. Statist.* **45** 128–148. MR3646193 https://doi.org/10.1002/cjs.11320

[13] RHEE, S.-Y., TAYLOR, J., WADHERA, G., BEN-HUR, A., BRUTLAG, D. L. and SHAFER, R. W. (2006). Genotypic predictors of human immunodeficiency virus type 1 drug resistance. *Proc. Natl. Acad. Sci. USA* **103** 17355–17360.

[14] TIAN, X. and TAYLOR, J. (2018). Selective inference with a randomized response. *Ann. Statist.* **46** 679–710. MR3782381 https://doi.org/10.1214/17-AOS1564

[15] TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689 https://doi.org/10.1080/01621459.2015.1108848

[16] TIBSHIRANI, R., TIBSHIRANI, R., TAYLOR, J., LOFTUS, J., REID, S. and MARKOVIC, J. (2019). selectiveInference: Tools for post-selection inference. R package version 1.2.5.

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

[18] WEINSTEIN, A., FITHIAN, W. and BENJAMINI, Y. (2013). Selection adjusted confidence intervals with more power to determine the sign. *J. Amer. Statist. Assoc.* **108** 165–176. MR3174610 https://doi.org/10.1080/01621459.2012.737740

[19] YEKUTIELI, D. (2012). Adjusted Bayesian inference for selected parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 515–541. MR2925372 https://doi.org/10.1111/j.1467-9868.2011.01016.x

[20] YU, G., BIEN, J. and TIBSHIRANI, R. (2019). Reluctant interaction modeling. arXiv preprint, arXiv:1907.08414.

[21] ZÖLLNER, S. and PRITCHARD, J. K. (2007). Overcoming the winner's curse: Estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* **80** 605–615.

The Annals of Statistics
2021, Vol. 49, No. 5, 2825–2850
https://doi.org/10.1214/21-AOS2058

THE COST OF PRIVACY: OPTIMAL RATES OF CONVERGENCE FOR PARAMETER ESTIMATION WITH DIFFERENTIAL PRIVACY

BY T. TONY CAI[1,*], YICHEN WANG[1,†] AND LINJUN ZHANG[2,??]

[1]*Department of Statistics, The Wharton School, University of Pennsylvania,* [*]*tcai@wharton.upenn.edu,* [†]*wangyc@wharton.upenn.edu*

[2]*Department of Statistics, Rutgers University,* [??]*linjun.zhang@rutgers.edu*

Privacy-preserving data analysis is a rising challenge in contemporary statistics, as the privacy guarantees of statistical methods are often achieved at the expense of accuracy. In this paper, we investigate the tradeoff between statistical accuracy and privacy in mean estimation and linear regression, under both the classical low-dimensional and modern high-dimensional settings. A primary focus is to establish minimax optimality for statistical estimation with the (ε, δ)-differential privacy constraint. By refining the "tracing adversary" technique for lower bounds in the theoretical computer science literature, we improve existing minimax lower bound for low-dimensional mean estimation and establish new lower bounds for high-dimensional mean estimation and linear regression problems. We also design differentially private algorithms that attain the minimax lower bounds up to logarithmic factors. In particular, for high-dimensional linear regression, a novel private iterative hard thresholding algorithm is proposed. The numerical performance of differentially private algorithms is demonstrated by simulation studies and applications to real data sets.

REFERENCES

[1] ABADI, M., CHU, A., GOODFELLOW, I., MCMAHAN, H. B., MIRONOV, I., TALWAR, K. and ZHANG, L. (2016). Deep learning with differential privacy. In *ACM CCS* 2016 308–318. ACM, New York.

[2] ABOWD, J. M. (2016). The challenge of scientific reproducibility and privacy protection for statistical agencies. In *Census Scientific Advisory Committee*.

[3] AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2010). Fast global convergence of gradient methods for high-dimensional statistical recovery. In *Advances in Neural Information Processing Systems* 37–45.

[4] BASSILY, R., FELDMAN, V., TALWAR, K. and THAKURTA, A. G. (2019). Private stochastic convex optimization with optimal rates. In *Advances in Neural Information Processing Systems* 11282–11291.

[5] BASSILY, R., SMITH, A. and THAKURTA, A. (2014). Private empirical risk minimization: Efficient algorithms and tight error bounds. In 55*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2014 464–473. IEEE Computer Soc., Los Alamitos, CA. MR3344896 https://doi.org/10.1109/FOCS.2014.56

[6] BISWAS, S., DONG, Y., KAMATH, G. and COINPRESS, J. U. (2020). Practical private mean and covariance estimation. Preprint. Available at arXiv:2006.06618.

[7] BLUMENSATH, T. and DAVIES, M. E. (2009). Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.* **27** 265–274. MR2559726 https://doi.org/10.1016/j.acha.2009.04.002

[8] BU, Z., DONG, J. LONG, Q. and SU, W. J. (2019). Deep learning with gaussian differential privacy. Preprint. Available at arXiv:1911.11607.

[9] BUN, M., ULLMAN, J. and VADHAN, S. (2014). Fingerprinting codes and the price of approximate differential privacy. In *STOC'*14*—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 1–10. ACM, New York. MR3238925 https://doi.org/10.1145/2591796.2591877

[10] CAI, T. T., WANG, Y. and ZHANG, L. (2021). Supplement to "The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy." https://doi.org/10.1214/21-AOS2058SUPP

MSC2020 subject classifications. Primary 62F30; secondary 62F12, 62J05.

Key words and phrases. High-dimensional data, differential privacy, mean estimation, linear regression, minimax optimality.

[11] DING, B., KULKARNI, J. and YEKHANIN, S. (2017). Collecting telemetry data privately. In *NeurIPS* 2017 3571–3580.

[12] DONG, J. ROTH, A. and SU, W. J. (2019). Gaussian differential privacy. Preprint. Availabel at arXiv:1905.02383.

[13] DONOHO, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270. MR1272082 https://doi.org/10.1214/aos/1176325367

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

[15] DWORK, C. and FELDMAN, V. (2018). Privacy-preserving prediction. Preprint. Available at arXiv:1803.10266.

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

[17] DWORK, C. and ROTH, A. (2013). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–487. MR3254020 https://doi.org/10.1561/0400000042

[18] DWORK, C. and ROTHBLUM, G. N. (2016). Concentrated differential privacy. Preprint. Available at arXiv:1603.01887.

[19] DWORK, C., SMITH, A., STEINKE, T. and ULLMAN, J. (2017). Exposed! A survey of attacks on private data. *Annu. Rev. Stat. Appl.* **4** 61–84.

[20] DWORK, C., SMITH, A., STEINKE, T., ULLMAN, J. and VADHAN, S. (2015). Robust traceability from trace amounts. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 650–669. IEEE Computer Soc., Los Alamitos, CA. MR3473333 https://doi.org/10.1109/FOCS.2015.46

[21] DWORK, C., SU, W. J. and ZHANG, L. (2018). Differentially private false discovery rate control. Preprint. Available at arXiv:1807.04209.

[22] DWORK, C., TALWAR, K., THAKURTA, A. and ZHANG, L. (2014). Analyze Gauss: Optimal bounds for privacy-preserving principal component analysis. In *STOC'*14*—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 11–20. ACM, New York. MR3238926 https://doi.org/10.1145/2591796.2591883

[23] ERLINGSSON, Ú., PIHUR, V. and RAPPOR, A. K. (2014). Randomized aggregatable privacy-preserving ordinal response. In *ACM CCS* 2014 1054–1067. ACM, New York.

[24] HOMER, N., SZELINGER, S., REDMAN, M., DUGGAN, D., TEMBE, W., MUEHLING, J., PEARSON, J. V., STEPHAN, D. A., NELSON, S. F. et al. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* **4** e1000167. https://doi.org/10.1371/journal.pgen.1000167

[25] JAIN, P., TEWARI, A. and KAR, P. (2014). On iterative hard thresholding methods for high-dimensional m-estimation. In *NeurIPS* 2014 685–693.

[26] JOHNSTONE, I. M. (1994). On minimax estimation of a sparse normal mean vector. *Ann. Statist.* **22** 271–289. MR1272083 https://doi.org/10.1214/aos/1176325368

[27] KAMATH, G., LI, J., SINGHAL, V. and ULLMAN, J. (2018). Privately learning high-dimensional distributions. Preprint. Available at arXiv:1805.00216.

[28] KARWA, V. and VADHAN, S. (2017). Finite sample differentially private confidence intervals. Preprint. Available at arXiv:1711.03908.

[29] KASIVISWANATHAN, S. P., LEE, H. K., NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2011). What can we learn privately? *SIAM J. Comput.* **40** 793–826. MR2823508 https://doi.org/10.1137/090756090

[30] KIFER, D., SMITH, A. and THAKURTA, A. G. (2012). Private convex empirical risk minimization and high-dimensional regression. In *COLT* 2012 25.1–25.40.

[31] LEI, J. (2011). Differentially private m-estimators. In *NeurIPS* 2011 361–369.

[32] LOH, P.-L. and WAINWRIGHT, M. J. (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **16** 559–616. MR3335800

[33] LOWTHER, C., MERICO, D., COSTAIN, G., WASERMAN, J., BOYD, K., NOOR, A., SPEEVAK, M., STAVROPOULOS, D. J., WEI, J. et al. (2017). Impact of IQ on the diagnostic yield of chromosomal microarray in a community sample of adults with schizophrenia. *Gen. Med.* **9** 105.

[34] MCSHERRY, F. and TALWAR, K. (2007). Mechanism design via differential privacy. In *FOCS* 2007 **7** 94–103.

[35] MIRONOV, I. (2017). Renyi differential privacy. In *CSF* 2017 263–275. IEEE, New York.

[36] NEGAHBAN, S., YU, B., WAINWRIGHT, M. J. and RAVIKUMAR, P. K. (2009). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. In *Advances in Neural Information Processing Systems* 1348–1356.

[37] PACE, R. K. and BARRY, R. (1997). Sparse spatial autoregressions. *Statist. Probab. Lett.* **33** 291–297.

[38] RAAB, M. and STEGER, A. (1998). "Balls into bins"—A simple and tight analysis. In *Randomization and Approximation Techniques in Computer Science* (*Barcelona*, 1998). *Lecture Notes in Computer Science* **1518** 159–170. Springer, Berlin. MR1729169 https://doi.org/10.1007/3-540-49543-6_13

[39] ROHDE, A. and STEINBERGER, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.* **48** 2646–2670. MR4152116 https://doi.org/10.1214/19-AOS1901

[40] SHEFFET, O. (2017). Differentially private ordinary least squares. In *Proceedings of the* 34*th International Conference on Machine Learning* **70** 3105–3114. JMLR.org.

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

[42] STEINKE, T. and ULLMAN, J. (2017). Between pure and approximate differential privacy. *J. Priv. Confid.* **7**.

[43] STEINKE, T. and ULLMAN, J. (2017). Tight lower bounds for differentially private selection. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 552–563. IEEE Computer Soc., Los Alamitos, CA. MR3734260 https://doi.org/10.1109/FOCS.2017.57

[44] TALWAR, K., THAKURTA, A. G. and ZHANG, L. (2015). Nearly optimal private lasso. In *NeurIPS* 2015 3025–3033.

[45] APPLE DIFFERENTIAL PRIVACY TEAM (2017). Privacy at scale.

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

[47] WANG, Y.-X., FIENBERG, S. and SMOLA, A. (2015). Privacy for free: Posterior sampling and stochastic gradient Monte Carlo. In *ICML* 2493–2502.

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

The Annals of Statistics
2021, Vol. 49, No. 5, 2851–2869
https://doi.org/10.1214/21-AOS2062

DISTRIBUTED STATISTICAL INFERENCE FOR MASSIVE DATA

BY SONG XI CHEN[1,??] AND LIUHUA PENG[2,??]

[1]*Guanghua School of Management and Center for Statistical Science, Peking University,* [??]*songxichen@pku.edu.cn*
[2]*School of Mathematics and Statistics, University of Melbourne,* [??]*liuhua.peng@unimelb.edu.au*

This paper considers distributed statistical inference for general symmetric statistics in the context of massive data with efficient computation. Estimation efficiency and asymptotic distributions of the distributed statistics are provided, which reveal different results between the nondegenerate and degenerate cases, and show the number of the data subsets plays an important role. Two distributed bootstrap methods are proposed and analyzed to approximation the underlying distribution of the distributed statistics with improved computation efficiency over existing methods. The accuracy of the distributional approximation by the bootstrap are studied theoretically. One of the methods, the pseudo-distributed bootstrap, is particularly attractive if the number of datasets is large as it directly resamples the subset-based statistics, assumes less stringent conditions and its performance can be improved by studentization.

REFERENCES

ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586 https://doi.org/10.1214/aos/1176348650

ATTA-ASIAMAH, E. and YUAN, M. (2019). Distributed inference for degenerate U-statistics. *Stat* **8** e234, 8. MR3978408 https://doi.org/10.1002/sta4.234

BATTEY, H., FAN, J., LIU, H., LU, J. and ZHU, Z. (2018). Distributed testing and estimation under sparse high dimensional models. *Ann. Statist.* **46** 1352–1382. MR3798006 https://doi.org/10.1214/17-AOS1587

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. MR471142 https://doi.org/10.1214/aos/1176344134

CHANG, J. and HALL, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika* **102** 203–214. MR3335106 https://doi.org/10.1093/biomet/asu060

CHEN, S. X. and HALL, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21** 1166–1181. MR1241263 https://doi.org/10.1214/aos/1176349256

CHEN, S. X and PENG, L. (2021). Supplement to "Distributed statistical inference for massive data." https://doi.org/10.1214/21-AOS2062SUPP

CHEN, X. and XIE, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statist. Sinica* **24** 1655–1684. MR3308656

DAVIDSON, R. and MACKINNON, J. G. (2002). Fast double bootstrap tests of nonnested linear regression models. *Econometric Rev.* **21** 419–429. MR1951651 https://doi.org/10.1081/ETC-120015384

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR515681 https://doi.org/10.1214/aos/1176344552

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7

HE, X. and SHAO, Q.-M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* **24** 2608–2630. MR1425971 https://doi.org/10.1214/aos/1032181172

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I: Statistics* **5.1** 221–233. Univ. California Press, Berkeley, CA. MR0216620

JING, B.-Y. and WANG, Q. (2010). A unified approach to Edgeworth expansions for a general class of statistics. *Statist. Sinica* **20** 613–636. MR2682633

MSC2020 subject classifications. Primary 62G09; secondary 62G20.
Key words and phrases. Distributed bootstrap, distributed statistics, massive data, pseudo-distributed bootstrap.

KLEINER, A., TALWALKAR, A., SARKAR, P. and JORDAN, M. I. (2014). A scalable bootstrap for massive data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 795–816. MR3248677 https://doi.org/10.1111/rssb.12050

LAHIRI, S. N. (1994). On two-term Edgeworth expansions and bootstrap approximations for Studentized multivariate M-estimators. *Sankhyā Ser. A* **56** 201–226. MR1664912

LAI, T. L. and WANG, J. Q. (1993). Edgeworth expansions for symmetric statistics with applications to bootstrap methods. *Statist. Sinica* **3** 517–542. MR1243399

LIN, N. and XI, R. (2010). Fast surrogates of U-statistics. *Comput. Statist. Data Anal.* **54** 16–24. MR2558454 https://doi.org/10.1016/j.csda.2009.08.009

LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. MR964947 https://doi.org/10.1214/aos/1176351062

SENGUPTA, S., VOLGUSHEV, S. and SHAO, X. (2016). A subsampled double bootstrap for massive data. *J. Amer. Statist. Assoc.* **111** 1222–1232. MR3561944 https://doi.org/10.1080/01621459.2015.1080709

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165 https://doi.org/0.1002/9780470316481

SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. MR1201705 https://doi.org/10.2307/2951780

SKOVGAARD, I. M. (1981). Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Stat.* **8** 207–217. MR0642801

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

VOLGUSHEV, S., CHAO, S.-K. and CHENG, G. (2019). Distributed inference for quantile regression processes. *Ann. Statist.* **47** 1634–1662. MR3911125 https://doi.org/10.1214/18-AOS1730

ZHANG, Y., DUCHI, J. C. and WAINWRIGHT, M. J. (2013). Communication-efficient algorithms for statistical optimization. *J. Mach. Learn. Res.* **14** 3321–3363. MR3144464

The Annals of Statistics
2021, Vol. 49, No. 5, 2870–2884
https://doi.org/10.1214/21-AOS2063

CONSTRUCTION OF MIXED ORTHOGONAL ARRAYS WITH HIGH STRENGTH

BY SHANQI PANG[1,??], JING WANG[2,??], DENNIS K. J. LIN[3,??] AND MIN-QIAN LIU[4,??]

[1]*College of Mathematics and Information Science, Henan Normal University,* [??]*pangshanqi@263.net*
[2]*College of Mathematics and Science, Shanghai Normal University,* [??]*wangjing19@126.com*
[3]*Department of Statistics, Purdue University,* [??]*dkjlin@purdue.edu*
[4]*School of Statistics and Data Science, LPMC & KLMDASR, Nankai University,* [??]*mqliu@nankai.edu.cn*

A considerable portion of the work on mixed orthogonal arrays applies specifically to arrays of strength 2. Although strength $t = 2$ is arguably the most important case for statistical applications, there is an urgent need for better methods for $t \geq 3$. However, the knowledge on the existence of arrays for $t \geq 3$ is rather limited. In this paper, new construction methods for symmetric and asymmetric orthogonal arrays (OAs) with high strength are proposed by using lower strength orthogonal partitions of spaces and OAs. A positive answer is provided to the open problem in Hedayat, Sloane and Stufken (*Orthogonal Arrays: Theory and Applications* (1999) Springer) on developing better methods and tools for the construction of mixed orthogonal arrays with strength $t \geq 3$. Not only are the methods straightforward, but also they are useful for constructing symmetric or asymmetric OAs of arbitrary strengths, numbers of levels and various sizes. The constructed OAs can be utilized to generate more OAs. The resulting OAs have a high degree of flexibility and many other desirable properties. Some selective OAs are tabulated for practical uses.

REFERENCES

AGGARWAL, M. L. and BUDHRAJA, V. (2002). On construction of some new symmetric and asymmetric orthogonal arrays. *J. Discrete Math. Sci. Cryptogr.* **5** 215–225. MR1986731 https://doi.org/10.1080/09720529.2002.10697950

BIERBRAUER, J. (2005). *Introduction to Coding Theory. Discrete Mathematics and Its Applications*. CRC Press/CRC, Boca Raton, FL. MR2079734

CARLET, C. and CHEN, X. (2018). Constructing low-weight dth-order correlation-immune Boolean functions through the Fourier–Hadamard transform. *IEEE Trans. Inf. Theory* **64** 2969–2978. MR3784600 https://doi.org/10.1109/TIT.2017.2785775

CHEN, G., JI, L. and LEI, J. (2014). The existence of mixed orthogonal arrays with four and five factors of strength two. *J. Combin. Des.* **22** 323–342. MR3218310 https://doi.org/10.1002/jcd.21350

CHEN, G. Z. and LEI, J. G. (2017). Constructions of mixed orthogonal arrays of strength three. *Sci. Sin., Math.* **47** 545–564 (in Chinese).

CHÊNG, C. S. (1980). Orthogonal arrays with variable numbers of symbols. *Ann. Statist.* **8** 447–453. MR0560740

COLBOURN, C. J. and DINITZ, J. H. (2007). *The CRC Handbook of Combinatorial Designs*. Chapman & Hall/CRC, Boca Raton, FL.

DEAN, A., MORRIS, M., STUFKEN, J. and BINGHAM, D., eds. (2015). *Handbook of Design and Analysis of Experiments. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3642972

GOYENECHE, D. and ŻYCZKOWSKI, K. (2014). Genuinely multipartite entangled states and orthogonal arrays. *Phys. Rev. A* **90** 022316.

HE, Y. and TANG, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100** 254–260. MR3034340 https://doi.org/10.1093/biomet/ass065

HE, Y. and TANG, B. (2014). A characterization of strong orthogonal arrays of strength three. *Ann. Statist.* **42** 1347–1360. MR3226159 https://doi.org/10.1214/14-AOS1225

MSC2020 subject classifications. Primary 62K15; secondary 05B15.
Key words and phrases. High strength orthogonal arrays, Kronecker product, orthogonal partitions.

HEDAYAT, A. S., SLOANE, N. J. A. and STUFKEN, J. (1999). *Orthogonal Arrays*: *Theory and Applications*. *Springer Series in Statistics*. Springer, New York. With a foreword by C. R. Rao. MR1693498 https://doi.org/10.1007/978-1-4612-1478-6

HEDAYAT, A. S., STUFKEN, J. and SU, G. (1996). On difference schemes and orthogonal arrays of strength t. *J. Statist. Plann. Inference* **56** 307–324. MR1436014 https://doi.org/10.1016/S0378-3758(96)00026-2

JI, L. and YIN, J. (2010). Constructions of new orthogonal arrays and covering arrays of strength three. *J. Combin. Theory Ser. A* **117** 236–247. MR2592899 https://doi.org/10.1016/j.jcta.2009.06.002

JIANG, B. and AI, M. (2017). Construction of uniform U-designs. *J. Statist. Plann. Inference* **181** 1–10. MR3567993 https://doi.org/10.1016/j.jspi.2016.08.003

JIANG, L. and YIN, J. (2013). An approach of constructing mixed-level orthogonal arrays of strength ≥ 3. *Sci. China Math.* **56** 1109–1115. MR3063958 https://doi.org/10.1007/s11425-013-4616-y

KUHFELD, W. F. (2018). Orthogonal arrays. Available at http://support.sas.com/techsup/technote/ts723.html.

LIN, C. D., MUKERJEE, R. and TANG, B. (2009). Construction of orthogonal and nearly orthogonal Latin hypercubes. *Biometrika* **96** 243–247. MR2482150 https://doi.org/10.1093/biomet/asn064

LIN, C. D., BINGHAM, D., SITTER, R. R. and TANG, B. (2010). A new and flexible method for constructing designs for computer experiments. *Ann. Statist.* **38** 1460–1477. MR2662349 https://doi.org/10.1214/09-AOS757

LIU, H. and LIU, M.-Q. (2015). Column-orthogonal strong orthogonal arrays and sliced strong orthogonal arrays. *Statist. Sinica* **25** 1713–1734. MR3409089

LO, H. K., CURTY, M. and QI, B. (2012). Measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **108** 130503.

MUKERJEE, R., SUN, F. and TANG, B. (2014). Nearly orthogonal arrays mappable into fully orthogonal arrays. *Biometrika* **101** 957–963. MR3286928 https://doi.org/10.1093/biomet/asu042

PANG, S. (2004). Construction methods of orthogonal arrays and their applications. Ph.D. thesis, Xidian Univ. Xian, China (in Chinese).

PANG, S., WANG, X., WANG, J., DU, J. and FENG, M. (2018). Construction and count of 1-resilient rotation symmetric Boolean functions. *Inform. Sci.* **450** 336–342. MR3789898 https://doi.org/10.1016/j.ins.2018.03.052

PANG, S., ZHANG, X., LIN, X. and ZHANG, Q. (2019). Two and three-uniform states from irredundant orthogonal arrays. *npj Quantum Inf.* **5** 52.

PANG, S., WANG, J., LIN, D. K. J. and LIU, M.-Q. (2021). Supplement to "Construction of mixed orthogonal arrays with high strength." https://doi.org/10.1214/21-AOS2063SUPP

PENG, J., MUKERJEE, R. and LIN, D. K. J. (2019). Design of order-of-addition experiments. *Biometrika* **106** 683–694. MR3992397 https://doi.org/10.1093/biomet/asz025

RIEBE, M., HÄFFNER, H., ROOS, C. F., HÄNSEL, W., BENHELM, J., LANCASTER, G. P. T., KÖRBER, T. W., BECHER, C., SCHMIDT-KALER, F. et al. (2004). Deterministic quantum teleportation with atoms. *Nature* **429** 734–737.

SCHOEN, E. D., EENDEBAK, P. T. and NGUYEN, M. V. M. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *J. Combin. Des.* **18** 123–140. MR2604638 https://doi.org/10.1002/jcd.20236

STINSON, D. R. (2004). *Combinatorial Designs*: *Constructions and Analysis*. Springer, New York. With a foreword by Charles J. Colbourn. MR2029249

STINSON, D. R. (2018). Ideal ramp schemes and related combinatorial objects. *Discrete Math.* **341** 299–307. MR3724097 https://doi.org/10.1016/j.disc.2017.08.041

SUEN, C., DAS, A. and DEY, A. (2001). On the construction of asymmetric orthogonal arrays. *Statist. Sinica* **11** 241–260. MR1820009

SUEN, C. and DEY, A. (2003). Construction of asymmetric orthogonal arrays through finite geometries. *J. Statist. Plann. Inference* **115** 623–635. MR1985887 https://doi.org/10.1016/S0378-3758(02)00165-9

SUN, F., LIU, M.-Q. and QIAN, P. Z. G. (2014). On the construction of nested space-filling designs. *Ann. Statist.* **42** 1394–1425. MR3226161 https://doi.org/10.1214/14-AOS1229

SUN, F. and TANG, B. (2017). A method of constructing space-filling orthogonal designs. *J. Amer. Statist. Assoc.* **112** 683–689. MR3671762 https://doi.org/10.1080/01621459.2016.1159211

TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* **88** 1392–1397. MR1245375

VOELKEL, J. G. (2019). The design of order-of-addition experiments. *J. Qual. Technol.* **51** 230–241.

WANG, J. C. and WU, C. F. J. (1992). Nearly orthogonal arrays with mixed levels and small runs. *Technometrics* **34** 409–422.

XU, H. and WU, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Ann. Statist.* **29** 549–560. MR1863969 https://doi.org/10.1214/aos/1009210552

YIN, Y., LIN, D. K. J. and LIU, M.-Q. (2014). Sliced Latin hypercube designs via orthogonal arrays. *J. Statist. Plann. Inference* **149** 162–171. MR3199902 https://doi.org/10.1016/j.jspi.2014.02.008

YIN, J., WANG, J., JI, L. and LI, Y. (2011). On the existence of orthogonal arrays $\mathrm{OA}(3, 5, 4n+2)$. *J. Combin. Theory Ser. A* **118** 270–276. MR2737200 https://doi.org/10.1016/j.jcta.2010.03.013

ZHANG, Y. (2006). Data analysis and constructions of orthogonal arrays. Ph.D. thesis, East China Normal Univ., Shanghai, China (in Chinese).

ZHANG, Y. (2007). Orthogonal arrays obtained by repeating-column difference matrices. *Discrete Math*. **307** 246–261. MR2285195 https://doi.org/10.1016/j.disc.2006.06.029

ZHANG, T., DENG, Q. and DEY, A. (2017). Construction of asymmetric orthogonal arrays of strength three via a replacement method. *J. Combin*. *Des*. **25** 339–348. MR3666988 https://doi.org/10.1002/jcd.21557

ZHANG, Y., LU, Y. and PANG, S. (1999). Orthogonal arrays obtained by orthogonal decomposition of projection matrices. *Statist*. *Sinica* **9** 595–604. MR1707856

ZHANG, Y., PANG, S. and WANG, Y. (2001). Orthogonal arrays obtained by generalized Hadamard product. *Discrete Math*. **238** 151–170. MR1843905 https://doi.org/10.1016/S0012-365X(00)00421-0

ZHANG, T., ZONG, Y. and DEY, A. (2016). On the construction of asymmetric orthogonal arrays. *J. Statist*. *Plann*. *Inference* **170** 77–82. MR3427861 https://doi.org/10.1016/j.jspi.2015.09.004

ZHAO, Z., CHEN, Y. A., ZHANG, A. N., YANG, T., BRIEGEL, H. and PAN, J. W. (2004). Experimental demonstration of five-photon entanglement and open-destination teleportation. *Nature* **430** 54–58.

ZHOU, Y.-D. and XU, H. (2014). Space-filling fractional factorial designs. *J. Amer. Statist*. *Assoc*. **109** 1134–1144. MR3265686 https://doi.org/10.1080/01621459.2013.873367

The Annals of Statistics
2021, Vol. 49, No. 5, 2885–2915
https://doi.org/10.1214/21-AOS2064

FOUNDATIONS OF STRUCTURAL CAUSAL MODELS WITH CYCLES AND LATENT VARIABLES

BY STEPHAN BONGERS[1,*], PATRICK FORRÉ[1,†], JONAS PETERS[2,??] AND JORIS M. MOOIJ[3,??]

[1]*Informatics Institute, University of Amsterdam,* [*]*s.r.bongers@uva.nl,* [†]*p.d.forre@uva.nl*
[2]*Department of Mathematical Sciences, University of Copenhagen,* [??]*jonas.peters@math.ku.dk*
[3]*Korteweg–De Vries Institute, University of Amsterdam,* [??]*j.m.mooij@uva.nl*

Structural causal models (SCMs), also known as (nonparametric) structural equation models (SEMs), are widely used for causal modeling purposes. In particular, acyclic SCMs, also known as recursive SEMs, form a well-studied subclass of SCMs that generalize causal Bayesian networks to allow for latent confounders. In this paper, we investigate SCMs in a more general setting, allowing for the presence of both latent confounders and cycles. We show that in the presence of cycles, many of the convenient properties of acyclic SCMs do not hold in general: they do not always have a solution; they do not always induce unique observational, interventional and counterfactual distributions; a marginalization does not always exist, and if it exists the marginal model does not always respect the latent projection; they do not always satisfy a Markov property; and their graphs are not always consistent with their causal semantics. We prove that for SCMs in general each of these properties does hold under certain solvability conditions. Our work generalizes results for SCMs with cycles that were only known for certain special cases so far. We introduce the class of simple SCMs that extends the class of acyclic SCMs to the cyclic setting, while preserving many of the convenient properties of acyclic SCMs. With this paper, we aim to provide the foundations for a general theory of statistical causal modeling with SCMs.

REFERENCES

[1] BALKE, A. and PEARL, J. (1994). Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence* (*AAAI*-94) **1** 230–237. AAAI Press, Menlo Park.

[2] BECKERS, S. and HALPERN, J. Y. (2019). Abstracting causal models. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence* (*AAAI*-19) **33** 2678–2685. AAAI Press, Menlo Park.

[3] BLOM, T., BONGERS, S. and MOOIJ, J. M. (2019). Beyond structural causal models: Causal constraints models. In *Proceedings of the* 35*th Conference on Uncertainty in Artificial Intelligence* (*UAI*-19) (R. P. Adams and V. Gogate, eds.). AUAI Press.

[4] BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0996025 https://doi.org/10.1002/9781118619179

[5] BONGERS, S., BLOM, T. and MOOIJ, J. M. (2021). Causal modeling of dynamical systems. Preprint. Available at arXiv:1803.08784v3 [cs.AI].

[6] BONGERS, S., FORRÉ, P., PETERS, J. and MOOIJ, J. M. (2021). Supplement to "Foundations of structural causal models with cycles and latent variables." https://doi.org/10.1214/21-AOS2064SUPP

[7] BÜHLMANN, P., PETERS, J. and ERNEST, J. (2014). CAM: Causal additive models, high-dimensional order search and penalized regression. *Ann*. *Statist*. **42** 2526–2556. MR3277670 https://doi.org/10.1214/14-AOS1260

[8] BYRNE, R. M. J. (2007). *The Rational Imagination*: *How People Create Alternatives to Reality*. *A Bradford Book*. MIT Press, Cambridge, MA.

MSC2020 subject classifications. Primary 62A09, 68T30; secondary 68T37.
Key words and phrases. Structural causal models, causal graph, cycles, interventions, counterfactuals, solvability, Markov properties, marginalization.

[9] COOPER, G. F. (1997). A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Min. Knowl. Discov.* **1** 203–224.

[10] DAWID, A. P. (2002). Influence diagrams for causal modelling and inference. *Int. Stat. Rev.* **70** 161–189.

[11] DUNCAN, O. D. (1975). *Introduction to Structural Equation Models: Studies in Population*. Academic Press, New York. MR0398558

[12] EATON, D. and MURPHY, K. (2007). Exact Bayesian structure learning from uncertain interventions. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics* (M. Meila and X. Shen, eds.). *Proceedings of Machine Learning Research* **2** 107–114.

[13] EBERHARDT, F., HOYER, P. and SCHEINES, R. (2010). Combining experiments to discover linear cyclic models with latent variables. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. Teh and M. Titterington, eds.). *Proceedings of Machine Learning Research* **9** 185–192.

[14] EVANS, R. J. (2016). Graphs for margins of Bayesian networks. *Scand. J. Stat.* **43** 625–648. MR3543314 https://doi.org/10.1111/sjos.12194

[15] EVANS, R. J. (2018). Margins of discrete Bayesian networks. *Ann. Statist.* **46** 2623–2656. MR3851750 https://doi.org/10.1214/17-AOS1631

[16] FISHER, F. M. (1970). A correspondence principle for simultaneous equation models. *Econometrica* **38** 73–92.

[17] FORRÉ, P. and MOOIJ, J. M. (2017). Markov properties for graphical models with cycles and latent variables. Preprint. Available at arXiv:1710.08775 [math.ST].

[18] FORRÉ, P. and MOOIJ, J. M. (2018). Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. In *Proceedings of the* 34*th Conference on Uncertainty in Artificial Intelligence* (*UAI*-18) (A. Globerson and R. Silva, eds.). AUAI Press.

[19] FORRÉ, P. and MOOIJ, J. M. (2019). Causal calculus in the presence of cycles, latent confounders and selection bias. In *Proceedings of the* 35*th Conference on Uncertainty in Artificial Intelligence* (*UAI*-19) (R. P. Adams and V. Gogate, eds.). AUAI Press.

[20] FOYGEL, R., DRAISMA, J. and DRTON, M. (2012). Half-trek criterion for generic identifiability of linear structural equation models. *Ann. Statist.* **40** 1682–1713. MR3015040 https://doi.org/10.1214/12-AOS1012

[21] GOLDBERGER, A. S. and DUNCAN, O. D. (1973). *Structural Equation Models in the Social Sciences*. Seminar Press, New York.

[22] HAAVELMO, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11** 1–12. MR0007954 https://doi.org/10.2307/1905714

[23] HALPERN, J. (1998). Axiomatizing causal reasoning. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (*UAI*-98) (G. Cooper and S. Moral, eds.) 202–210. Morgan Kaufmann, San Francisco, CA, USA.

[24] HYTTINEN, A., EBERHARDT, F. and HOYER, P. O. (2012). Learning linear cyclic causal models with latent variables. *J. Mach. Learn. Res.* **13** 3387–3439. MR3005891

[25] HYTTINEN, A., HOYER, P. O., EBERHARDT, F. and JÄRVISALO, M. (2013). Discovering cyclic causal models with latent variables: A general SAT-based procedure. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence* (*UAI*-13) (A. Nicholson and P. Smyth, eds.) 301–310. AUAI Press, Corvallis, OR, USA.

[26] IWASAKI, Y. and SIMON, H. A. (1994). Causality and model abstraction. *Artificial Intelligence* **67** 143–194. MR1281657 https://doi.org/10.1016/0004-3702(94)90014-0

[27] KECHRIS, A. S. (1995). *Classical Descriptive Set Theory*. *Graduate Texts in Mathematics* **156**. Springer, New York. MR1321597 https://doi.org/10.1007/978-1-4612-4190-4

[28] KOSTER, J. T. A. (1996). Markov properties of nonrecursive causal models. *Ann. Statist.* **24** 2148–2177. MR1421166 https://doi.org/10.1214/aos/1069362315

[29] KOSTER, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Stat.* **26** 413–431. MR1712043 https://doi.org/10.1111/1467-9469.00157

[30] LACERDA, G., SPIRTES, P. L., RAMSEY, J. and HOYER, P. O. (2008). Discovering cyclic causal models by independent components analysis. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence* (*UAI*-08) (D. McAllester and P. Myllymaki, eds.) 366–374. AUAI Press, Corvallis, OR, USA.

[31] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Clarendon Press, Oxford. MR1419991

[32] LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H. G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491–505. MR1064735 https://doi.org/10.1002/net.3230200503

[33] LEWIS, D. K. (1979). Counterfactual dependence and time's arrow. *Noûs* **13** 455–476. https://doi.org/10.2307/2215339

[34] MAATHUIS, M. H., KALISCH, M. and BÜHLMANN, P. (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Statist.* **37** 3133–3164. MR2549555 https://doi.org/10.1214/09-AOS685

[35] MANI, S. (2006). A Bayesian local causal discovery framework. PhD thesis, Univ. Pittsburg.

[36] MASON, S. J. (1953). Feedback theory—Some properties of signal flow graphs. In *Proceedings of the IRE* **41** 1144–1156. IEEE.

[37] MASON, S. J. (1956). Feedback theory—Further properties of signal flow graphs. In *Proceedings of the IRE* **44** 920–926. IEEE.

[38] MEEK, C. (1995). Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (*UAI*-95) (P. Besnard and S. Hanks, eds.) 411–418. Morgan Kaufmann, San Francisco, CA, USA.

[39] MOGENSEN, S. W. and HANSEN, N. R. (2020). Markov equivalence of marginalized local independence graphs. *Ann. Statist.* **48** 539–559. MR4065173 https://doi.org/10.1214/19-AOS1821

[40] MOGENSEN, S. W., MALINSKY, D. and HANSEN, N. R. (2018). Causal learning for partially observed stochastic dynamical systems. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence* (*UAI*-18) (A. Globerson and R. Silva, eds.) AUAI Press.

[41] MOOIJ, J. M. and CLAASSEN, T. (2020). Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In *Proceedings of the* 36*th Conference on Uncertainty in Artificial Intelligence* (*UAI*-20) (J. Peters and D. Sontag, eds.) **124** 1159–1168. PMLR.

[42] MOOIJ, J. M. and HESKES, T. (2013). Cyclic causal discovery from continuous equilibrium data. In *Proceedings of the* 29*th Conference on Uncertainty in Artificial Intelligence* (*UAI*-13) (A. Nicholson and P. Smyth, eds.) 431–439. AUAI Press, Corvallis, OR, USA.

[43] MOOIJ, J. M., JANZING, D. and SCHÖLKOPF, B. (2013). From ordinary differential equations to structural causal models: The deterministic case. In *Proceedings of the* 29*th Conference on Uncertainty in Artificial Intelligence* (*UAI*-13) (A. Nicholson and P. Smyth, eds.) 440–448. AUAI Press.

[44] MOOIJ, J. M., MAGLIACANE, S. and CLAASSEN, T. (2020). Joint causal inference from multiple contexts. *J. Mach. Learn. Res.* **21** Paper No. 99, 108. MR4119167

[45] MOOIJ, J. M., PETERS, J., JANZING, D., ZSCHEISCHLER, J. and SCHÖLKOPF, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *J. Mach. Learn. Res.* **17** Paper No. 32, 102. MR3491126

[46] NEAL, R. M. (2000). On deducing conditional independence from d-separation in causal graphs with feedback. *J. Artificial Intelligence Res.* **12** 87–91. MR1745047 https://doi.org/10.1613/jair.689

[47] PEARL, J. (1985). A constraint propagation approach to probabilistic reasoning. In *Proceedings of the First Conference on Uncertainty in Artificial Intelligence* (*UAI*-85) (L. Kanal and J. Lemmer, eds.) 31–42. AUAI Press, Corvallis, OR, USA.

[48] PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

[49] PEARL, J. and DECHTER, R. (1996). Identifying independence in causal graphs with feedback. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence* (*UAI*-96) (E. Horvitz and F. Jensen, eds.) 420–426. Morgan Kaufmann, San Francisco, CA, USA.

[50] PEARL, J. and MACKENZIE, D. (2018). *The Book of Why*: *The New Science of Cause and Effect*. Basic Books, New York. MR3791446

[51] PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2017). *Elements of Causal Inference*: *Foundations and Learning Algorithms*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3822088

[52] PETERS, J., MOOIJ, J. M., JANZING, D. and SCHÖLKOPF, B. (2014). Causal discovery with continuous additive noise models. *J. Mach. Learn. Res.* **15** 2009–2053. MR3231600

[53] PFISTER, N., BAUER, S. and PETERS, J. (2019). Learning stable and predictive structures in kinetic systems. *Proc. Natl. Acad. Sci. USA* **116** 25405–25411. MR4047351 https://doi.org/10.1073/pnas.1905688116

[54] RICHARDSON, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30** 145–157. MR1963898 https://doi.org/10.1111/1467-9469.00323

[55] RICHARDSON, T. and SPIRTES, P. (1999). Automated discovery of linear feedback models. In *Computation*, *Causation*, *and Discovery* (C. Glymour and G. F. Cooper, eds.) 253–302. AAAI Press, Menlo Park, CA. MR1696455

[56] RICHARDSON, T. and SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015

[57] RICHARDSON, T. S. (1996). A discovery algorithm for directed cyclic graphs. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence* (*UAI*-96) (E. Horvitz and F. Jensen, eds.) 454–461. Morgan Kaufmann, San Francisco, CA, USA. MR1617227

[58] RICHARDSON, T. S. (1996). Discovering cyclic causal structure. Technical Report No. CMU-PHIL-68, Carnegie Mellon Univ.

[59] RICHARDSON, T. S. (1996). Models of feedback: Interpretation and discovery. Ph.D. thesis, Carnegie Mellon Univ.

[60] RICHARDSON, T. S. and ROBINS, J. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Technical Report No. 128, Center for Statistics and the Social Sciences.

[61] ROESE, N. J. (1997). Counterfactual thinking. *Psychol*. *Bull*. **121** 133–148.

[62] RUBENSTEIN, P. K., WEICHWALD, S., BONGERS, S., MOOIJ, J. M., JANZING, D., GROSSE-WENTRUP, M. and SCHÖLKOPF, B. (2017). Causal consistency of structural equation models. In *Proceedings of the* 33*rd Conference on Uncertainty in Artificial Intelligence* (*UAI*-17) (G. Elidan and K. Kersting, eds.). AUAI Press.

[63] RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ*. *Psychol*. **66** 688–701.

[64] SHPITSER, I. and PEARL, J. (2008). Complete identification methods for the causal hierarchy. *J. Mach*. *Learn*. *Res*. **9** 1941–1979. MR2447308

[65] SPIRTES, P. (1993). Directed cyclic graphs, conditional independence, and non-recursive linear structural equation models. Technical Report No. CMU-PHIL-35, Carnegie Mellon Univ.

[66] SPIRTES, P. (1994). Conditional independence in directed cyclic graphical models for feedback. Technical Report No. CMU-PHIL-54, Carnegie Mellon Univ.

[67] SPIRTES, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (*UAI*-95) (P. Besnard and S. Hanks, eds.) 499–506. Morgan Kaufmann, San Francisco, CA, USA.

[68] SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation*, *Prediction*, *and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR1815675

[69] SPIRTES, P., MEEK, C. and RICHARDSON, T. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. In *Computation*, *Causation*, *and Discovery* (C. Glymour and G. F. Cooper, eds.) 211–252. AAAI Press, Menlo Park, CA. MR1689972

[70] SPIRTES, P., RICHARDSON, T., MEEK, C., SCHEINES, R. and GLYMOUR, C. (1998). Using path diagrams as a structural equation modelling tool. *Sociol*. *Methods Res*. **27** 182–225.

[71] TIAN, J. (2002). Studies in causal reasoning and learning. Technical Report No. R-309, Cognitive Systems Laboratory, Univ. California, Los Angeles, USA.

[72] TIAN, J. and PEARL, J. (2001). Causal discovery from changes. In *Proceedings of the* 17*th Conference in Uncertainty in Artificial Intelligence* (*UAI*-01) (J. Breese and D. Koller, eds.) 512–521. Morgan Kaufmann, San Francisco, CA, USA.

[73] VERMA, T. S. (1993). Graphical aspects of causal models. Technical Report No. R-191. Computer Science Department, Univ. California, Los Angeles, USA.

[74] WRIGHT, S. (1921). Correlation and causation. *J. Agric*. *Res*. **20** 557–585.

[75] ZHANG, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence* **172** 1873–1896. MR2459793 https://doi.org/10.1016/j.artint.2008.08.001

The Annals of Statistics
2021, Vol. 49, No. 5, 2916–2947
https://doi.org/10.1214/21-AOS2065

MEASURING DEPENDENCE IN THE WASSERSTEIN DISTANCE FOR BAYESIAN NONPARAMETRIC MODELS

BY MARTA CATALANO[1,??], ANTONIO LIJOI[2,*] AND IGOR PRÜNSTER[2,†]

[1]*Department of Statistics, University of Warwick,* [??]*marta.catalano@warwick.ac.uk*
[2]*Department of Decision Sciences, Bocconi University,* [*]*antonio.lijoi@unibocconi.it,* [†]*igor.pruenster@unibocconi.it*

The proposal and study of dependent Bayesian nonparametric models has been one of the most active research lines in the last two decades, with random vectors of measures representing a natural and popular tool to define them. Nonetheless, a principled approach to understand and quantify the associated dependence structure is still missing. We devise a general, and not model-specific, framework to achieve this task for random measure based models, which consists in: (a) quantify dependence of a random vector of probabilities in terms of closeness to exchangeability, which corresponds to the maximally dependent coupling with the same marginal distributions, that is, the comonotonic vector; (b) recast the problem in terms of the underlying random measures (in the same Fréchet class) and quantify the closeness to comonotonicity; (c) define a distance based on the Wasserstein metric, which is ideally suited for spaces of measures, to measure the dependence in a principled way. Several results, which represent the very first in the area, are obtained. In particular, useful bounds in terms of the underlying Lévy intensities are derived relying on compound Poisson approximations. These are then specialized to popular models in the Bayesian literature leading to interesting insights.

REFERENCES

[1] BACALLADO, S., DIACONIS, P. and HOLMES, S. (2015). de Finetti priors using Markov chain Monte Carlo computations. *Stat. Comput.* **25** 797–808. MR3360493 https://doi.org/10.1007/s11222-015-9562-9

[2] BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

[3] CAMERLENGHI, F., LIJOI, A. and PRÜNSTER, I. (2021). Survival analysis via hierarchically dependent mixture hazards. *Ann. Statist.* **49** 863 – 884. https://doi.org/10.1214/20-AOS1982

[4] CATALANO, M., LIJOI, A. and PRÜNSTER, I. (2020). Approximation of Bayesian models for time-to-event data. *Electron. J. Stat.* **14** 3366–3395. MR4148234 https://doi.org/10.1214/20-EJS1746

[5] CIFARELLI, D. M. and REGAZZINI, E. (1978). Nonparametric statistical problems under partial exchangeability: The role of associative means. *Quaderni Istituto Matematica Finanziaria Dell'Università di Torino Serie III* **12** 1–36.

[6] CONT, R. and TANKOV, P. (2004). *Financial Modelling with Jump Processes. Chapman & Hall/CRC Financial Mathematics Series*. CRC Press/CRC, Boca Raton, FL. MR2042661

[7] CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* **26** 2292–2300.

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

[9] DE BLASI, P., PECCATI, G. and PRÜNSTER, I. (2009). Asymptotics for posterior hazards. *Ann. Statist.* **37** 1906–1945. MR2533475 https://doi.org/10.1214/08-AOS631

[10] DE FINETTI, B. (1938). Sur la condition d'Equivalence partielle. *Actual. Sci. Ind.* **739** 5–18.

[11] DOKSUM, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201. MR0373081 https://doi.org/10.1214/aop/1176996703

MSC2020 subject classifications. 62C10, 60G57, 60G09.
Key words and phrases. Bayesian nonparametrics, completely random measures, completely random vectors, compound Poisson, dependence, independent increments, Lévy copula, Wasserstein distance.

[12] DYKSTRA, R. L. and LAUD, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367. MR0606619

[13] EPIFANI, I. and LIJOI, A. (2010). Nonparametric priors for vectors of survival functions. *Statist. Sinica* **20** 1455–1484. MR2777332

[14] FLAMARY, R. and COURTY, N. (2017). POT Python Optimal Transport library.

[15] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[16] GINI, C. (1914). Di una misura delle relazioni tra le graduatorie di due caratteri. *Saggi Monografici del Comune di Roma*, *Tip. Cecchini*.

[17] GRIFFIN, J. E. and LEISEN, F. (2017). Compound random measures and their use in Bayesian nonparametrics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 525–545. MR3611758 https://doi.org/10.1111/rssb.12176

[18] GRIFFITHS, R. C. and MILNE, R. K. (1978). A class of bivariate Poisson processes. *J. Multivariate Anal.* **8** 380–395. MR0512608 https://doi.org/10.1016/0047-259X(78)90061-1

[19] HJORT, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294. MR1062708 https://doi.org/10.1214/aos/1176347749

[20] HJORT, N. L., HOLMES, C. C., MÜLLER, P. and WALKER, S. G. (2010). *Bayesian Nonparametrics*. *Camb. Ser. Stat. Probab. Math.* **28**. Cambridge University Press. Cambridge. MR2722987 https://doi.org/10.1017/CBO9780511802478

[21] ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729 https://doi.org/10.1198/016214501750332758

[22] ISHWARAN, H. and JAMES, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data. *J. Amer. Statist. Assoc.* **99** 175–190. MR2054297 https://doi.org/10.1198/016214504000000179

[23] JAMES, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33** 1771–1799. MR2166562 https://doi.org/10.1214/009053605000000336

[24] KALLENBERG, O. (2017). *Random Measures*, *Theory and Applications*. *Probability Theory and Stochastic Modelling* **77**. Springer, Cham. MR3642325 https://doi.org/10.1007/978-3-319-41598-7

[25] KALLSEN, J. and TANKOV, P. (2006). Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *J. Multivariate Anal.* **97** 1551–1572. MR2275419 https://doi.org/10.1016/j.jmva.2005.11.001

[26] KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. MR0210185

[27] KNOTT, M. and SMITH, C. S. (1984). On the optimal mapping of distributions. *J. Optim. Theory Appl.* **43** 39–49. MR0745785 https://doi.org/10.1007/BF00934745

[28] LAU, J. W., CRIPPS, E. and HUI, W. (2020). Variational inference for multiplicative intensity models. *Statist. Probab. Lett.* **161** 108720, 6. MR4065484 https://doi.org/10.1016/j.spl.2020.108720

[29] LEISEN, F. and LIJOI, A. (2011). Vectors of two-parameter Poisson–Dirichlet processes. *J. Multivariate Anal.* **102** 482–495. MR2755010 https://doi.org/10.1016/j.jmva.2010.10.008

[30] LEISEN, F., LIJOI, A. and SPANÓ, D. (2013). A vector of Dirichlet processes. *Electron. J. Stat.* **7** 62–90. MR3020414 https://doi.org/10.1214/12-EJS764

[31] LIJOI, A. and NIPOTI, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *J. Amer. Statist. Assoc.* **109** 802–814. MR3223751 https://doi.org/10.1080/01621459.2013.869499

[32] LIJOI, A., NIPOTI, B. and PRÜNSTER, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20** 1260–1291. MR3217444 https://doi.org/10.3150/13-BEJ521

[33] LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics*. *Camb. Ser. Stat. Probab. Math.* **28** 80–136. Cambridge Univ. Press, Cambridge. MR2730661

[34] LO, A. Y. and WENG, C.-S. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41** 227–245. MR1006487 https://doi.org/10.1007/BF00049393

[35] MACEACHERN, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association, Alexandria, VA.

[36] MACEACHERN, S. N. (2000). Dependent Dirichlet processes. Technical Report, The Ohio State Univ.

[37] MARIUCCI, E. and REISS, M. (2018). Wasserstein and total variation distance between marginals of Lévy processes. *Electron. J. Stat.* **12** 2482–2514. MR3833470 https://doi.org/10.1214/18-EJS1456

[38] MÜLLER, P., QUINTANA, F. A. and PAGE, G. (2018). Nonparametric Bayesian inference in applications (with discussion). *Stat. Methods Appl.* **27** 175–251. MR3807363 https://doi.org/10.1007/s10260-017-0405-z

[39] Müller, P., Quintana, F. A., Jara, A. and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. *Springer Series in Statistics*. Springer, Cham. MR3309338 https://doi.org/10.1007/978-3-319-18968-0

[40] Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400. MR3059422 https://doi.org/10.1214/12-AOS1065

[41] Nguyen, X. (2016). Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure. *Bernoulli* **22** 1535–1571. MR3474825 https://doi.org/10.3150/15-BEJ703

[42] Panaretos, V. M. and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

[43] Peccati, G. and Prünster, I. (2008). Linear and quadratic functionals of random hazard rates: An asymptotic analysis. *Ann. Appl. Probab.* **18** 1910–1943. MR2462554 https://doi.org/10.1214/07-AAP509

[44] Rachev, S. (1985). The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory Probab. Appl.* **29** 647–676. MR0773434

[45] Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist.* **31** 560–585. MR1983542 https://doi.org/10.1214/aos/1051027881

[46] Riva-Palacio, A. and Leisen, F. (2021). Compound vectors of subordinators and their associated positive Lévy copulas. *J. Multivariate Anal.* **183** 104728. MR4219330 https://doi.org/10.1016/j.jmva.2021.104728

[47] Rüschendorf, L. (1991). Fréchet-bounds and their applications. In *Advances in Probability Distributions with Given Marginals* (*Rome*, 1990). *Math. Appl.* **67** 151–187. Kluwer Academic, Dordrecht. MR1215951

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

[49] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

[50] Tankov, P. (2003). Dependence structure of spectrally positive multidimensional Lévy processes. Unpublished Manuscript.

[51] Villani, C. (2008). *Optimal Transport*: *Old and New*. Springer, Berlin Heidelberg.

The Annals of Statistics
2021, Vol. 49, No. 5, 2948–2971
https://doi.org/10.1214/21-AOS2066

BRIDGING CONVEX AND NONCONVEX OPTIMIZATION IN ROBUST PCA: NOISE, OUTLIERS AND MISSING DATA

BY YUXIN CHEN[1,??], JIANQING FAN[2,*], CONG MA[3,??] AND YULING YAN[2,†]

[1]*Department of Electrical and Computer Engineering, Princeton University,* [??]*yuxin.chen@princeton.edu*
[2]*Department of Operations Research and Financial Engineering, Princeton University,* [*]*jqfan@princeton.edu,* [†]*yulingy@princeton.edu*
[3]*Department of Statistics, University of Chicago,* [??]*congm@uchicago.edu*

This paper delivers improved theoretical guarantees for the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers and (3) missing data. This problem, often dubbed as *robust principal component analysis (robust PCA)*, finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis in the presence of random noise) remains highly suboptimal, which we strengthen in this paper. When the unknown matrix is well conditioned, incoherent and of constant rank, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the ℓ_∞ loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. The key analysis idea lies in bridging the convex program in use and an auxiliary nonconvex optimization algorithm, and hence the title of this paper.

REFERENCES

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Ann. Statist.* **40** 1171–1197. MR2985947 https://doi.org/10.1214/12-AOS1000

AHMED, A., RECHT, B. and ROMBERG, J. (2014). Blind deconvolution using convex programming. *IEEE Trans. Inf. Theory* **60** 1711–1732. MR3168432 https://doi.org/10.1109/TIT.2013.2294644

CAI, H., CAI, J.-F. and WEI, K. (2019). Accelerated alternating projections for robust principal component analysis. *J. Mach. Learn. Res.* **20** Paper No. 20, 33. MR3911427

CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458 https://doi.org/10.1214/13-AOS1178

CAI, J.-F. and WEI, K. (2018). Solving systems of phaseless equations via Riemannian optimization with optimal sampling complexity. Preprint. arXiv:1809.02773.

CANDÈS, E. J., LI, X. and SOLTANOLKOTABI, M. (2015). Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory* **61** 1985–2007. MR3332993 https://doi.org/10.1109/TIT.2015.2399924

CANDÈS, E. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.

CANDÈS, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5

CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061

CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** Art. 11, 37. MR2811000 https://doi.org/10.1145/1970392.1970395

CHANDRASEKARAN, V., PARRILO, P. A. and WILLSKY, A. S. (2012). Latent variable graphical model selection via convex optimization. *Ann. Statist.* **40** 1935–1967. MR3059067 https://doi.org/10.1214/11-AOS949

CHANDRASEKARAN, V., SANGHAVI, S., PARRILO, P. A. and WILLSKY, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* **21** 572–596. MR2817479 https://doi.org/10.1137/090761793

CHARISOPOULOS, V., DAVIS, D., DÍAZ, M. and DRUSVYATSKIY, D. (2019). Composite optimization for robust blind deconvolution. Preprint. arXiv::1901.01624.

MSC2020 subject classifications. Primary 62F10; secondary 62B10.
Key words and phrases. Robust principal component analysis, convex relaxation, ℓ_∞ guarantees, leave-one-out analysis.

CHEN, Y. (2015). Incoherence-optimal matrix completion. *IEEE Trans. Inf. Theory* **61** 2909–2923. MR3342311 https://doi.org/10.1109/TIT.2015.2415195

CHEN, Y. and CANDÈS, E. J. (2017). Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Comm. Pure Appl. Math.* **70** 822–883. MR3628877 https://doi.org/10.1002/cpa.21638

CHEN, Y. and CANDÈS, E. J. (2018). The projected power method: An efficient algorithm for joint alignment from pairwise differences. *Comm. Pure Appl. Math.* **71** 1648–1714. MR3847751 https://doi.org/10.1002/cpa.21760

CHEN, Y. and CHI, Y. (2014). Robust spectral compressed sensing via structured matrix completion. *IEEE Trans. Inf. Theory* **60** 6576–6601. MR3265040 https://doi.org/10.1109/TIT.2014.2343623

CHEN, Y., GUIBAS, L. J. and HUANG, Q. (2014). Near-optimal joint optimal matching via convex relaxation. In *International Conference on Machine Learning* (*ICML*) 100–108.

CHEN, J., LIU, D. and LI, X. (2020). Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. *IEEE Trans. Inform. Theory* **66** 5806–5841. MR4158648 https://doi.org/10.1109/TIT.2020.2992234

CHEN, Y. and WAINWRIGHT, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. arXiv:1509.03025.

CHEN, Y., JALALI, A., SANGHAVI, S. and CARAMANIS, C. (2013). Low-rank matrix recovery from errors and erasures. *IEEE Transactions on Information Theory* **59** 4324–4337.

CHEN, Y., JALALI, A., SANGHAVI, S. and XU, H. (2014). Clustering partially observed graphs via convex optimization. *J. Mach. Learn. Res.* **15** 2213–2238. MR3231602

CHEN, Y., CHI, Y., FAN, J. and MA, C. (2019a). Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Math. Program.* **176** 5–37. MR3960803 https://doi.org/10.1007/s10107-019-01363-6

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019b). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2020a). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM J. Optim.* **30** 3098–3121. MR4167625 https://doi.org/10.1137/19M1290000

CHEN, Y., CHI, Y., FAN, J. and MA, C. (2020b). Spectral methods for data science: A statistical perspective. Preprint. arXiv::2012.08496.

CHEN, Y., FAN, J., WANG, B. and YAN, Y. (2020c). Convex and nonconvex optimization are both minimax-optimal for noisy blind deconvolution. Preprint. arXiv::2008.01724.

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2021). Supplement to "Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data." https://doi.org/10.1214/21-AOS2066SUPP

CHERAPANAMJERI, Y., GUPTA, K. and JAIN, P. (2017). Nearly optimal robust matrix completion. In *Proceedings of the* 34*th International Conference on Machine Learning* **70** 797–805. JMLR.org.

CHI, Y., LU, Y. M. and CHEN, Y. (2019). Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Trans. Signal Process.* **67** 5239–5269. MR4016283 https://doi.org/10.1109/TSP.2019.2937282

DAVENPORT, M. A. and ROMBERG, J. (2016). An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing* **10** 608–622.

DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257

DRUSVYATSKIY, D., IOFFE, A. D. and LEWIS, A. S. (2021). Nonsmooth optimization using Taylor-like models: Error bounds, convergence, and termination criteria. *Math. Program.* **185** 357–383. MR4201717 https://doi.org/10.1007/s10107-019-01432-w

FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991 https://doi.org/10.1016/j.jeconom.2008.09.017

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., WANG, W. and ZHONG, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** Paper No. 207, 42. MR3827095

FAN, J., WANG, W. and ZHONG, Y. (2019). Robust covariance estimation for approximate factor models. *J. Econometrics* **208** 5–22. MR3906959 https://doi.org/10.1016/j.jeconom.2018.09.003

FAN, J., SUN, Q., ZHOU, W.-X. and ZHU, Z. (2018). Principal component analysis for big data. Preprint. arXiv::1801.01602.

FENG, J., XU, H. and YAN, S. (2013). Online robust PCA via stochastic optimization. In *Advances in Neural Information Processing Systems* 404–412.

GANESH, A., WRIGHT, J., LI, X., CANDES, E. J. and MA, Y. (2010). Dense error correction for low-rank matrices via principal component pursuit. In 2010 *IEEE International Symposium on Information Theory* 1513–1517. IEEE.

GOLDFARB, D., MA, S. and SCHEINBERG, K. (2013). Fast alternating linearization methods for minimizing the sum of two convex functions. *Math. Program.* **141** 349–382. MR3097290 https://doi.org/10.1007/s10107-012-0530-2

GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory* **57** 1548–1566. MR2815834 https://doi.org/10.1109/TIT.2011.2104999

GU, Q., WANG, Z. W. and LIU, H. (2016). Low-rank and sparse structure pursuit via alternating minimization. In *Artificial Intelligence and Statistics* 600–609.

GUO, H., QIU, C. and VASWANI, N. (2014). An online algorithm for separating sparse and low-dimensional signal sequences from their sum. *IEEE Trans. Signal Process.* **62** 4284–4297. MR3260427 https://doi.org/10.1109/TSP.2014.2331612

HSU, D., KAKADE, S. M. and ZHANG, T. (2011). Robust matrix decomposition with sparse corruptions. *IEEE Trans. Inform. Theory* **57** 7221–7234. MR2883652 https://doi.org/10.1109/TIT.2011.2158250

HUANG, Q. and GUIBAS, L. (2013). Consistent shape maps via semidefinite programming. *Computer Graphics Forum* **32** 177–186.

JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing* 665–674. ACM, New York. MR3210828 https://doi.org/10.1145/2488608.2488693

JOLLIFFE, I. T. (1986). *Principal Component Analysis. Springer Series in Statistics*. Springer, New York. MR0841268 https://doi.org/10.1007/978-1-4757-1904-8

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inform. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205

KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20** 282–303. MR3160583 https://doi.org/10.3150/12-BEJ486

KLOPP, O., LOUNICI, K. and TSYBAKOV, A. B. (2017). Robust matrix completion. *Probab. Theory Related Fields* **169** 523–564. MR3704775 https://doi.org/10.1007/s00440-016-0736-y

KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869 https://doi.org/10.1214/11-AOS894

KRAHMER, F. and STÖGER, D. (2021). On the convex geometry of blind deconvolution and matrix completion. *Comm. Pure Appl. Math.* **74** 790–832. MR4221934

LI, X. (2013). Compressed sensing and matrix completion with constant proportion of corruptions. *Constr. Approx.* **37** 73–99. MR3010211 https://doi.org/10.1007/s00365-012-9176-9

LI, Y., MA, C., CHEN, Y. and CHI, Y. (2019). Nonconvex matrix factorization from rank-one measurements. In *The 22nd International Conference on Artificial Intelligence and Statistics* 1496–1505.

MA, S. and AYBAT, N. S. (2018). Efficient optimization algorithms for robust principal component analysis and its variants. *Proceedings of the IEEE* **106** 1411–1426.

MA, C., WANG, K., CHI, Y. and CHEN, Y. (2020). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Found. Comput. Math.* **20** 451–632. MR4099988 https://doi.org/10.1007/s10208-019-09429-9

MAZUMDER, R., HASTIE, T. and TIBSHIRANI, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11** 2287–2322. MR2719857

NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

NETRAPALLI, P., JAIN, P. and SANGHAVI, S. (2015). Phase retrieval using alternating minimization. *IEEE Trans. Signal Process.* **63** 4814–4826. MR3385838 https://doi.org/10.1109/TSP.2015.2448516

NETRAPALLI, P., NIRANJAN, U., SANGHAVI, S., ANANDKUMAR, A. and JAIN, P. (2014). Non-convex robust PCA. In *Advances in Neural Information Processing Systems* 1107–1115.

PEARSON, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2** 559–572.

QIU, C. and VASWANI, N. (2010). Real-time robust principal components' pursuit. In 2010 *48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 591–598. IEEE.

QIU, C., VASWANI, N., LOIS, B. and HOGBEN, L. (2014). Recursive robust PCA or recursive sparse recovery in large but structured noise. *IEEE Trans. Inform. Theory* **60** 5007–5039. MR3245369 https://doi.org/10.1109/TIT.2014.2331344

SHEN, Y., WEN, Z. and ZHANG, Y. (2014). Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optim. Methods Softw.* **29** 239–263. MR3175484 https://doi.org/10.1080/10556788.2012.700713

SINGER, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.* **30** 20–36. MR2737931 https://doi.org/10.1016/j.acha.2010.02.001

SREBRO, N. and SHRAIBMAN, A. (2005). Rank, trace-norm and max-norm. In *Learning Theory. Lecture Notes in Computer Science* **3559** 545–560. Springer, Berlin. MR2203286 https://doi.org/10.1007/11503415_37

SUN, R. and LUO, Z.-Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Inform. Theory* **62** 6535–6579. MR3565131 https://doi.org/10.1109/TIT.2016.2598574

TAO, M. and YUAN, X. (2011). Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.* **21** 57–81. MR2765489 https://doi.org/10.1137/100781894

VASWANI, N. and NARAYANAMURTHY, P. (2018). Static and dynamic robust PCA and matrix completion: A review. *Proceedings of the IEEE* **106** 1359–1379.

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

WANG, G., GIANNAKIS, G. B. and ELDAR, Y. C. (2018). Solving systems of random quadratic equations via truncated amplitude flow. *IEEE Trans. Inform. Theory* **64** 773–794. MR3762591 https://doi.org/10.1109/TIT.2017.2756858

WEI, K., CAI, J.-F., CHAN, T. F. and LEUNG, S. (2016). Guarantees of Riemannian optimization for low rank matrix recovery. *SIAM J. Matrix Anal. Appl.* **37** 1198–1222. MR3543156 https://doi.org/10.1137/15M1050525

WONG, R. K. W. and LEE, T. C. M. (2017). Matrix completion with noisy entries and outliers. *J. Mach. Learn. Res.* **18** Paper No. 147, 25. MR3763781

YI, X., PARK, D., CHEN, Y. and CARAMANIS, C. (2016). Fast algorithms for robust PCA via gradient descent. In *NIPS* 4152–4160.

ZHAN, J., LOIS, B., GUO, H. and VASWANI, N. (2016). Online (and offline) robust PCA: Novel algorithms and performance guarantees. In *Artificial Intelligence and Statistics* 1488–1496.

ZHANG, H., CHI, Y. and LIANG, Y. (2016). Provable non-convex phase retrieval with outliers: Median truncated Wirtinger flow. In *International Conference on Machine Learning* 1022–1031.

ZHANG, X., WANG, L. W. and GU, Q. (2018). A unified framework for nonconvex low-rank plus sparse matrix recovery. In *International Conference on Artificial Intelligence and Statistics*.

ZHENG, Q. and LAFFERTY, J. (2016). Convergence analysis for rectangular matrix completion using Burer–Monteiro factorization and gradient descent. arXiv:1605.07051.

ZHOU, Z., LI, X., WRIGHT, J., CANDÈS, E. and MA, Y. (2010). Stable principal component pursuit. In *International Symposium on Information Theory* 1518–1522.

The Annals of Statistics
2021, Vol. 49, No. 5, 2972–2990
https://doi.org/10.1214/21-AOS2068

EFFICIENCY OF DELAYED-ACCEPTANCE RANDOM WALK METROPOLIS ALGORITHMS

BY CHRIS SHERLOCK[1,??], ALEXANDRE H. THIERY[2,??] AND ANDREW GOLIGHTLY[3,??]

[1]*Department of Mathematics and Statistics, Lancaster University,* [??]*c.sherlock@lancaster.ac.uk*
[2]*Department of Statistics and Applied Probability, National University of Singapore,* [??]*a.h.thiery@gmail.com*
[3]*School of Mathematics, Statistics and Physics, Newcastle University,* [??]*andrew.golightly@ncl.ac.uk*

Delayed-acceptance Metropolis–Hastings and delayed-acceptance pseudo-marginal Metropolis–Hastings algorithms can be applied when it is computationally expensive to calculate the true posterior or an unbiased stochastic approximation thereof, but a computationally cheap deterministic approximation is available. An initial accept–reject stage uses the cheap approximation for computing the Metropolis–Hastings ratio; proposals which are accepted at this stage are subjected to a further accept–reject step, which corrects for the error in the approximation. Since the expensive posterior, or the approximation thereof, is only evaluated for proposals which are accepted at the first stage, the cost of the algorithm is reduced and larger scalings may be used.

We focus on the random walk Metropolis (RWM) and consider the delayed-acceptance RWM and the delayed-acceptance pseudo-marginal RWM. We provide a framework for incorporating relatively general deterministic approximations into the theoretical analysis of high-dimensional targets. Justified by diffusion-approximation arguments, we derive expressions for the limiting efficiency and acceptance rates in high-dimensional settings. Finally, these theoretical insights are leveraged to formulate practical guidelines for the efficient tuning of the algorithms. The robustness of these guidelines and predicted properties are verified against simulation studies, all of which are strictly outside of the domain of validity of our limit results.

REFERENCES

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

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

[3] AU, S.-K. and BECK, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probab. Eng. Mech.* **16** 263–277. https://doi.org/10.1016/S0266-8920(01)00019-4

[4] BANTERLE, M., GRAZIAN, C., LEE, A. and ROBERT, C. P. (2019). Accelerating Metropolis–Hastings algorithms by delayed acceptance. *Foundations of Data Science* **1** 103–128.

[5] BEAUMONT, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164** 1139–1160.

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

[7] BÉDARD, M., DOUC, R. and MOULINES, E. (2012). Scaling analysis of multiple-try MCMC methods. *Stochastic Process. Appl.* **122** 758–786. MR2891436 https://doi.org/10.1016/j.spa.2011.11.004

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

MSC2020 subject classifications. 65C05, 65C40, 60F05.
Key words and phrases. Markov chain Monte Carlo, delayed-acceptance, diffusion limit, Pseudo-marginal MCMC, random-walk Metropolis.

[9] BÉRARD, J., DEL MORAL, P. and DOUCET, A. (2014). A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Probab.* **19** no. 94, 28. MR3272327 https://doi.org/10.1214/EJP.v19-3428

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

[11] BOYS, R. J., WILKINSON, D. J. and KIRKWOOD, T. B. L. (2008). Bayesian inference for a discretely observed stochastic kinetic model. *Stat. Comput.* **18** 125–135. MR2390814 https://doi.org/10.1007/s11222-007-9043-x

[12] BREYER, L. A., PICCIONI, M. and SCARLATTI, S. (2004). Optimal scaling of MaLa for nonlinear regression. *Ann. Appl. Probab.* **14** 1479–1505. MR2071431 https://doi.org/10.1214/105051604000000369

[13] BROOKS, S., GELMAN, A., JONES, G. L. and MENG, X.-L., eds. (2011) *Handbook of Markov chain Monte Carlo*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR10.1201/b10905

[14] CATANACH, T. A. and BECK, J. L. (2018). Bayesian updating and uncertainty quantification using sequential tempered MCMC with the rank-one modified Metropolis algorithm. ArXiv E-prints. Available at arXiv:1804.08738.

[15] CHRISTEN, J. A. and FOX, C. (2005). Markov chain Monte Carlo using an approximation. *J. Comput. Graph. Statist.* **14** 795–810. MR2211367 https://doi.org/10.1198/106186005X76983

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

[17] DAHLIN, J. and SCHÖN, T. B. (2019). Getting started with particle Metropolis–Hastings for inference in nonlinear dynamical models. ArXiv E-prints. Available at arXiv:1511.01707.

[18] DEL MORAL, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. *Probability and Its Applications* (*New York*). Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

[19] DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005 https://doi.org/10.1093/biomet/asu075

[20] EFENDIEV, Y., DATTA-GUPTA, A., GINTING, V., MA, X. and MALLICK, B. (2005). An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resour. Res.* **41**.

[21] EFENDIEV, Y., HOU, T. and LUO, W. (2006). Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM J. Sci. Comput.* **28** 776–803. MR2231730 https://doi.org/10.1137/050628568

[22] EVERITT, R. G. and ROWIŃSKA, P. A. (2017). Delayed acceptance ABC-SMC. ArXiv E-prints. Available at arXiv:1708.02230.

[23] FEARNHEAD, P. and SHERLOCK, C. (2006). An exact Gibbs sampler for the Markov-modulated Poisson process. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 767–784. MR2301294 https://doi.org/10.1111/j.1467-9868.2006.00566.x

[24] FILIPPONE, M. and GIROLAMI, M. (2014). Pseudo-marginal Bayesian inference for Gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 2214–2226.

[25] FLURY, T. and SHEPHARD, N. (2011). Bayesian inference based only on simulated likelihood: Particle filter analysis of dynamic economic models. *Econometric Theory* **27** 933–956. MR2843833 https://doi.org/10.1017/S0266466610000599

[26] FRANKS, J. and VIHOLA, M. (2020). Importance sampling correction versus standard averages of reversible MCMCs in terms of the asymptotic variance. *Stochastic Process. Appl.* **130** 6157–6183. MR4140030 https://doi.org/10.1016/j.spa.2020.05.006

[27] GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J., eds. (1996) *Markov Chain Monte Carlo in Practice*. *Interdisciplinary Statistics*. CRC Press, London. MR1397966 https://doi.org/10.1007/978-1-4899-4485-6

[28] GOLIGHTLY, A., HENDERSON, D. A. and SHERLOCK, C. (2015). Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Stat. Comput.* **25** 1039–1055. MR3375634 https://doi.org/10.1007/s11222-014-9469-x

[29] GOLIGHTLY, A. and WILKINSON, D. J. (2011). Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* **1** 807–820. https://doi.org/10.1098/rsfs.2011.0047

[30] HIGDON, D. C., REESE, S. J., MOULTON, D., VRUGT, J. A. and FOX, C. (2011). Posterior exploration for computationally intensive forward models. In *Handbook of Markov chain Monte Carlo*, *chapter* 16 (S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.) 401–418. CRC Press, Boca Raton, FL.

[31] KAIPIO, J. and SOMERSALO, E. (2005). *Statistical and Computational Inverse Problems*. *Applied Mathematical Sciences* **160**. Springer, New York. MR2102218
[32] KNAPE, J. and DE VALPINE, P. (2012). Fitting complex population models by combining particle filters with Markov chain Monte Carlo. *Ecology* **93** 256–263.
[33] LIU, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. Springer, New York. MR1842342
[34] MATTINGLY, J. C., PILLAI, N. S. and STUART, A. M. (2012). Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Probab.* **22** 881–930. MR2977981 https://doi.org/10.1214/10-AAP754
[35] MOULTON, J. D., FOX, C. and SVYATSKIY, D. (2008). Multilevel approximations in sample-based inversion from the Dirichlet-to-Neumann map. *J. Phys.*, *Conf. Ser.* **124**.
[36] PASARICA, C. and GELMAN, A. (2010). Adaptively scaling the Metropolis algorithm using expected squared jumped distance. *Statist. Sinica* **20** 343–364. MR2640698
[37] PITT, M. K., DOS SANTOS SILVA, R., GIORDANI, P. and KOHN, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *J. Econometrics* **171** 134–151. MR2991856 https://doi.org/10.1016/j.jeconom.2012.06.004
[38] QUIROZ, M., TRAN, M.-N., VILLANI, M. and KOHN, R. (2018). Speeding up MCMC by delayed acceptance and data subsampling. *J. Comput. Graph. Statist.* **27** 12–22. MR3788297 https://doi.org/10.1080/10618600.2017.1307117
[39] ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254
[40] ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123
[41] ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450 https://doi.org/10.1214/ss/1015346320
[42] ROBERTS, G. O. and ROSENTHAL, J. S. (2014). Complexity bounds for Markov chain Monte Carlo algorithms via diffusion limits. *J. Appl. Probab.* **53** 410–420. MR3514287
[43] ROBERTS, G. O. and ROSENTHAL, J. S. (2014). Minimising MCMC variance via diffusion limits, with an application to simulated tempering. *Ann. Appl. Probab.* **24** 131–149. MR3161644 https://doi.org/10.1214/12-AAP918
[44] SCHMON, S. M., DELIGIANNIDIS, G., DOUCET, A. and PITT, M. K. (2021). Large-sample asymptotics of the pseudo-marginal method. *Biometrika* **108** 37–51. MR4226188 https://doi.org/10.1093/biomet/asaa044
[45] SHERLOCK, C. (2013). Optimal scaling of the random walk Metropolis: General criteria for the 0.234 acceptance rule. *J. Appl. Probab.* **50** 1–15. MR3076768 https://doi.org/10.1239/jap/1363784420
[46] SHERLOCK, C. (2016). Optimal scaling for the pseudo-marginal random walk Metropolis: Insensitivity to the noise generating mechanism. *Methodol. Comput. Appl. Probab.* **18** 869–884. MR3531025 https://doi.org/10.1007/s11009-015-9471-6
[47] SHERLOCK, C., FEARNHEAD, P. and ROBERTS, G. O. (2010). The random walk Metropolis: Linking theory and practice through a case study. *Statist. Sci.* **25** 172–190. MR2789988 https://doi.org/10.1214/10-STS327
[48] SHERLOCK, C., GOLIGHTLY, A. and HENDERSON, D. A. (2017). Adaptive, delayed-acceptance MCMC for targets with expensive likelihoods. *J. Comput. Graph. Statist.* **26** 434–444. MR3640199 https://doi.org/10.1080/10618600.2016.1231064
[49] SHERLOCK, C. and LEE, A. (2017). Variance bounding of delayed-acceptance kernels. ArXiv E-prints. Available at arXiv:1706.02142.
[50] SHERLOCK, C. and ROBERTS, G. (2009). Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli* **15** 774–798. MR2555199 https://doi.org/10.3150/08-BEJ176
[51] SHERLOCK, C., THIERY, A. H and GOLIGHTLY, A. (2021). Supplement to "Efficiency of delayed-acceptance random walk Metropolis algorithms." https://doi.org/10.1214/21-AOS2068SUPP
[52] SHERLOCK, C., THIERY, A. H. and LEE, A. (2017). Pseudo-marginal Metropolis–Hastings sampling using averages of unbiased estimators. *Biometrika* **104** 727–734. MR3694593 https://doi.org/10.1093/biomet/asx031
[53] SHERLOCK, C., THIERY, A. H., ROBERTS, G. O. and ROSENTHAL, J. S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606 https://doi.org/10.1214/14-AOS1278
[54] SMITH, M. E. (2011). Estimating nonlinear economic models using surrogate transitions. Available from https://files.nyu.edu/mes473/public/Smith_Surrogate.pdf.

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

[56] VIHOLA, M., HELSKE, J. and FRANKS, J. (2016). Importance sampling type estimators based on approximate marginal MCMC. Availabe at https://doi.org/10.1111/sjos.12492.

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

[58] ZANELLA, G., BÉDARD, M. and KENDALL, W. S. (2017). A Dirichlet form approach to MCMC optimal scaling. *Stochastic Process*. *Appl*. **127** 4053–4082. MR3718106 https://doi.org/10.1016/j.spa.2017.03.021

The Annals of Statistics
2021, Vol. 49, No. 5, 2991–3014
https://doi.org/10.1214/21-AOS2070

SEMIPARAMETRIC OPTIMAL ESTIMATION WITH NONIGNORABLE NONRESPONSE DATA

BY KOSUKE MORIKAWA[1,??] AND JAE KWANG KIM[2,??]

[1]*Graduate School of Engineering Science, Osaka University,* [??]*morikawa@sigmath.es.osaka-u.ac.jp*
[2]*Department of Statistics, Iowa State University,* [??]*jkim@iastate.edu*

When the response mechanism is believed to be not missing at random (NMAR), a valid analysis requires stronger assumptions on the response mechanism than standard statistical methods would otherwise require. Semiparametric estimators have been developed under the parametric model assumptions on the response mechanism. In this paper, a new statistical test is proposed to guarantee model identifiability without using instrumental variable assumption. Furthermore, we develop optimal semiparametric estimation for parameters such as the population mean. Specifically, we propose two semiparametric optimal estimators that do not require any model assumptions other than the response mechanism. Asymptotic properties of the proposed estimators are discussed. An extensive simulation study is presented to compare with some existing methods. We present an application of our method using Korean labor and income panel survey data.

REFERENCES

[1] BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

[2] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. MR1623559

[3] CHANG, T. and KOTT, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* **95** 555–571. MR2443175 https://doi.org/10.1093/biomet/asn022

[4] CHEN, S. and HAZIZA, D. (2017). Multiply robust imputation procedures for the treatment of item nonresponse in surveys. *Biometrika* **104** 439–453. MR3698264 https://doi.org/10.1093/biomet/asx007

[5] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

[6] D'HAULTFOEUILLE, X. (2010). A new instrumental method for dealing with endogenous selection. *J. Econometrics* **154** 1–15. MR2558947 https://doi.org/10.1016/j.jeconom.2009.06.005

[7] DIGGLE, P. and KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **43** 49–93.

[8] DOMÍNGUEZ, M. A. and LOBATO, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72** 1601–1615. MR2078215 https://doi.org/10.1111/j.1468-0262.2004.00545.x

[9] EUBANK, R. L. and HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20** 1412–1425. MR1186256 https://doi.org/10.1214/aos/1176348775

[10] FITZMAURICE, G. M., LIPSITZ, S. R., MOLENBERGHS, G. and IBRAHIM, J. G. (2005). A protective estimator for longitudinal binary data subject to non-ignorable non-monotone missingness. *J. Roy. Statist. Soc. Ser. A* **168** 723–735. MR2205403 https://doi.org/10.1111/j.1467-985X.2005.00374.x

[11] GREENLEES, J. S., REECE, W. S. and ZIESCHANG, K. D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *J. Amer. Statist. Assoc.* **77** 251–261.

[12] GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2008). A kernel statistical test of independence. *NeurIPS* **20** 585–592.

[13] GRETTON, A., HERBRICH, R., SMOLA, A., BOUSQUET, O. and SCHÖLKOPF, B. (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6** 2075–2129. MR2249882

MSC2020 subject classifications. Primary 62F35, 62G20; secondary 62G10.
Key words and phrases. Estimating functions, identification, incomplete data, not missing at random (NMAR), semiparametric efficient estimation.

[14] HAN, P. (2014). Multiply robust estimation in regression analysis with missing data. *J. Amer. Statist. Assoc.* **109** 1159–1173. MR3265688 https://doi.org/10.1080/01621459.2014.880058

[15] KANAMORI, T., SUZUKI, T. and SUGIYAMA, M. (2012). Statistical analysis of kernel-based least-squares density-ratio estimation. *Mach. Learn.* **86** 335–367. MR2897527 https://doi.org/10.1007/s10994-011-5266-3

[16] KIM, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika* **98** 119–132. MR2804214 https://doi.org/10.1093/biomet/asq073

[17] KIM, J. K. and YU, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *J. Amer. Statist. Assoc.* **106** 157–165. MR2816710 https://doi.org/10.1198/jasa.2011.tm10104

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

[19] LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR1925014 https://doi.org/10.1002/9781119013563

[20] LIU, C. (2004). Robit regression: A simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. *Wiley Ser. Probab. Stat.* 227–238. Wiley, Chichester. MR2138259 https://doi.org/10.1002/0470090456.ch21

[21] LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44** 226–233. MR0676213

[22] MA, W.-Q., GENG, Z. and HU, Y.-H. (2003). Identification of graphical models for nonignorable non-response of binary outcomes in longitudinal studies. *J. Multivariate Anal.* **87** 24–45. MR2007260 https://doi.org/10.1016/S0047-259X(03)00043-5

[23] MIAO, W., DING, P. and GENG, Z. (2016). Identifiability of normal and normal mixture models with nonignorable missing data. *J. Amer. Statist. Assoc.* **111** 1673–1683. MR3601726 https://doi.org/10.1080/01621459.2015.1105808

[24] MOLENBERGHS, G., BEUNCKENS, C., SOTTO, C. and KENWARD, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 371–388. MR2424758 https://doi.org/10.1111/j.1467-9868.2007.00640.x

[25] MORIKAWA, K., KIM, J. K. and KANO, Y. (2017). Semiparametric maximum likelihood estimation with data missing not at random. *Canad. J. Statist.* **45** 393–409. MR3729977 https://doi.org/10.1002/cjs.11340

[26] QIN, J., LEUNG, D. and SHAO, J. (2002). Estimation with survey data under nonignorable nonresponse or informative sampling. *J. Amer. Statist. Assoc.* **97** 193–200. MR1947279 https://doi.org/10.1198/016214502753479338

[27] RIDDLES, M. K., KIM, J. K. and IM, J. (2016). Propensity-score-adjustment method for nonignorable nonresponse. *J. Surv. Stat. Methodol.* **97** 215–245.

[28] ROBINS, J. M., ROTNITZKY, A. and SCHARFSTEIN, D. O. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (*Minneapolis*, *MN*, 1997). *IMA Vol. Math. Appl.* **116** 1–94. Springer, New York. MR1731681 https://doi.org/10.1007/978-1-4612-1284-3_1

[29] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

[30] ROTNITZKY, A. and ROBINS, J. M. (1997). Analysis of semi-parametric regression models with non-ignorable non-response. *Stat. Med.* **16** 81–102.

[31] ROTNITZKY, A., SCHARFSTEIN, D., SU, T.-L. and ROBINS, J. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57** 103–113. MR1833295 https://doi.org/10.1111/j.0006-341X.2001.00103.x

[32] RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

[33] SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. MR1731478 https://doi.org/10.2307/2669923

[34] SEN, A. and SEN, B. (2014). Testing independence and goodness-of-fit in linear models. *Biometrika* **101** 927–942. MR3286926 https://doi.org/10.1093/biomet/asu026

[35] SHAO, J. and WANG, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika* **103** 175–187. MR3465829 https://doi.org/10.1093/biomet/asv071

[36] SKRONDAL, A. and RABE-HESKETH, S. (2014). Protective estimation of mixed-effects logistic regression when data are not missing at random. *Biometrika* **101** 175–188. MR3180664 https://doi.org/10.1093/biomet/ast054

[37] SVERCHKOV, M. (2014). A new approach to estimation of response probabilities when missing data are not missing at random. In *Proc*. *Survey Res*. *Meth*. *Sect*., *Am*. *Statist*. *Ass*. 867–874. Amer. Statist. Assoc., Alexandria, VA.

[38] TANG, G., LITTLE, R. J. A. and RAGHUNATHAN, T. E. (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika* **90** 747–764. MR2024755 https://doi.org/10.1093/biomet/90.4.747

[39] TANG, N., ZHAO, P. and ZHU, H. (2014). Empirical likelihood for estimating equations with nonignorably missing data. *Statist*. *Sinica* **24** 723–747. MR3235396

[40] TERAN HIDALGO, S. J., WU, M. C., ENGEL, S. M. and KOSOROK, M. R. (2018). Goodness-of-fit test for nonparametric regression models: Smoothing spline ANOVA models as example. *Comput*. *Statist*. *Data Anal*. **122** 135–155. MR3765820 https://doi.org/10.1016/j.csda.2018.01.004

[41] TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. *Springer Series in Statistics*. Springer, New York. MR2233926

[42] VERBEKE, G., MOLENBERGHS, G., THIJS, H., LESAFFRE, E. and KENWARD, M. G. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics* **57** 7–14. MR1833286 https://doi.org/10.1111/j.0006-341X.2001.00007.x

[43] WANG, S., SHAO, J. and KIM, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statist*. *Sinica* **24** 1097–1116. MR3241279

[44] ZHAO, J. and SHAO, J. (2015). Semiparametric pseudo-likelihoods in generalized linear models with nonignorable missing data. *J*. *Amer*. *Statist*. *Assoc*. **110** 1577–1590. MR3449056 https://doi.org/10.1080/01621459.2014.983234

[45] ZHAO, P., TANG, N., QU, A. and JIANG, D. (2017). Semiparametric estimating equations inference with nonignorable missing data. *Statist*. *Sinica* **27** 89–113. MR3618161

The Annals of Statistics
2021, Vol. 49, No. 5, 3015–3042
https://doi.org/10.1214/21-AOS2071

TWO-LEVEL PARALLEL FLATS DESIGNS

BY CHUNYAN WANG[1,??] AND ROBERT W. MEE[2,??]

[1] *School of Statistics and Data Science, LPMC & KLMDASR, Nankai University,* [??] *2120150092@mail.nankai.edu.cn*
[2] *Department of Business Analytics and Statistics, University of Tennessee,* [??] *rmee@utk.edu*

Regular 2^{n-p} designs are also known as single flat designs. Parallel flats designs (PFDs) consisting of three parallel flats (3-PFDs) are the most frequently utilized PFDs, due to their simple structure. Generalizing to f-PFD with $f > 3$ is more challenging. This paper aims to study the general theory for the f-PFD for any $f \geq 3$. We propose a method for obtaining the confounding frequency vectors for all nonequivalent f-PFDs, and to find the least G-aberration (or highest D-efficiency) f-PFD constructed from any single flat. PFDs are particularly useful for constructing nonregular fraction, split-plot or randomized block designs. We also characterize the quaternary code design series as PFDs. Finally, we show how designs constructed by concatenating regular fractions from different families may also have a parallel flats structure. Examples are given throughout to illustrate the results.

REFERENCES

[1] ADDELMAN, S. (1961). Irregular fractions of the 2^n factorial experiments. *Technometrics* **3** 479–496. MR0132663 https://doi.org/10.2307/1266555
[2] ADDELMAN, S. (1964). Some two-level factorial plans with split plot confounding. *Technometrics* **6** 253–258.
[3] BINGHAM, D. and SITTER, R. R. (1999). Minimum-aberration two-level fractional factorial split-plot designs. *Technometrics* **41** 62–70.
[4] BLOCK, R. M. and MEE, R. W. (1999). Second order saturated resolution IV designs. *J. Stat. Theory Appl.* **2** 96–112.
[5] BOX, G. and TYSSEDAL, J. (1996). Projective properties of certain orthogonal arrays. *Biometrika* **83** 950–955. MR1440059 https://doi.org/10.1093/biomet/83.4.950
[6] BUTLER, N. A. (2003). Minimum aberration construction results for nonregular two-level fractional factorial designs. *Biometrika* **90** 891–898. MR2024764 https://doi.org/10.1093/biomet/90.4.891
[7] CHAI, F. S. and LIAO, C. T. (2001). Three parallel flats designs for two-level factorial experiments. *Aust. N. Z. J. Stat.* **43** 113–119. MR1837500 https://doi.org/10.1111/1467-842X.00159
[8] CHEN, J., SUN, D. X. and WU, C. F. J. (1993). A catalogue of two-level and three-level fractional factorial designs with small runs. *Int. Stat. Rev.* **61** 135–145.
[9] CHENG, S.-W., LI, W. and YE, K. Q. (2004). Blocked nonregular two-level factorial designs. *Technometrics* **46** 269–279. MR2082497 https://doi.org/10.1198/004017004000000301
[10] CONNOR, W. S. and YOUNG, S. (1961). *Fractional Factorial Designs for Experiments with Factors at Two and Three Levels. National Bureau of Standards Applied Mathematics Series* **58**. National Bureau of Standards, Gaithersburg.
[11] DENG, L.-Y. and TANG, B. (1999). Generalized resolution and minimum aberration criteria for Plackett–Burman and other nonregular factorial designs. *Statist. Sinica* **9** 1071–1082. MR1744824
[12] EDWARDS, D. J. and MEE, R. W. (2021). Structure of nonregular two-level designs. *J. Amer. Statist. Assoc.* to appear. https://doi.org/10.1080/01621459.2021.1984927
[13] FRIES, A. and HUNTER, W. G. (1996). Minimum aberration 2^{k-p} designs. *Technometrics* **22** 601–608.
[14] HEDAYAT, A. S., SLOANE, N. J. A. and STUFKEN, J. (1999). *Orthogonal Arrays. Springer Series in Statistics*. Springer, New York. Theory and applications, With a foreword by C. R. Rao. MR1693498 https://doi.org/10.1007/978-1-4612-1478-6
[15] JOHN, P. W. M. (1962). Three-quarter replicates of 2^4 and 2^5 designs. *Biometrics* **17** 319–321.

MSC2020 subject classifications. Primary 62K15; secondary 62K05.
Key words and phrases. Confounding frequency vector, equivalent design, G-aberration, nonregular fractional factorial design, quaternary code, Sylvester Hadamard matrix, two-factor interaction model.

[16] JOHN, P. W. M. (1969). Some non-orthogonal fractions of 2^n designs. *J. Roy. Statist. Soc. Ser. B* **31** 270–275. MR0269041

[17] JONES, B. and GOOS, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **56** 347–364. MR2370994 https://doi.org/10.1111/j.1467-9876.2007.00581.x

[18] JONES, B., LEKIVETZ, R., MAJUMDAR, D., NACHTSHEIM, C. J. and STALLRICH, J. W. (2020). Construction, properties, and analysis of group-orthogonal supersaturated designs. *Technometrics* **62** 403–414. MR4125505 https://doi.org/10.1080/00401706.2019.1654926

[19] LIAO, C. T., IYER, H. K. and VECCHIA, D. F. (1996). Construction of orthogonal two-level designs of user-specified resolution where $N \neq 2^k$. *Technometrics* **38** 342–353. MR1422001 https://doi.org/10.2307/1271305

[20] MEE, R. W. (2004). Efficient two-level designs for estimating all main effects and two-factor interactions. *J. Qual. Technol.* **36** 400–412.

[21] MEE, R. W. (2009). *A Comprehensive Guide to Factorial Two-Level Experimentation*. Springer, New York.

[22] MEE, R. W. (2013). Tips for analyzing nonregular fractional factorial experiments. *J. Qual. Technol.* **45** 330–349.

[23] PAJAK, T. F. (1972). *Minimum Full Sequences of Two-Level Fractional Factorial Plans*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—State University of New York at Buffalo. MR2622593

[24] PAJAK, T. F. and ADDELMAN, S. (1975). Minimum full sequences of 2^{n-m} resolution III plans. *J. Roy. Statist. Soc. Ser. B* **37** 88–95. MR0381177

[25] PHOA, F. K. H. (2012). A code arithmetic approach for quaternary code designs and its application to (1/64)th-fractions. *Ann. Statist.* **40** 3161–3175. MR3097973 https://doi.org/10.1214/12-AOS1069

[26] PHOA, F. K. H., MUKERJEE, R. and XU, H. (2012). One-eighth- and one-sixteenth-fraction quaternary code designs with high resolution. *J. Statist. Plann. Inference* **142** 1073–1080. MR2879752 https://doi.org/10.1016/j.jspi.2011.11.012

[27] PHOA, F. K. H. and XU, H. (2009). Quarter-fraction factorial designs constructed via quaternary codes. *Ann. Statist.* **37** 2561–2581. MR2543703 https://doi.org/10.1214/08-AOS656

[28] SRIVASTAVA, J. N. and CHOPRA, D. A. (1973). Balanced arrays and orthogonal arrays. In *A Survey of Combinatorial Theory* (J. N. Srivastava, ed.), Amsterdam, North-Holland.

[29] SRIVASTAVA, J. N. and LI, J. (1996). Orthogonal designs of parallel flats type. *J. Statist. Plann. Inference* **53** 261–283. MR1412090 https://doi.org/10.1016/0378-3758(95)00145-X

[30] TANG, B. and DENG, L.-Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Ann. Statist.* **27** 1914–1926. MR1765622 https://doi.org/10.1214/aos/1017939244

[31] TSAI, S.-F. and LIAO, C.-T. (2018). D-optimal two-level parallel-flats block designs with partial replication. *J. Statist. Plann. Inference* **193** 89–102. MR3713465 https://doi.org/10.1016/j.jspi.2017.08.004

[32] VAZQUEZ, A. R. and XU, H. (2019). Construction of two-level nonregular designs of strength three with large run sizes. *Technometrics* **61** 341–353. MR3987624 https://doi.org/10.1080/00401706.2018.1513381

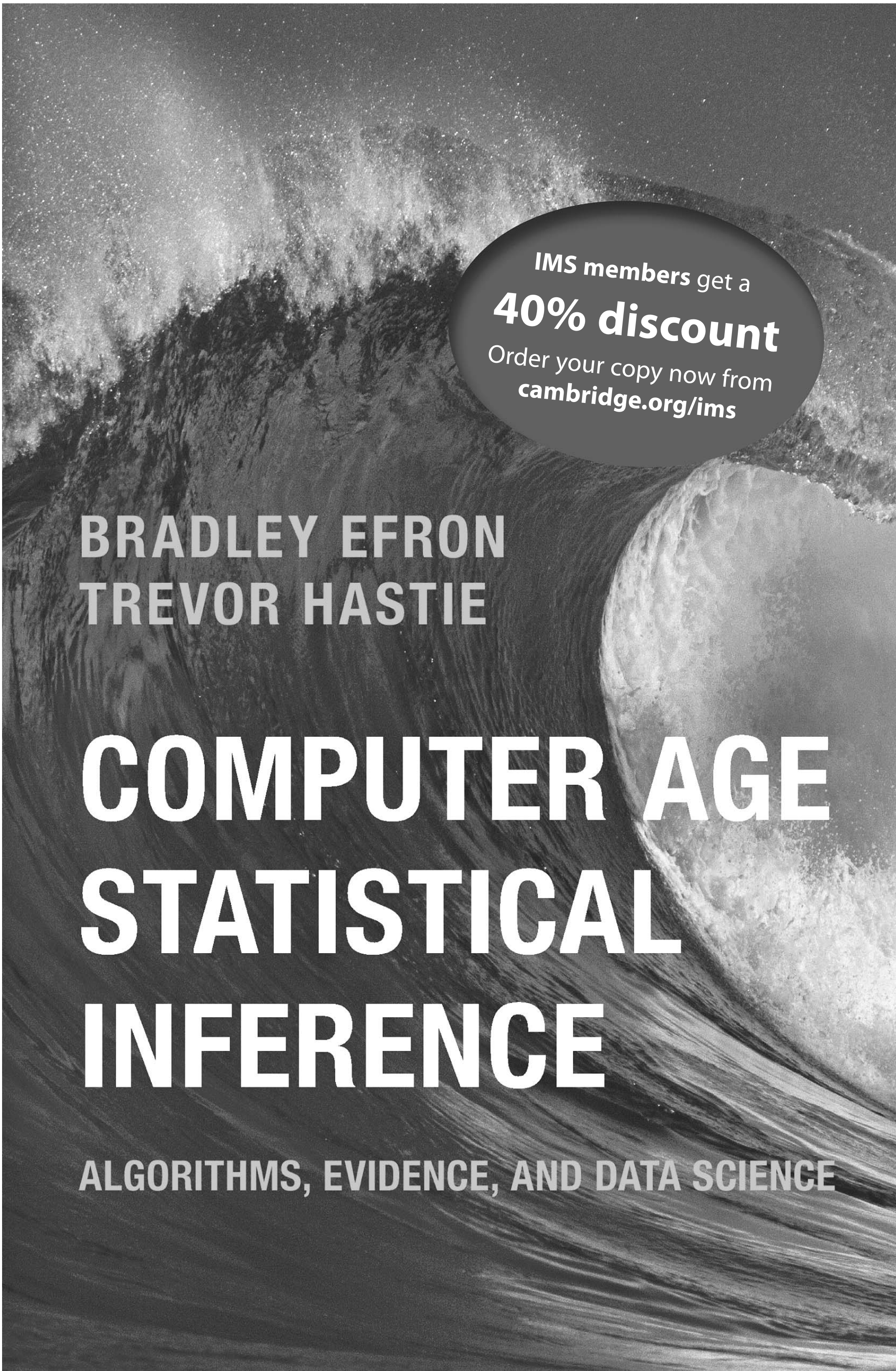
[33] WANG, C. and MEE, R. W (2021). Supplement to "Two-level parallel flats designs." https://doi.org/10.1214/21-AOS2071SUPP

[34] WOODING, W. M. (1973). The split-plot design. *J. Qual. Technol.* **5** 16–33.

[35] XU, H. (2005). Some nonregular designs from the Nordstrom–Robinson code and their statistical properties. *Biometrika* **92** 385–397. MR2201366 https://doi.org/10.1093/biomet/92.2.385

[36] XU, H. and WONG, A. (2007). Two-level nonregular designs from quaternary linear codes. *Statist. Sinica* **17** 1191–1213. MR2397390

[37] ZHANG, R., PHOA, F. K. H., MUKERJEE, R. and XU, H. (2011). A trigonometric approach to quaternary code designs with application to one-eighth and one-sixteenth fractions. *Ann. Statist.* **39** 931–955. MR2816343 https://doi.org/10.1214/10-AOS815



IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE