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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

Vol. 48, No. 1—February 2020

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. 48, No. 1—February 2020

THE ANNALS OF STATISTICS Vol. 48, No. 1, pp. 1–627 February 2020

***The Annals of Statistics* [ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 48, Number 1, February 2020. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, 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, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

The Annals of Statistics
2020, Vol. 48, No. 1, 1–26
https://doi.org/10.1214/18-AOS1788

TWO-STEP SEMIPARAMETRIC EMPIRICAL LIKELIHOOD INFERENCE

BY FRANCESCO BRAVO[1], JUAN CARLOS ESCANCIANO[2] AND INGRID VAN KEILEGOM[3]

[1]*Department of Economics, University of York, francesco.bravo@york.ac.uk*
[2]*Department of Economics, Universidad Carlos III de Madrid, jescanci@eco.uc3m.es*
[3]*ORSTAT, KU Leuven, ingrid.vankeilegom@kuleuven.be*

In both parametric and certain nonparametric statistical models, the empirical likelihood ratio satisfies a nonparametric version of Wilks' theorem. For many semiparametric models, however, the commonly used two-step (plug-in) empirical likelihood ratio is not asymptotically distribution-free, that is, its asymptotic distribution contains unknown quantities, and hence Wilks' theorem breaks down. This article suggests a general approach to restore Wilks' phenomenon in two-step semiparametric empirical likelihood inferences. The main insight consists in using as the moment function in the estimating equation the influence function of the plug-in sample moment. The proposed method is general; it leads to a chi-squared limiting distribution with known degrees of freedom; it is efficient; it does not require undersmoothing; and it is less sensitive to the first-step than alternative methods, which is particularly appealing for high-dimensional settings. Several examples and simulation studies illustrate the general applicability of the procedure and its excellent finite sample performance relative to competing methods.

REFERENCES

[1] ACKERBERG, D., CHEN, X., HAHN, J. and LIAO, Z. (2014). Asymptotic efficiency of semiparametric two-step GMM. *Rev. Econ. Stud.* **81** 919–943. MR3338225 https://doi.org/10.1093/restud/rdu011
[2] BANERJEE, M. (2013). Current status data in the twenty-first century: Some interesting developments. In *Interval-Censored Time-to-Event Data. Chapman & Hall/CRC Biostat. Ser.* 45–90. CRC Press, Boca Raton, FL. MR3014204
[3] BERAN, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.
[4] BERTAIL, P. (2006). Empirical likelihood in some semiparametric models. *Bernoulli* **12** 299–331. MR2218557 https://doi.org/10.3150/bj/1145993976
[5] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941
[6] BICKEL, P. J., RITOV, Y. and STOKER, T. M. (2006). Tailor-made tests for goodness of fit to semiparametric hypotheses. *Ann. Statist.* **34** 721–741. MR2281882 https://doi.org/10.1214/009053606000000137
[7] BRAVO, F. (2004). Empirical likelihood based inference with applications to some econometric models. *Econometric Theory* **20** 231–264. MR2044271 https://doi.org/10.1017/S0266466604202018
[8] BRAVO, F. (2018). Second order asymptotics for nonparametric conditional moment restrictions. Working paper.
[9] BRAVO, F., CHU, B. M. and JACHO-CHÁVEZ, D. T. (2017). Semiparametric estimation of moment condition models with weakly dependent data. *J. Nonparametr. Stat.* **29** 108–136. MR3597220 https://doi.org/10.1080/10485252.2016.1254781
[10] BRAVO, F., ESCANCIANO, J. C. and VAN KEILEGOM, I. (2020). Supplement to "Two-step semiparametric empirical likelihood inference." https://doi.org/10.1214/18-AOS1788SUPP.
[11] CHEN, S. X. and CUI, H. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* **93** 215–220. MR2277752 https://doi.org/10.1093/biomet/93.1.215
[12] CHEN, S. X. and CUI, H. (2007). On the second-order properties of empirical likelihood with moment restrictions. *J. Econometrics* **141** 492–516. MR2413478 https://doi.org/10.1016/j.jeconom.2006.10.006

MSC2010 subject classifications. Primary 62M10; secondary 62G10.
Key words and phrases. Empirical likelihood, semiparametric inference, high-dimensional parameters, Wilks' phenomenon.

[13] CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18** 415–447. MR2566404 https://doi.org/10.1007/s11749-009-0159-5

[14] CHEN, X., HONG, H. and TAROZZI, A. (2008). Semiparametric efficiency in GMM models with auxiliary data. *Ann. Statist.* **36** 808–843. MR2396816 https://doi.org/10.1214/009053607000000947

[15] CHEN, X., LINTON, O. and VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71** 1591–1608. MR2000259 https://doi.org/10.1111/1468-0262.00461

[16] CHERNOZHUKOV, V., ESCANCIANO, J. C., NEWEY, W. K., ICHIMURA, H. and ROBINS, J. (2017). Locally Robust Semiparametric Estimation. Working paper.

[17] CHOI, S., HALL, W. J. and SCHICK, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841–861. MR1394992 https://doi.org/10.1214/aos/1032894469

[18] DICICCIO, T., HALL, P. and ROMANO, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19** 1053–1061. MR1105861 https://doi.org/10.1214/aos/1176348137

[19] DICICCIO, T. J. and ROMANO, J. P. (1989). On adjustments based on the signed root of the empirical likelihood ratio statistic. *Biometrika* **76** 447–456. MR1040639 https://doi.org/10.1093/biomet/76.3.447

[20] DU, Y. and AKRITAS, M. G. (2002). Uniform strong representation of the conditional Kaplan–Meier process. *Math. Methods Statist.* **11** 152–182. MR1941314

[21] ESCANCIANO, J. C., JACHO-CHÁVEZ, D. T. and LEWBEL, A. (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *J. Econometrics* **178** 426–443. MR3132442 https://doi.org/10.1016/j.jeconom.2013.06.004

[22] FAN, J. and ZHANG, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Ann. Statist.* **32** 1858–1907. MR2102496 https://doi.org/10.1214/009053604000000210

[23] GOLDSTEIN, L. and MESSER, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *Ann. Statist.* **20** 1306–1328. MR1186251 https://doi.org/10.1214/aos/1176348770

[24] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5

[25] HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. thesis, Univ. California, Berkeley.

[26] HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657

[27] HE, S., LIANG, W., SHEN, J. and YANG, G. (2016). Empirical likelihood for right censored lifetime data. *J. Amer. Statist. Assoc.* **111** 646–655. MR3538694 https://doi.org/10.1080/01621459.2015.1024058

[28] HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826 https://doi.org/10.1111/1468-0262.00442

[29] HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068 https://doi.org/10.1214/07-AOS555

[30] HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568. MR1394975 https://doi.org/10.1214/aos/1032894452

[31] HUANG, J. and WELLNER, J. A. (1997). *Interval Censored Survival Data*: *A Review of Recent Progress* (D. Lin and T. Fleming, eds.). Springer, New York.

[32] ICHIMURA, H. and LINTON, O. (2005). Asymptotic expansions for some semiparametric program evaluation estimators. In *Identification and Inference for Econometric Models* 149–170. Cambridge Univ. Press, Cambridge. MR2232143 https://doi.org/10.1017/CBO9780511614491.009

[33] ICHIMURA, H. and NEWEY, W. (2017). The Influence Function of Semiparametric Estimators. CEMMAP working paper.

[34] INGLOT, T. and LEDWINA, T. (2006). Data-driven score tests for homoscedastic linear regression model: Asymptotic results. *Probab. Math. Statist.* **26** 41–61. MR2301887

[35] JEWELL, N. P. and VAN DER LAAN, M. (2004). Current status data: Review, recent developments and open problems. In *Advances in Survival Analysis*. *Handbook of Statist.* **23** 625–642. Elsevier, Amsterdam. MR2065792 https://doi.org/10.1016/S0169-7161(03)23035-2

[36] KOŠEVNIK, JU. A. and LEVIT, B. J. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753.

[37] LENG, C. and TONG, X. (2013). A quantile regression estimator for censored data. *Bernoulli* **19** 344–361. MR3019498 https://doi.org/10.3150/11-BEJ388

[38] LEPAGE, Y. (1973). A maximin test for means. *Stat. Neerl.* **27** 55–57. MR0373161 https://doi.org/10.1111/j.1467-9574.1973.tb00208.x

[39] LEWBEL, A., MCFADDEN, D. and LINTON, O. (2011). Estimating features of a distribution from binomial data. *J. Econometrics* **162** 170–188. MR2795611 https://doi.org/10.1016/j.jeconom.2010.11.006

[40] LI, G., LIN, L. and ZHU, L. (2012). Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters. *J. Multivariate Anal.* **105** 85–111. MR2877505 https://doi.org/10.1016/j.jmva.2011.08.010

[41] LI, Q., RACINE, J. and WOOLDRIDGE, J. (2008). Estimating average treatment effects with continuous and discrete covariates: The case of Swan–Ganz catherization. *Am. Econ. Rev. Pap. Proc.* **98** 357–362.

[42] MATSUSHITA, Y. and OTSU, T. (2016). Likelihood inference on semiparametric models with generated regressors. LSE working paper.

[43] MATSUSHITA, Y. and OTSU, T. (2018). Likelihood inference on semiparametric models: Average derivative and treatment effect. *Jpn. Econ. Rev.* **69** 133–155. MR3803072

[44] NEUMEYER, N. and VAN KEILEGOM, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.* **101** 1067–1078. MR2595293 https://doi.org/10.1016/j.jmva.2010.01.007

[45] NEWEY, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* **10** 233–253. MR1293201 https://doi.org/10.1017/S0266466600008409

[46] NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382. MR1303237 https://doi.org/10.2307/2951752

[47] NEWEY, W. K., HSIEH, F. and ROBINS, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica* **72** 947–962. MR2051442 https://doi.org/10.1111/j.1468-0262.2004.00518.x

[48] OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120. MR1041387 https://doi.org/10.1214/aos/1176347494

[49] OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049 https://doi.org/10.1093/biomet/75.2.237

[50] OWEN, A. B. (2001). *Empirical Likelihood*. CRC Press, London.

[51] PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory*. *Lecture Notes in Statistics* **13**. Springer, New York. MR0675954

[52] POWELL, J. L., STOCK, J. H. and STOKER, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57** 1403–1430. MR1035117 https://doi.org/10.2307/1913713

[53] QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085 https://doi.org/10.1214/aos/1176325370

[54] RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley, New York. MR0338013

[55] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

[56] RUBIN, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.

[57] SHAO, J. and SITTER, R. R. (1996). Bootstrap for imputed survey data. *J. Amer. Statist. Assoc.* **91** 1278–1288. MR1424624 https://doi.org/10.2307/2291746

[58] STRASSER, H. (1985). *Mathematical Theory of Statistics*. *De Gruyter Studies in Mathematics* **7**. de Gruyter, Berlin. MR0812467 https://doi.org/10.1515/9783110850826

[59] SUN, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. *Statistics for Biology and Health*. Springer, New York. MR2287318

[60] TANG, C. Y. and QIN, Y. (2012). An efficient empirical likelihood approach for estimating equations with missing data. *Biometrika* **99** 1001–1007. MR2999176 https://doi.org/10.1093/biomet/ass045

[61] TANG, X., LI, J. and LIAN, H. (2013). Empirical likelihood for partially linear proportional hazards models with growing dimensions. *J. Multivariate Anal.* **121** 22–32. MR3090466 https://doi.org/10.1016/j.jmva.2013.06.002

[62] VAN DER VAART, A. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–204. MR1091845 https://doi.org/10.1214/aos/1176347976

[63] 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

[64] 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

[65] WANG, D. and CHEN, S. X. (2009). Empirical likelihood for estimating equations with missing values. *Ann. Statist.* **37** 490–517. MR2488360 https://doi.org/10.1214/07-AOS585

[66] WANG, Q., LINTON, O. and HÄRDLE, W. (2004). Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.* **99** 334–345. MR2062820 https://doi.org/10.1198/016214504000000449
[67] WANG, Q. and RAO, J. N. K. (2001). Empirical likelihood for linear regression models under imputation for missing responses. *Canad. J. Statist.* **29** 597–608. MR1888507 https://doi.org/10.2307/3316009
[68] WANG, Q. and RAO, J. N. K. (2002). Empirical likelihood-based inference in linear models with missing data. *Scand. J. Stat.* **29** 563–576. MR1925575 https://doi.org/10.1111/1467-9469.00306
[69] WANG, Q. and RAO, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.* **30** 896–924. MR1922545 https://doi.org/10.1214/aos/1028674845
[70] WANG, Q.-H. and JING, B.-Y. (2003). Empirical likelihood for partial linear models. *Ann. Inst. Statist. Math.* **55** 585–595. MR2007800 https://doi.org/10.1007/BF02517809
[71] WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9** 60–62.
[72] WU, T. T., LI, G. and TANG, C. (2015). Empirical likelihood for censored linear regression and variable selection. *Scand. J. Stat.* **42** 798–812. MR3391693 https://doi.org/10.1111/sjos.12137
[73] XUE, L. (2009). Empirical likelihood confidence intervals for response mean with data missing at random. *Scand. J. Stat.* **36** 671–685. MR2573302 https://doi.org/10.1111/j.1467-9469.2009.00651.x
[74] XUE, L. and WANG, Q. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli* **18** 836–856. MR2948904 https://doi.org/10.3150/11-BEJ365
[75] XUE, L. and XUE, D. (2011). Empirical likelihood for semiparametric regression model with missing response data. *J. Multivariate Anal.* **102** 723–740. MR2772332 https://doi.org/10.1016/j.jmva.2010.11.001
[76] XUE, L. and ZHU, L. (2007). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika* **94** 921–937. MR2416799 https://doi.org/10.1093/biomet/asm066
[77] XUE, L. and ZHU, L. (2012). Empirical likelihood in some nonparametric and semiparametric models. *Stat. Interface* **5** 367–378. MR2979216 https://doi.org/10.4310/SII.2012.v5.n3.a9
[78] XUE, L.-G. and ZHU, L. (2006). Empirical likelihood for single-index models. *J. Multivariate Anal.* **97** 1295–1312. MR2279674 https://doi.org/10.1016/j.jmva.2005.09.004
[79] YANG, G., CUI, X. and HOU, S. (2017). Empirical likelihood confidence regions in the single-index model with growing dimensions. *Comm. Statist. Theory Methods* **46** 7562–7579. MR3648584 https://doi.org/10.1080/03610926.2016.1157190
[80] ZHENG, M., ZHAO, Z. and YU, W. (2012). Empirical likelihood methods based on influence functions. *Stat. Interface* **5** 355–366. MR2979215 https://doi.org/10.4310/SII.2012.v5.n3.a8
[81] ZHU, L., LIN, L., CUI, X. and LI, G. (2010). Bias-corrected empirical likelihood in a multi-link semiparametric model. *J. Multivariate Anal.* **101** 850–868. MR2584904 https://doi.org/10.1016/j.jmva.2009.08.009
[82] ZHU, L. and XUE, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 549–570. MR2278341 https://doi.org/10.1111/j.1467-9868.2006.00556.x

The Annals of Statistics
2020, Vol. 48, No. 1, 27–42
https://doi.org/10.1214/18-AOS1789

THE PHASE TRANSITION FOR THE EXISTENCE OF THE MAXIMUM LIKELIHOOD ESTIMATE IN HIGH-DIMENSIONAL LOGISTIC REGRESSION

BY EMMANUEL J. CANDÈS[1,2] AND PRAGYA SUR[2]

[1]*Department of Mathematics, Stanford University, candes@stanford.edu*
[2]*Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, pragya@seas.harvard.edu*

This paper rigorously establishes that the existence of the maximum likelihood estimate (MLE) in high-dimensional logistic regression models with Gaussian covariates undergoes a sharp "phase transition." We introduce an explicit boundary curve h_{MLE}, parameterized by two scalars measuring the overall magnitude of the unknown sequence of regression coefficients, with the following property: in the limit of large sample sizes n and number of features p proportioned in such a way that $p/n \to \kappa$, we show that if the problem is sufficiently high dimensional in the sense that $\kappa > h_{\mathrm{MLE}}$, then the MLE does not exist with probability one. Conversely, if $\kappa < h_{\mathrm{MLE}}$, the MLE asymptotically exists with probability one.

REFERENCES

[1] ALBERT, A. and ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71** 1–10. MR0738319 https://doi.org/10.1093/biomet/71.1.1
[2] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453 https://doi.org/10.1093/imaiai/iau005
[3] 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
[4] CHRISTMANN, A. and ROUSSEEUW, P. J. (2001). Measuring overlap in binary regression. *Comput. Statist. Data Anal.* **37** 65–75. MR1862480 https://doi.org/10.1016/S0167-9473(00)00063-3
[5] COVER, T. M. (1964). Geometrical and statistical properties of linear threshold devices. Ph.D. thesis.
[6] COVER, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* **3** 326–334.
[7] GORDON, Y. (1988). On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. MR0950977 https://doi.org/10.1007/BFb0081737
[8] KAUFMANN, H. (1988). On existence and uniqueness of a vector minimizing a convex function. *Z. Oper.-Res.* **32** 357–373. MR0976377 https://doi.org/10.1007/BF01920035
[9] KOLASSA, J. E. (1997). Infinite parameter estimates in logistic regression, with application to approximate conditional inference. *Scand. J. Stat.* **24** 523–530. MR1615343 https://doi.org/10.1111/1467-9469.00078
[10] KONIS, K. (2007). Linear programming algorithms for detecting separated data in binary logistic regression models. Ph.D. thesis, Univ. Oxford.
[11] LESAFFRE, E. and ALBERT, A. (1989). Partial separation in logistic discrimination. *J. Roy. Statist. Soc. Ser. B* **51** 109–116. MR0984997
[12] MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models. Monographs on Statistics and Applied Probability*. CRC Press, London. Second edition [of MR0727836]. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6
[13] NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A* **135** 370–384.
[14] OYMAK, S. and TROPP, J. A. (2018). Universality laws for randomized dimension reduction, with applications. *Inf. Inference* **7** 337–446. MR3858331 https://doi.org/10.1093/imaiai/iax011

MSC2010 subject classifications. 62J12, 62E20.
Key words and phrases. High-dimensional logistic regression, MLE phase transition.

[15] SANTNER, T. J. and DUFFY, D. E. (1986). A note on A. Albert and J. A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **73** 755–758. MR0897873 https://doi.org/10.1093/biomet/73.3.755

[16] SCHNEIDER, R. and WEIL, W. (2008). *Stochastic and Integral Geometry*. Springer, Berlin. MR2455326 https://doi.org/10.1007/978-3-540-78859-1

[17] SILVAPULLE, M. J. (1981). On the existence of maximum likelihood estimators for the binomial response models. *J. Roy. Statist. Soc. Ser. B* **43** 310–313. MR0637943

[18] SILVAPULLE, M. J. and BURRIDGE, J. (1986). Existence of maximum likelihood estimates in regression models for grouped and ungrouped data. *J. Roy. Statist. Soc. Ser. B* **48** 100–106. MR0848055

[19] SUR, P. and CANDÈS, E. J. (2018). A modern maximum-likelihood theory for high-dimensional logistic regression. ArXiv Preprint ArXiv:1803.06964.

[20] SUR, P., CHEN, Y. and CANDÈS, E. J. (2018). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probab. Theory Related Fields*. To Appear.

[21] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

The Annals of Statistics
2020, Vol. 48, No. 1, 43–63
https://doi.org/10.1214/18-AOS1790

RERANDOMIZATION IN 2^K FACTORIAL EXPERIMENTS

BY XINRAN LI[1], PENG DING[2] AND DONALD B. RUBIN[3]

[1]*Department of Statistics, University of Pennsylvania, lixinran@wharton.upenn.edu*
[2]*Department of Statistics, University of California, Berkeley, pengdingpku@berkeley.edu*
[3]*Department of Statistical Science, Temple University, donald.rubin@temple.edu*

With many pretreatment covariates and treatment factors, the classical factorial experiment often fails to balance covariates across multiple factorial effects simultaneously. Therefore, it is intuitive to restrict the randomization of the treatment factors to satisfy certain covariate balance criteria, possibly conforming to the tiers of factorial effects and covariates based on their relative importances. This is rerandomization in factorial experiments. We study the asymptotic properties of this experimental design under the randomization inference framework without imposing any distributional or modeling assumptions of the covariates and outcomes. We derive the joint asymptotic sampling distribution of the usual estimators of the factorial effects, and show that it is symmetric, unimodal and more "concentrated" at the true factorial effects under rerandomization than under the classical factorial experiment. We quantify this advantage of rerandomization using the notions of "central convex unimodality" and "peakedness" of the joint asymptotic sampling distribution. We also construct conservative large-sample confidence sets for the factorial effects.

REFERENCES

ANGRIST, J., LANG, D. and OREOPOULOS, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *Am. Econ. J. Appl. Econ.* **1** 136–163.

BAILEY, R. A. (1983). Restricted randomization. *Biometrika* **70** 183–198. MR0742988 https://doi.org/10.1093/biomet/70.1.183

BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2011). The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika* **98** 519–535. MR2836404 https://doi.org/10.1093/biomet/asr021

BICKEL, P. J. and LEHMANN, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Ann. Statist.* **4** 1139–1158. MR0474620

BIRNBAUM, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Stat.* **19** 76–81. MR0024099 https://doi.org/10.1214/aoms/1177730293

BOSE, R. C. (1947). Mathematical theory of the symmetrical factorial design. *Sankhyā* **8** 107–166. MR0026781

BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2140250

BRANSON, Z., DASGUPTA, T. and RUBIN, D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education high school study. *Ann. Appl. Stat.* **10** 1958–1976. MR3592044 https://doi.org/10.1214/16-AOAS959

BRUHN, M. and MCKENZIE, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *Am. Econ. J. Appl. Econ.* **1** 200–232.

COCHRAN, W. G. and COX, G. M. (1950). *Experimental Designs*. Wiley, New York. MR0034558

COX, D. R. (1982). Randomization and concomitant variables in the design of experiments. In *Statistics and Probability: Essays in Honor of C. R. Rao* (P. R. K. G. Kallianpur and J. K. Ghosh, eds.) 197–202. North-Holland, Amsterdam. MR0659470

COX, D. R. (2009). Randomization in the design of experiments. *Int. Stat. Rev.* **77** 415–429.

COX, D. R. and REID, N. (2000). *The Theory of the Design of Experiments*. Chapman and Hall/CRC, New York.

MSC2010 subject classifications. Primary 62K15, 62K10; secondary 62K05.
Key words and phrases. Covariate balance, tiers of covariates, tiers of factorial effects.

DASGUPTA, T., PILLAI, N. S. and RUBIN, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 727–753. MR3382595 https://doi.org/10.1111/rssb.12085

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications. Probability and Mathematical Statistics*. Academic Press, Boston, MA. MR0954608

DHARMADHIKARI, S. W. and JOGDEO, K. (1976). Multivariate unimodality. *Ann. Statist.* **4** 607–613. MR0415719

FINNEY, D. J. (1943). The fractional replication of factorial arrangements. *Ann. Hum. Genet.* **12** 291–301.

FISHER, R. A. (1926). The arrangement of field experiments. *J. Minist. Agric. G.B.* **33** 503–513.

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

GIOVAGNOLI, A. and WYNN, H. P. (1995). Multivariate dispersion orderings. *Statist. Probab. Lett.* **22** 325–332. MR1333191 https://doi.org/10.1016/0167-7152(94)00084-L

GRUNDY, P. M. and HEALY, M. J. R. (1950). Restricted randomization and quasi-Latin squares. *J. Roy. Statist. Soc. Ser. B* **12** 286–291.

HANSEN, B. B. and BOWERS, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statist. Sci.* **23** 219–236. MR2516821 https://doi.org/10.1214/08-STS254

HINKELMANN, K. and KEMPTHORNE, O. (2007). *Design and Analysis of Experiments, Vol.* 1, *Introduction to Experimental Design*, 2nd ed. Wiley, New York.

KANTER, M. (1977). Unimodality and dominance for symmetric random vectors. *Trans. Amer. Math. Soc.* **229** 65–85. MR0445580 https://doi.org/10.2307/1998500

KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments*. Wiley, New York. MR0045368

LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

LI, X., DING, P. and RUBIN, D. B. (2018). Asymptotic theory of rerandomization in treatment-control experiments. *Proc. Natl. Acad. Sci. USA* **115** 9157–9162. MR3856113 https://doi.org/10.1073/pnas.1808191115

LI, X., DING, P. and RUBIN, D. B. (2020). Supplement to "Rerandomization in 2^K factorial experiments." https://doi.org/10.1214/18-AOS1790SUPP.

LU, J. (2016). Covariate adjustment in randomization-based causal inference for 2^K factorial designs. *Statist. Probab. Lett.* **119** 11–20. MR3555264 https://doi.org/10.1016/j.spl.2016.07.010

MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (2009). *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. Springer, New York.

MORGAN, K. L. and RUBIN, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Ann. Statist.* **40** 1263–1282. MR2985950 https://doi.org/10.1214/12-AOS1008

MORGAN, K. L. and RUBIN, D. B. (2015). Rerandomization to balance tiers of covariates. *J. Amer. Statist. Assoc.* **110** 1412–1421. MR3449036 https://doi.org/10.1080/01621459.2015.1079528

SHAKED, M. (1985). Ordering distributions by dispersion. In *Encyclopedia of Statistical Sciences, Vol.* 6 (N. L. Johnson and S. Kotz, eds.) 485–490. Wiley, New York.

SHERMAN, S. (1955). A theorem on convex sets with applications. *Ann. Math. Stat.* **26** 763–767. MR0074845 https://doi.org/10.1214/aoms/1177728435

SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles (with discussion). Section 9. *Statist. Sci.* **5** 465–472. Translated from the Polish and edited by D. M. Dąbrowska and T. P. Speed. MR1092986

STUDENT (1938). Comparison between balanced and random arrangements of field plots. *Biometrika* **29** 363–378.

WU, C. F. J. (2015). Post-Fisherian experimentation: From physical to virtual. *J. Amer. Statist. Assoc.* **110** 612–620. MR3367251 https://doi.org/10.1080/01621459.2014.914441

WU, C. F. J. and HAMADA, M. S. (2011). *Experiments: Planning, Analysis, and Optimization*. Wiley, Hoboken, NJ.

YATES, F. (1937). The design and analysis of factorial experiments. Technical Communication No. 35, Imperial Bureau of Soil Sciences, Harpenden.

YATES, F. (1948). Comment to F. J. Anscombe. *J. R. Stat. Soc., A* **111** 204–205.

YOUDEN, W. J. (1972). Randomization and experimentation. *Technometrics* **14** 13–22.

The Annals of Statistics
2020, Vol. 48, No. 1, 64–85
https://doi.org/10.1214/18-AOS1791

SPARSE SIR: OPTIMAL RATES AND ADAPTIVE ESTIMATION

BY KAI TAN[1,*], LEI SHI[2] AND ZHOU YU[1,**]

[1]*School of Statistics, East China Normal University,* [*]*kaitan@stu.ecnu.edu.cn;* [**]*zyu@stat.ecnu.edu.cn*
[2]*School of Mathematical Sciences, Fudan University, leishi@fudan.edu.cn*

Sliced inverse regression (SIR) is an innovative and effective method for sufficient dimension reduction and data visualization. Recently, an impressive range of penalized SIR methods has been proposed to estimate the central subspace in a sparse fashion. Nonetheless, few of them considered the sparse sufficient dimension reduction from a decision-theoretic point of view. To address this issue, we in this paper establish the minimax rates of convergence for estimating the sparse SIR directions under various commonly used loss functions in the literature of sufficient dimension reduction. We also discover the possible trade-off between statistical guarantee and computational performance for sparse SIR. We finally propose an adaptive estimation scheme for sparse SIR which is computationally tractable and rate optimal. Numerical studies are carried out to confirm the theoretical properties of our proposed methods.

REFERENCES

BONDELL, H. D. and LI, L. (2009). Shrinkage inverse regression estimation for model-free variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 287–299. MR2655534 https://doi.org/10.1111/j.1467-9868.2008.00686.x

BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3** 1–122.

BURA, E. and COOK, R. D. (2001). Extending sliced inverse regression: The weighted chi-squared test. *J. Amer. Statist. Assoc.* **96** 996–1003. MR1946367 https://doi.org/10.1198/016214501753208979

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

CHEN, C.-H. and LI, K.-C. (1998). Can SIR be as popular as multiple linear regression? *Statist. Sinica* **8** 289–316. MR1624402

CHEN, X., ZOU, C. and COOK, R. D. (2010). Coordinate-independent sparse sufficient dimension reduction and variable selection. *Ann. Statist.* **38** 3696–3723. MR2766865 https://doi.org/10.1214/10-AOS826

COOK, R. D. (1996). Graphics for regressions with a binary response. *J. Amer. Statist. Assoc.* **91** 983–992. MR1424601 https://doi.org/10.2307/2291717

COOK, R. D. (2004). Testing predictor contributions in sufficient dimension reduction. *Ann. Statist.* **32** 1062–1092. MR2065198 https://doi.org/10.1214/009053604000000292

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

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

COOK, R. D. and WEISBERG, S. (1991). Discussion of sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 328–332.

COOK, R. D. and YIN, X. (2001). Dimension reduction and visualization in discriminant analysis. *Aust. N. Z. J. Stat.* **43** 147–199. MR1839361 https://doi.org/10.1111/1467-842X.00164

DONG, Y. and LI, B. (2010). Dimension reduction for non-elliptically distributed predictors: Second-order methods. *Biometrika* **97** 279–294. MR2650738 https://doi.org/10.1093/biomet/asq016

FERRÉ, L. (1998). Determining the dimension in sliced inverse regression and related methods. *J. Amer. Statist. Assoc.* **93** 132–140. MR1614604 https://doi.org/10.2307/2669610

FUKUMIZU, K., BACH, F. R. and JORDAN, M. I. (2009). Kernel dimension reduction in regression. *Ann. Statist.* **37** 1871–1905. MR2533474 https://doi.org/10.1214/08-AOS637

MSC2010 subject classifications. Primary 62H12; secondary 62G08, 62G09.
Key words and phrases. Sliced inverse regression, sufficient dimension reduction, sparsity.

GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. MR3718162 https://doi.org/10.1214/16-AOS1519

GAO, C., MA, Z., REN, Z. and ZHOU, H. H. (2015). Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.* **43** 2168–2197. MR3396982 https://doi.org/10.1214/15-AOS1332

HSING, T. and CARROLL, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20** 1040–1061. MR1165605 https://doi.org/10.1214/aos/1176348669

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

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

LI, L. (2007). Sparse sufficient dimension reduction. *Biometrika* **94** 603–613. MR2410011 https://doi.org/10.1093/biomet/asm044

LI, L., COOK, R. D. and NACHTSHEIM, C. J. (2005). Model-free variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 285–299. MR2137326 https://doi.org/10.1111/j.1467-9868.2005.00502.x

LI, B. and DONG, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *Ann. Statist.* **37** 1272–1298. MR2509074 https://doi.org/10.1214/08-AOS598

LI, L. and NACHTSHEIM, C. J. (2006). Sparse sliced inverse regression. *Technometrics* **48** 503–510. MR2328619 https://doi.org/10.1198/004017006000000129

LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102** 997–1008. MR2354409 https://doi.org/10.1198/016214507000000536

LI, L. and YIN, X. (2008). Sliced inverse regression with regularizations. *Biometrics* **64** 124–131, 323. MR2422826 https://doi.org/10.1111/j.1541-0420.2007.00836.x

LI, B., ZHA, H. and CHIAROMONTE, F. (2005). Contour regression: A general approach to dimension reduction. *Ann. Statist.* **33** 1580–1616. MR2166556 https://doi.org/10.1214/009053605000000192

LIN, Q., ZHAO, Z. and LIU, J. S. (2018a). On consistency and sparsity for sliced inverse regression in high dimensions. *Ann. Statist.* **46** 580–610. MR3782378 https://doi.org/10.1214/17-AOS1561

LIN, Q., ZHAO, Z. and LIU, J. S. (2018b). Sparse sliced inverse regression via lasso. *J. Amer. Statist. Assoc.* To appear.

LIN, Q., LI, X., HUANG, D. and LIU, J. S. (2017). On the optimality of sliced inverse regression in high dimensions. ArXiv Preprint ArXiv:1701.06009.

MA, Y. and ZHU, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.* **107** 168–179. MR2949349 https://doi.org/10.1080/01621459.2011.646925

NI, L., COOK, D. and TSAI, C.-L. (2005). A note on shrinkage sliced inverse regression. *Biometrika* **92** 242–247. MR2158624 https://doi.org/10.1093/biomet/92.1.242

TAN, K., SHI, L. and YU, Z. (2020). Supplement to "Sparse SIR: Optimal rates and adaptive estimation." https://doi.org/10.1214/18-AOS1791SUPP.

TAN, K. M., WANG, Z., ZHANG, T., LIU, H. and COOK, R. D. (2018a). A convex formulation for high-dimensional sparse sliced inverse regression. *Biometrika* **105** 769–782. MR3876161 https://doi.org/10.1093/biomet/asy049

TAN, K. M., WANG, Z., LIU, H. and ZHANG, T. (2018b). Sparse generalized eigenvalue problem: Optimal statistical rates via truncated Rayleigh flow. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 1057–1086.

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

WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438 https://doi.org/10.1214/15-AOS1369

WANG, H. and XIA, Y. (2008). Sliced regression for dimension reduction. *J. Amer. Statist. Assoc.* **103** 811–821. MR2524332 https://doi.org/10.1198/016214508000000418

XIA, Y., TONG, H., LI, W. K. and ZHU, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 363–410. MR1924297 https://doi.org/10.1111/1467-9868.03411

YANG, Z., BALASUBRAMANIAN, K. and LIU, H. (2017). On Stein's identity and near-optimal estimation in high-dimensional index models. ArXiv Preprint ArXiv:1709.08795.

YIN, X. and COOK, R. D. (2005). Direction estimation in single-index regressions. *Biometrika* **92** 371–384. MR2201365 https://doi.org/10.1093/biomet/92.2.371

YIN, X. and HILAFU, H. (2015). Sequential sufficient dimension reduction for large p, small n problems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 879–892. MR3382601 https://doi.org/10.1111/rssb.12093

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

YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963

YU, Z., DONG, Y. and SHAO, J. (2016). On marginal sliced inverse regression for ultrahigh dimensional model-free feature selection. *Ann. Statist.* **44** 2594–2623. MR3576555 https://doi.org/10.1214/15-AOS1424

YU, Z., ZHU, L., PENG, H. and ZHU, L. (2013). Dimension reduction and predictor selection in semiparametric models. *Biometrika* **100** 641–654. MR3094442 https://doi.org/10.1093/biomet/ast005

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

ZHOU, J. and HE, X. (2008). Dimension reduction based on constrained canonical correlation and variable filtering. *Ann. Statist.* **36** 1649–1668. MR2435451 https://doi.org/10.1214/07-AOS529

ZHU, L., MIAO, B. and PENG, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101** 630–643. MR2281245 https://doi.org/10.1198/016214505000001285

ZHU, L. X. and NG, K. W. (1995). Asymptotics of sliced inverse regression. *Statist. Sinica* **5** 727–736. MR1347616

ZHU, Y. and ZENG, P. (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. *J. Amer. Statist. Assoc.* **101** 1638–1651. MR2279485 https://doi.org/10.1198/016214506000000140

The Annals of Statistics
2020, Vol. 48, No. 1, 86–110
https://doi.org/10.1214/18-AOS1793

ROBUST SPARSE COVARIANCE ESTIMATION BY THRESHOLDING TYLER'S M-ESTIMATOR

BY JOHN GOES[1,*], GILAD LERMAN[1,**] AND BOAZ NADLER[2]

[1]*School of Mathematics, University of Minnesota, [*]johngoes@umn.edu; [**]lerman@umn.edu*
[2]*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, boaz.nadler@weizmann.ac.il*

Estimating a high-dimensional sparse covariance matrix from a limited number of samples is a fundamental task in contemporary data analysis. Most proposals to date, however, are not robust to outliers or heavy tails. Toward bridging this gap, in this work we consider estimating a sparse shape matrix from n samples following a possibly heavy-tailed elliptical distribution. We propose estimators based on thresholding either Tyler's M-estimator or its regularized variant. We prove that in the joint limit as the dimension p and the sample size n tend to infinity with $p/n \to \gamma > 0$, our estimators are minimax rate optimal. Results on simulated data support our theoretical analysis.

REFERENCES

ABRAMOVICH, Y. I. and SPENCER, N. K. (2007). Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering. In *Proceedings of the IEEE* 32*nd Intl*. *Conf*. *on Acoustics*, *Speech*, *and Signal Proc*. (*ICASSP*) 1105–1108.

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

AVELLA-MEDINA, M., BATTEY, H. S., FAN, J. and LI, Q. (2018). Robust estimation of high-dimensional covariance and precision matrices. *Biometrika* **105** 271–284. MR3804402 https://doi.org/10.1093/biomet/asy011

BALAKRISHNAN, S., DU, S. S., LI, J. and SINGH, A. (2017). Computationally efficient robust sparse estimation in high dimensions. In *Conference on Learning Theory* 169–212.

BHATIA, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

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

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

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

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

CAI, T. T. and ZHOU, H. H. (2012b). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statist*. *Sinica* **22** 1319–1349. MR3027084

CAMBANIS, S., HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J*. *Multivariate Anal*. **11** 368–385. MR0629795 https://doi.org/10.1016/0047-259X(81)90082-8

CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann*. *Statist*. **46** 1932–1960. MR3845006 https://doi.org/10.1214/17-AOS1607

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

COUILLET, R., KAMMOUN, A. and PASCAL, F. (2016). Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals. *J*. *Multivariate Anal*. **143** 249–274. MR3431431 https://doi.org/10.1016/j.jmva.2015.08.021

MSC2010 subject classifications. Primary 62H12; secondary 62G35, 62G20.
Key words and phrases. Covariance matrix estimation, spectral norm, elliptical distribution, sparsity, Tyler's M-estimator, thresholding.

COUILLET, R. and MCKAY, M. (2014). Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *J. Multivariate Anal.* **131** 99–120. MR3252638 https://doi.org/10.1016/j.jmva.2014.06.018

COUILLET, R., PASCAL, F. and SILVERSTEIN, J. W. (2014). Robust estimates of covariance matrices in the large dimensional regime. *IEEE Trans. Inform. Theory* **60** 7269–7278. MR3273053 https://doi.org/10.1109/TIT.2014.2354045

COUILLET, R., PASCAL, F. and SILVERSTEIN, J. W. (2015). The random matrix regime of Maronna's M-estimator with elliptically distributed samples. *J. Multivariate Anal.* **139** 56–78. MR3349480 https://doi.org/10.1016/j.jmva.2015.02.020

DAVIDSON, K. R. and SZAREK, S. J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach Spaces, Vol. I* 317–366. North-Holland, Amsterdam. MR1863696 https://doi.org/10.1016/S1874-5849(01)80010-3

DÜMBGEN, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.* **50** 471–491. MR1664575 https://doi.org/10.1023/A:1003573311481

DÜMBGEN, L., NORDHAUSEN, K. and SCHUHMACHER, H. (2016). New algorithms for M-estimation of multivariate scatter and location. *J. Multivariate Anal.* **144** 200–217. MR3434950 https://doi.org/10.1016/j.jmva.2015.11.009

DÜMBGEN, L., PAULY, M. and SCHWEIZER, T. (2015). M-functionals of multivariate scatter. *Stat. Surv.* **9** 32–105. MR3324616 https://doi.org/10.1214/15-SS109

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

FALK, M. (2002). The sample covariance is not efficient for elliptical distributions. *J. Multivariate Anal.* **80** 358–377. MR1889781 https://doi.org/10.1006/jmva.2000.1983

FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions. Monographs on Statistics and Applied Probability* **36**. CRC Press, London. MR1071174 https://doi.org/10.1007/978-1-4899-2937-2

FRAHM, G. (2004). Generalized elliptical distributions: Theory and applications. Ph.D. thesis, Univ. Köln.

FRAHM, G. and JAEKEL, U. (2007). Tyler's M-estimator, random matrix theory, and generalized elliptical distributions with applications to finance. Discussion Papers in Econometrics and Statistics, No. 2/07, Institute of Econometrics and Statistics, Univ. Cologne.

FRAHM, G. and JAEKEL, U. (2010). A generalization of Tyler's M-estimators to the case of incomplete data. *Comput. Statist. Data Anal.* **54** 374–393. MR2756433 https://doi.org/10.1016/j.csda.2009.08.019

GOES, J., LERMAN, G. and NADLER, B. (2020). Supplement to "Robust sparse covariance estimation by thresholding Tyler's M-estimator." https://doi.org/10.1214/18-AOS1793SUPP.

GUIONNET, A. and ZEITOUNI, O. (2000). Concentration of the spectral measure for large matrices. *Electron. Commun. Probab.* **5** 119–136. MR1781846 https://doi.org/10.1214/ECP.v5-1026

HAN, F., LU, J. and LIU, H. (2014). Robust scatter matrix estimation for high dimensional distributions with heavy tails. Technical Report, Princeton Univ., Princeton, NJ.

HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

KAMMOUN, A., COUILLET, R., PASCAL, F. and ALOUINI, M.-S. (2018). Optimal design of the adaptive normalized matched filter detector using regularized Tyler estimators. *IEEE Trans. Aerosp. Electron. Syst.* **54** 755–769. MR3456682 https://doi.org/10.1109/TSP.2015.2494858

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–438. MR0287628

KENT, J. T. and TYLER, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Stat.* **15** 247–254.

KENT, J. T. and TYLER, D. E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.* **19** 2102–2119. MR1135166 https://doi.org/10.1214/aos/1176348388

LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459 https://doi.org/10.1214/09-AOS720

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, London. MR0560319

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67. MR0388656

MARONNA, R. A. and YOHAI, V. J. (2017). Robust and efficient estimation of multivariate scatter and location. *Comput. Statist. Data Anal.* **109** 64–75. MR3603641 https://doi.org/10.1016/j.csda.2016.11.006

MORALES-JIMENEZ, D., COUILLET, R. and MCKAY, M. R. (2015). Large dimensional analysis of robust M-estimators of covariance with outliers. *IEEE Trans. Signal Process.* **63** 5784–5797. MR3405886 https://doi.org/10.1109/TSP.2015.2460225

NORDHAUSEN, K. and TYLER, D. E. (2015). A cautionary note on robust covariance plug-in methods. *Biometrika* **102** 573–588. MR3394276 https://doi.org/10.1093/biomet/asv022

OLLILA, E. and KOIVUNEN, V. (2003). Robust antenna array processing using M-estimators of pseudo-covariance. In *Proceedings of the IEEE* 14*th International Symposium on Personal*, *Indoor and Mobile Radio Communications* (*PIMRC*) 2659–2663.

OLLILA, E. and TYLER, D. E. (2012). Distribution-free detection under complex elliptically symmetric clutter distribution. In *IEEE* 7*th Sensor Array and Multichannel Signal Processing Workshop*, (*SAM*) 413–416.

OLLILA, E. and TYLER, D. E. (2014). Regularized M-estimators of scatter matrix. *IEEE Trans*. *Signal Process*. **62** 6059–6070. MR3281544 https://doi.org/10.1109/TSP.2014.2360826

PASCAL, F., CHITOUR, Y. and QUEK, Y. (2014). Generalized robust shrinkage estimator and its application to STAP detection problem. *IEEE Trans*. *Signal Process*. **62** 5640–5651. MR3273520 https://doi.org/10.1109/TSP.2014.2355779

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J*. *Amer. Statist*. *Assoc*. **104** 177–186. MR2504372 https://doi.org/10.1198/jasa.2009.0101

RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron*. *Commun*. *Probab.* **18** no. 82, 9. MR3125258 https://doi.org/10.1214/ECP.v18-2865

SIRKIÄ, S., TASKINEN, S. and OJA, H. (2007). Symmetrised M-estimators of multivariate scatter. *J. Multivariate Anal*. **98** 1611–1629. MR2370110 https://doi.org/10.1016/j.jmva.2007.06.005

SOLOVEYCHIK, I. and WIESEL, A. (2014). Tyler's covariance matrix estimator in elliptical models with convex structure. *IEEE Trans*. *Signal Process*. **62** 5251–5259. MR3268109 https://doi.org/10.1109/TSP.2014.2348951

SUN, Y., BABU, P. and PALOMAR, D. P. (2014). Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms. *IEEE Trans*. *Signal Process*. **62** 5143–5156. MR3260359 https://doi.org/10.1109/TSP.2014.2348944

SUN, Y., BABU, P. and PALOMAR, D. P. (2016). Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions. *IEEE Trans*. *Signal Process*. **64** 3576–3590. MR3515702 https://doi.org/10.1109/TSP.2016.2546222

TYLER, D. E. (1987a). A distribution-free M-estimator of multivariate scatter. *Ann*. *Statist*. **15** 234–251. MR0885734 https://doi.org/10.1214/aos/1176350263

TYLER, D. E. (1987b). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika* **74** 579–589. MR0909362 https://doi.org/10.1093/biomet/74.3.579

WIESEL, A. (2012). Unified framework to regularized covariance estimation in scaled Gaussian models. *IEEE Trans*. *Signal Process*. **60** 29–38. MR2932100 https://doi.org/10.1109/TSP.2011.2170685

WIESEL, A. and ZHANG, T. (2014). Structured robust covariance estimation. *Found*. *Trends Signal Process*. **8** 127–216. MR3438278 https://doi.org/10.1561/2000000053

ZHANG, T., CHENG, X. and SINGER, A. (2016). Marčenko–Pastur law for Tyler's M-estimator. *J. Multivariate Anal*. **149** 114–123. MR3507318 https://doi.org/10.1016/j.jmva.2016.03.010

The Annals of Statistics
2020, Vol. 48, No. 1, 111–137
https://doi.org/10.1214/18-AOS1794

MODEL ASSISTED VARIABLE CLUSTERING: MINIMAX-OPTIMAL RECOVERY AND ALGORITHMS

BY FLORENTINA BUNEA[1], CHRISTOPHE GIRAUD[2,*], XI LUO[3], MARTIN ROYER[2,**] AND NICOLAS VERZELEN[4]

[1]*Department of Statistical Science, Cornell University, fb238@cornell.edu*

[2]*Laboratoire de Mathématiques d'Orsay, CNRS, Université Paris-Sud, Université Paris-Saclay,* [*]*christophe.giraud@math.u-psud.fr;* [**]*martin.royer@math.u-psud.fr*

[3]*Department of Biostatistics and Data Science, School of Public Health, University of Texas Health Science Center at Houston, xi.rossi.luo@gmail.com*

[4]*INRA, Montpellier SupAgro, MISTEA, Université de Montpellier, nicolas.verzelen@inra.fr*

The problem of variable clustering is that of estimating groups of similar components of a p-dimensional vector $X = (X_1, \dots, X_p)$ from n independent copies of X. There exists a large number of algorithms that return data-dependent groups of variables, but their interpretation is limited to the algorithm that produced them. An alternative is model-based clustering, in which one begins by defining population level clusters relative to a model that embeds notions of similarity. Algorithms tailored to such models yield estimated clusters with a clear statistical interpretation. We take this view here and introduce the class of G-block covariance models as a background model for variable clustering. In such models, two variables in a cluster are deemed similar if they have similar associations will all other variables. This can arise, for instance, when groups of variables are noise corrupted versions of the same latent factor. We quantify the difficulty of clustering data generated from a G-block covariance model in terms of cluster proximity, measured with respect to two related, but different, cluster separation metrics. We derive minimax cluster separation thresholds, which are the metric values below which no algorithm can recover the model-defined clusters exactly, and show that they are different for the two metrics. We therefore develop two algorithms, COD and PECOK, tailored to G-block covariance models, and study their minimax-optimality with respect to each metric. Of independent interest is the fact that the analysis of the PECOK algorithm, which is based on a corrected convex relaxation of the popular K-means algorithm, provides the first statistical analysis of such algorithms for variable clustering. Additionally, we compare our methods with another popular clustering method, spectral clustering. Extensive simulation studies, as well as our data analyses, confirm the applicability of our approach.

REFERENCES

[1] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2017). Entrywise eigenvector analysis of random matrices with low expected rank. ArXiv Preprint arXiv:1709.09565.

[2] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 670–688. IEEE Computer Soc., Los Alamitos, CA. MR3473334

[3] AMINI, A. A. and LEVINA, E. (2018). On semidefinite relaxations for the block model. *Ann. Statist.* **46** 149–179. MR3766949 https://doi.org/10.1214/17-AOS1545

[4] ARTHUR, D. and VASSILVITSKII, S. (2007). `k-means++`: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 1027–1035. ACM, New York. MR2485254

MSC2010 subject classifications. Primary 62H30; secondary 62C20.
Key words and phrases. Convergence rates, convex optimization, covariance matrices, high-dimensional inference.

[5] AWASTHI, P., CHARIKAR, M., KRISHNASWAMY, R. and SINOP, A. K. (2015). The hardness of approximations of Euclidean k-means. In *31st International Symposium on Computational Geometry*. *LIPIcs. Leibniz Int. Proc. Inform.* **34** 754–767. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3392820

[6] BANERJEE, O., EL GHAOUI, L. and D'ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. MR2417243

[7] BELLEC, P., PERLBARG, V., JBABDI, S., PÉLÉGRINI-ISSAC, M., ANTON, J.-L., DOYON, J. and BENALI, H. (2006). Identification of large-scale networks in the brain using fMRI. *NeuroImage* **29** 1231–1243.

[8] BERNARDES, J. S., VIEIRA, F. R., COSTA, L. M. and ZAVERUCHA, G. (2015). Evaluation and improvements of clustering algorithms for detecting remote homologous protein families. *BMC Bioinform.* **16** 1–14.

[9] BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In *Proceedings of the 26th Annual Conference on Learning Theory* (S. Shalev-Shwartz and I. Steinwart, eds.). *Proceedings of Machine Learning Research* **30** 1046–1066. PMLR, Princeton, NJ.

[10] BERTHET, Q., RIGOLLET, P. and SRIVASTAVA, P. (2018). Exact recovery in the Ising blockmodel. *Ann. Statist.* To appear. arXiv:1612.03880.

[11] BING, M., BUNEA, F., NING, Y. and WEGKAMP, M. (2018). Adaptive estimation in structured factor models with applications to overlapping clustering. ArXiv E-prints.

[12] BOUVEYRON, C. and BRUNET-SAUMARD, C. (2014). Model-based clustering of high-dimensional data: A review. *Comput. Statist. Data Anal.* **71** 52–78. MR3131954 https://doi.org/10.1016/j.csda.2012.12.008

[13] BUNEA, F., GIRAUD, C. and LUO, X. (2015). Minimax optimal variable clustering in G-models via cord. ArXiv Preprint arXiv:1508.01939.

[14] BUNEA, F., GIRAUD, C., LUO, X., ROYER, M. and VERZELEN, N. (2020). Supplement to "Model assisted variable clustering: Minimax-optimal recovery and algorithms." https://doi.org/10.1214/18-AOS1794SUPP.

[15] BUNEA, F., GIRAUD, C., ROYER, M. and VERZELEN, N. (2016). PECOK: A convex optimization approach to variable clustering. ArXiv Preprint arXiv:1606.05100.

[16] CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** Paper No. 27, 57. MR3491121

[17] CHONG, M., BHUSHAN, C., JOSHI, A. A., CHOI, S., HALDAR, J. P., SHATTUCK, D. W., SPRENG, R. N. and LEAHY, R. M. (2017). Individual parcellation of resting fMRI with a group functional connectivity prior. *NeuroImage* **156** 87–100. https://doi.org/10.1016/j.neuroimage.2017.04.054

[18] CRADDOCK, R. C., JAMES, G. A., HOLTZHEIMER, P. E., HU, X. P. and MAYBERG, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* **33** 1914–1928.

[19] FREI, N., GARCIA, A. V., BIGEARD, J., ZAAG, R., BUESO, E., GARMIER, M., PATEYRON, S., DE TAUZIA-MOREAU, M. L., BRUNAUD, V. et al. (2014). Functional analysis of Arabidopsisimmune-related MAPKs uncovers a role for MPK3 as negative regulator of inducible defences. *Genome Biol.* **15** 1–22.

[20] GLASSER, M. F., COALSON, T. S., ROBINSON, E. C., HACKER, C. D., HARWELL, J., YACOUB, E., UGURBIL, K., ANDERSSON, J., BECKMANN, C. F. et al. (2016). A multi-modal parcelation of human cerebral cortex. *Nature* **536** 171–178.

[21] GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025 https://doi.org/10.1007/s00440-015-0659-z

[22] JAMES, G. A., HAZAROGLU, O. and BUSH, K. A. (2016). A human brain atlas derived via n-cut parcellation of resting-state and task-based fMRI data. *Magn. Reson. Imaging* **34** 209–218.

[23] JIANG, D., TANG, C. and ZHANG, A. (2004). Cluster analysis for gene expression data: A survey. *IEEE Trans. Knowl. Data Eng.* **16** 1370–1386.

[24] 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

[25] KONG, R., LI, J., SUN, N., SABUNCU, M., LIU, H., SCHAEFER, A., ZUO, X.-N., HOLMES, A., EICKHOFF, S. et al. (2018). Spatial topography of individual-specific cortical networks predicts human cognition, personality and emotion. https://www.biorxiv.org/content/early/2018/01/31/213041.

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

[27] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2016). Optimization via low-rank approximation for community detection in networks. *Ann. Statist.* **44** 373–400. MR3449772 https://doi.org/10.1214/15-AOS1360

[28] 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

[29] LEI, J. and ZHU, L. (2014). A generic sample splitting approach for refined community recovery in stochastic block models. ArXiv Preprint arXiv:1411.1469.

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

[31] LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of Lloyd's algorithm and its variants. ArXiv Preprint arXiv:1612.02099.

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

[33] MOSSEL, E., NEEMAN, J. and SLY, A. (2014). Consistency thresholds for binary symmetric block models. ArXiv Preprint arXiv:1407.1591.

[34] 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

[35] PERRY, A. and WEIN, A. S. (2015). A semidefinite program for unbalanced multisection in the stochastic block model. ArXiv E-prints arXiv:1507.05605.

[36] POLDRACK, R. A. (2007). Region of interest analysis for fMRI. *Soc. Cogn. Affect. Neurosci.* **2** 67–70. https://doi.org/10.1093/scan/nsm006

[37] POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678.

[38] ROYER, M. (2017). Adaptive clustering through semidefinite programming. In *Advances in Neural Information Processing Systems* (*NIPS*).

[39] XUE, G., ARON, A. R. and POLDRACK, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cereb. Cortex* **18** 1923–1932.

[40] YEO, B. T., KRIENEN, F. M., SEPULCRE, J., SABUNCU, M. R., LASHKARI, D., HOLLINSHEAD, M., ROFFMAN, J. L., SMOLLER, J. W., ZÖLLEI, L. et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106** 1125–1165.

[41] ZAAG, R., TAMBY, J., GUICHARD, C., TARIQ, Z., RIGAILL, G., DELANNOY, E., RENOU, J., BALZERGUE, S., MARY-HUARD, T. et al. (2015). GEM2Net: From gene expression modeling to -omics networks, a new CATdb module to investigate *Arabidopsis thaliana* genes involved in stress response. *Nucleic Acids Res.* **43** 1010–1017.

The Annals of Statistics
2020, Vol. 48, No. 1, 138–160
https://doi.org/10.1214/18-AOS1795

NEW *G*-FORMULA FOR THE SEQUENTIAL CAUSAL EFFECT AND BLIP EFFECT OF TREATMENT IN SEQUENTIAL CAUSAL INFERENCE

BY XIAOQIN WANG[1] AND LI YIN[2]

[1]*Department of Electrical Engineering, Mathematics and Science, University of Gävle, xwg@hig.se*
[2]*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, li.yin@ki.se*

In sequential causal inference, two types of causal effects are of practical interest, namely, the causal effect of the treatment regime (called the sequential causal effect) and the blip effect of treatment on the potential outcome after the last treatment. The well-known G-formula expresses these causal effects in terms of the standard parameters. In this article, we obtain a new G-formula that expresses these causal effects in terms of the point observable effects of treatments similar to treatment in the framework of single-point causal inference. Based on the new G-formula, we estimate these causal effects by maximum likelihood via point observable effects with methods extended from single-point causal inference. We are able to increase precision of the estimation without introducing biases by an unsaturated model imposing constraints on the point observable effects. We are also able to reduce the number of point observable effects in the estimation by treatment assignment conditions.

REFERENCES

[1] LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1925014 https://doi.org/10.1002/9781119013563

[2] MCCULLAGH, P. and NELDER, J. A. (1999). *Generalized Linear Models*. Chapman & Hall/CRC, Boca Raton, FL.

[3] PETERSEN, M. L., PORTER, K. E., GRUBER, S., WANG, Y. and VAN DER LAAN, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Stat. Methods Med. Res.* **21** 31–54. MR2867537 https://doi.org/10.1177/0962280210386207

[4] ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512. MR0877758 https://doi.org/10.1016/0270-0255(86)90088-6

[5] ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (*Los Angeles*, *CA*, 1994) (M. Berkane, ed.). *Lect. Notes Stat.* **120** 69–117. Springer, New York. MR1601279 https://doi.org/10.1007/978-1-4612-1842-5_4

[6] ROBINS, J. M. (1999). Association, causation, and marginal structural models. *Synthese* **121** 151–179. MR1766776 https://doi.org/10.1023/A:1005285815569

[7] ROBINS, J. M. (2009). Longitudinal data analysis. In *Handbooks of Modern Statistical Methods* (G. Fitzmaurice, ed.) 553–599. Chapman & Hall/CRC, Boca Raton, FL.

[8] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

[9] RUBIN, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100** 322–331. MR2166071 https://doi.org/10.1198/016214504000001880

[10] SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. Translated from the Polish and edited by D. M. Dąbrowska and T. P. Speed. MR1092986

[11] TAUBMAN, S. L., ROBINS, J. M., MITTLEMAN, M. A. and HERNÁN, M. A. (2009). Intervening on risk factors for coronary heart disease: An application of the parametric g-formula. *Int. J. Epidemiol.* **38** 1599–1611.

MSC2010 subject classifications. Primary 62H12; secondary 62H15, 62F03, 62F30.
Key words and phrases. Blip effect, curse of dimensionality, new G-formula, null paradox, point observable effect, sequential causal effect.

[12] WANG, X. and YIN, L. (2015). Identifying and estimating net effects of treatments in sequential causal inference. *Electron. J. Stat.* **9** 1608–1643. MR3379004 https://doi.org/10.1214/15-EJS1046
[13] WANG, X. and YIN, L. (2020). Supplement to "New G-formula for the sequential causal effect and blip effect of treatment in sequential causal inference." https://doi.org/10.1214/18-AOS1795SUPP.
[14] ZEGER, S. L. and DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

The Annals of Statistics
2020, Vol. 48, No. 1, 161–182
https://doi.org/10.1214/18-AOS1796

ENVELOPE-BASED SPARSE PARTIAL LEAST SQUARES

BY GUANGYU ZHU[1] AND ZHIHUA SU[2]

[1]*Department of Computer Science and Statistics, University of Rhode Island, guangyuzhu@uri.edu*
[2]*Department of Statistics, University of Florida, zhihuasu@ufl.edu*

Sparse partial least squares (SPLS) is widely used in applied sciences as a method that performs dimension reduction and variable selection simultaneously in linear regression. Several implementations of SPLS have been derived, among which the SPLS proposed in Chun and Keleş (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** (2010) 3–25) is very popular and highly cited. However, for all of these implementations, the theoretical properties of SPLS are largely unknown. In this paper, we propose a new version of SPLS, called the envelope-based SPLS, using a connection between envelope models and partial least squares (PLS). We establish the consistency, oracle property and asymptotic normality of the envelope-based SPLS estimator. The large-sample scenario and high-dimensional scenario are both considered. We also develop the envelope-based SPLS estimators under the context of generalized linear models, and discuss its theoretical properties including consistency, oracle property and asymptotic distribution. Numerical experiments and examples show that the envelope-based SPLS estimator has better variable selection and prediction performance over the SPLS estimator (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** (2010) 3–25).

REFERENCES

AGRESTI, A. (2013). *Categorical Data Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR3087436

CHEN, L. and HUANG, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *J. Amer. Statist. Assoc.* **107** 1533–1545. MR3036414 https://doi.org/10.1080/01621459.2012.734178

CHEN, X., ZOU, C. and COOK, R. D. (2010). Coordinate-independent sparse sufficient dimension reduction and variable selection. *Ann. Statist.* **38** 3696–3723. MR2766865 https://doi.org/10.1214/10-AOS826

CHUN, H. and KELEŞ, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 3–25. MR2751241 https://doi.org/10.1111/j.1467-9868.2009.00723.x

CHUNG, D. and KELEŞ, S. (2010). Sparse partial least squares classification for high dimensional data. *Stat. Appl. Genet. Mol. Biol.* **9** Art. 17, 32. MR2721697 https://doi.org/10.2202/1544-6115.1492

COOK, R. D., FORZANI, L. and SU, Z. (2016). A note on fast envelope estimation. *J. Multivariate Anal.* **150** 42–54. MR3534901 https://doi.org/10.1016/j.jmva.2016.05.006

COOK, R. D., HELLAND, I. S. and SU, Z. (2013). Envelopes and partial least squares regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 851–877. MR3124794 https://doi.org/10.1111/rssb.12018

COOK, R. D., LI, B. and CHIAROMONTE, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica* **20** 927–1010.

COOK, R. D. and ZHANG, X. (2015). Foundations for envelope models and methods. *J. Amer. Statist. Assoc.* **110** 599–611. MR3367250 https://doi.org/10.1080/01621459.2014.983235

COOK, R. D. and ZHANG, X. (2016). Algorithms for envelope estimation. *J. Comput. Graph. Statist.* **25** 284–300. MR3474048 https://doi.org/10.1080/10618600.2015.1029577

DE JONG, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **18** 251–263.

DIACONIS, P. and FREEDMAN, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12** 793–815. MR0751274 https://doi.org/10.1214/aos/1176346703

MSC2010 subject classifications. Primary 62F12, 62B05; secondary 62J05, 62J12.

Key words and phrases. Partial least squares, envelope model, sufficient dimension reduction, Grassmann manifold.

DING, B. and GENTLEMAN, R. (2005). Classification using generalized partial least squares. *J. Comput. Graph. Statist.* **14** 280–298. MR2160814 https://doi.org/10.1198/106186005X47697

HUANG, X., PAN, W., PARK, S., HAN, X., MILLER, L. W. and HALL, J. (2004). Modeling the relationship between LVAD support time and gene expression changes in the human heart by penalized partial least squares. *Bioinformatics* **20** 888–894.

KARABULUT, E. M. and IBRIKCI, T. (2014). Effective automated prediction of vertebral column pathologies based on logistic model tree with SMOTE preprocessing. *J. Med. Syst.* **38** 1–9.

KHARE, K., OH, S.-Y. and RAJARATNAM, B. (2015). A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 803–825. MR3382598 https://doi.org/10.1111/rssb.12088

KHARE, K., PAL, S. and SU, Z. (2017). A Bayesian approach for envelope models. *Ann. Statist.* **45** 196–222. MR3611490 https://doi.org/10.1214/16-AOS1449

LÊ CAO, K.-A., ROSSOUW, D., ROBERT-GRANIÉ, C. and BESSE, P. (2008). A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. Biol.* **7** Art. 35, 31. MR2457048 https://doi.org/10.2202/1544-6115.1390

LEE, D., LEE, W., LEE, Y. and PAWITAN, Y. (2011). Sparse partial least-squares regression and its applications to high-throughput data analysis. *Chemom. Intell. Lab. Syst.* **109** 1–8.

LICHMAN, M. (2013). UCI machine learning repository.

MA, Y. and ZHU, L. (2013). Efficiency loss and the linearity condition in dimension reduction. *Biometrika* **100** 371–383. MR3068440 https://doi.org/10.1093/biomet/ass075

MARX, B. D. (1996). Iteratively reweighted partial least squares estimation for generalized linear regression. *Technometrics* **38** 374–381.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models. Monographs on Statistics and Applied Probability*. CRC Press, London. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

PARK, P. J., TIAN, L. and KOHANE, I. S. (2002). Linking gene expression data with patient survival times using partial least squares. *Bioinformatics* **18** S120–S127.

PENG, J., WANG, P., ZHOU, N. and ZHU, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Amer. Statist. Assoc.* **104** 735–746. MR2541591 https://doi.org/10.1198/jasa.2009.0126

RAMSEY, F. and SCHAFER, D. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis*. Cengage Learning, Boston.

ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. MR2417391 https://doi.org/10.1214/08-EJS176

SU, Z. and COOK, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98** 133–146. MR2804215 https://doi.org/10.1093/biomet/asq063

SU, Z. and COOK, D. (2012). Inner envelopes: Efficient estimation in multivariate linear regression. *Biometrika* **99** 687–702. MR2966778 https://doi.org/10.1093/biomet/ass024

SU, Z., ZHU, G., CHEN, X. and YANG, Y. (2016). Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression. *Biometrika* **103** 579–593. MR3551785 https://doi.org/10.1093/biomet/asw036

WANG, L., CHEN, G. and LI, H. (2007). Group SCAD regression analysis for microarray time course gene expression data. *Bioinformatics* **23** 1486–1494.

WOLD, H. (1966). Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis* (*Proc. Internat. Sympos., Dayton, Ohio*, 1965) 391–420. Academic Press, New York. MR0220397

WOLD, H. (1975). *Path Models with Latent Variables: The NIPALS Approach*. Academic Press, Cambridge.

ZHANG, T. and ZOU, H. (2014). Sparse precision matrix estimation via lasso penalized D-trace loss. *Biometrika* **101** 103–120. MR3180660 https://doi.org/10.1093/biomet/ast059

ZHU, G. and SU, Z. (2019). Supplement to "Envelope-based sparse partial least squares." https://doi.org/10.1214/18-AOS1796SUPP.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

The Annals of Statistics
2020, Vol. 48, No. 1, 183–204
https://doi.org/10.1214/18-AOS1797

OPTIMAL RATES FOR COMMUNITY ESTIMATION IN THE WEIGHTED STOCHASTIC BLOCK MODEL

BY MIN XU[1], VARUN JOG[2,*] AND PO-LING LOH[2,**]

[1]*Department of Statistics, Rutgers University, mx76@stat.rutgers.edu*
[2]*Departments of ECE & Statistics, University of Wisconsin, Madison, *vjog@wisc.edu; **ploh@stat.wisc.edu*

Community identification in a network is an important problem in fields such as social science, neuroscience and genetics. Over the past decade, stochastic block models (SBMs) have emerged as a popular statistical framework for this problem. However, SBMs have an important limitation in that they are suited only for networks with unweighted edges; in various scientific applications, disregarding the edge weights may result in a loss of valuable information. We study a weighted generalization of the SBM, in which observations are collected in the form of a weighted adjacency matrix and the weight of each edge is generated independently from an unknown probability density determined by the community membership of its endpoints. We characterize the optimal rate of misclustering error of the weighted SBM in terms of the Renyi divergence of order 1/2 between the weight distributions of within-community and between-community edges, substantially generalizing existing results for unweighted SBMs. Furthermore, we present a computationally tractable algorithm based on discretization that achieves the optimal error rate. Our method is adaptive in the sense that the algorithm, without assuming knowledge of the weight densities, performs as well as the best algorithm that knows the weight densities.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[2] ABBE, E., BANDEIRA, A. S., BRACHER, A. and SINGER, A. (2014). Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Trans. Netw. Sci. Eng.* **1** 10–22. MR3349181 https://doi.org/10.1109/TNSE.2014.2368716

[3] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory* **62** 471–487. MR3447993 https://doi.org/10.1109/TIT.2015.2490670

[4] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS* 2015 670–688. IEEE Computer Soc., Los Alamitos, CA. MR3473334

[5] ABBE, E. and SANDON, C. (2015). Recovering communities in the general stochastic block model without knowing the parameters. In *Advances in Neural Information Processing Systems* 676–684.

[6] AICHER, C., JACOBS, A. Z. and CLAUSET, A. (2015). Learning latent block structure in weighted networks. *J. Complex Netw.* **3** 221–248. MR3365464 https://doi.org/10.1093/comnet/cnu026

[7] BALAKRISHNAN, S., XU, M., KRISHNAMURTHY, A. and SINGH, A. (2011). Noise thresholds for spectral clustering. In *Advances in Neural Information Processing Systems* 954–962.

[8] BARRAT, A., BARTHELEMY, M., PASTOR-SATORRAS, R. and VESPIGNANI, A. (2004). The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101** 3747–3752.

[9] BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. and LEFEBVRE, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **10**.

[10] BOCCALETTI, S., LATORA, V., MORENO, Y., CHAVEZ, M. and HWANG, D.-U. (2006). Complex networks: Structure and dynamics. *Phys. Rep.* **424** 175–308. MR2193621 https://doi.org/10.1016/j.physrep.2005.10.009

MSC2010 subject classifications. Primary 62H30, 91D30; secondary 62C20, 90B15.
Key words and phrases. Stochastic block models, network analysis, Renyi divergence, nonparametric estimation, optimal estimation rates.

[11] CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Proceedings of the* 28*th Conference on Learning Theory* 391–423.

[12] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys*. *Rev*. *E* **84**.

[13] EASLEY, D. and KLEINBERG, J. (2010). *Networks*, *Crowds*, *and Markets*: *Reasoning About a Highly Connected World*. Cambridge Univ. Press, Cambridge. MR2677125 https://doi.org/10.1017/CBO9780511761942

[14] FIENBERG, S. E., MEYER, M. M. and WASSERMAN, S. S. (1985). Statistical analysis of multiple sociometric relations. *J*. *Amer*. *Statist*. *Assoc*. **80** 51–67.

[15] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J*. *Mach*. *Learn*. *Res*. **18** Paper No. 60, 45. MR3687603

[16] GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found*. *Trends Mach*. *Learn*. **2** 129–233.

[17] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans*. *Inform*. *Theory* **62** 2788–2797. MR3493879 https://doi.org/10.1109/TIT.2016.2546280

[18] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans*. *Inform*. *Theory* **62** 5918–5937. MR3552431 https://doi.org/10.1109/TIT.2016.2594812

[19] HAJEK, B., WU, Y. and XU, J. (2017). Information limits for recovering a hidden community. *IEEE Trans*. *Inform*. *Theory* **63** 4729–4745. MR3683533 https://doi.org/10.1109/TIT.2017.2653804

[20] HAJEK, B., WU, Y. and XU, J. (2017). Submatrix localization via message passing. *J*. *Mach*. *Learn*. *Res*. **18** Paper No. 186, 52. MR3827074

[21] HARTUV, E. and SHAMIR, R. (2000). A clustering algorithm based on graph connectivity. *Inform*. *Process*. *Lett*. **76** 175–181. MR1807676 https://doi.org/10.1016/S0020-0190(00)00142-3

[22] HEIMLICHER, S., LELARGE, M. and MASSOULIÉ, L. (2012). Community detection in the labelled stochastic block model. Preprint. Available at arXiv:1209.2910.

[23] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc*. *Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[24] JACKSON, M. O. (2008). *Social and Economic Networks*. Princeton Univ. Press, Princeton, NJ. MR2435744

[25] JOG, V. and LOH, P. (2015). Information-theoretic bounds for exact recovery in weighted stochastic block models using the Renyi divergence. Preprint. Available at arXiv:1509.06418.

[26] 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

[27] LELARGE, M., MASSOULIÉ, L. and XU, J. (2015). Reconstruction in the labelled stochastic block model. *IEEE Trans*. *Netw*. *Sci*. *Eng*. **2** 152–163. MR3453283 https://doi.org/10.1109/TNSE.2015.2490580

[28] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'*14—*Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

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

[30] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. Preprint. Available at arXiv:1202.1499.

[31] MOSSEL, E., NEEMAN, J. and SLY, A. (2014). Consistency thresholds for binary symmetric block models. Preprint. Available at arXiv:1407.1591.

[32] MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. MR3876880 https://doi.org/10.1007/s00493-016-3238-8

[33] NEWMAN, M., BARABÁSI, A.-L. and WATTS, D. J., eds. (2006). *The Structure and Dynamics of Networks*. *Princeton Studies in Complexity*. Princeton Univ. Press, Princeton, NJ. MR2352222

[34] NEWMAN, M. E. J. (2004). Analysis of weighted networks. *Phys*. *Rev*. *E* **70**.

[35] NEWMAN, M. E. J. and GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Phys*. *Rev*. *E* **69**.

[36] PRITCHARD, J. K., STEPHENS, M. and DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155** 945–959.

[37] RUBINOV, M. and SPORNS, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52** 1059–1069.

[38] SADE, D. S. (1972). Sociometrics of Macaca mulatta: I. Linkages and cliques in grooming matrices. *Folia Primatologica* **18** 196–223.

[39] SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 888–905.

[40] XU, M., JOG, V. and LOH, P.-L (2020). Supplement to "Optimal rates for community estimation in the weighted stochastic block model." https://doi.org/10.1214/18-AOS1797SUPP.

[41] YUN, S. and PROUTIERE, A. (2016). Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems* 965–973.

[42] 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

[43] ZHANG, B. and HORVATH, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 17, 45. MR2170433 https://doi.org/10.2202/1544-6115.1128

The Annals of Statistics
2020, Vol. 48, No. 1, 205–229
https://doi.org/10.1214/18-AOS1799

ADAPTIVE RISK BOUNDS IN UNIVARIATE TOTAL VARIATION DENOISING AND TREND FILTERING

BY ADITYANAND GUNTUBOYINA[1,*], DONOVAN LIEU[1,**], SABYASACHI CHATTERJEE[2] AND BODHISATTVA SEN[3]

[1]*Department of Statistics, University of California, Berkeley,* [*]*aditya@stat.berkeley.edu;* [**]*dlieu333@berkeley.edu*
[2]*Department of Statistics, University of Illinois at Urbana–Champaign, sc1706@illinois.edu*
[3]*Department of Statistics, Columbia University, bodhi@stat.columbia.edu*

We study trend filtering, a relatively recent method for univariate nonparametric regression. For a given integer $r \geq 1$, the rth order trend filtering estimator is defined as the minimizer of the sum of squared errors when we constrain (or penalize) the sum of the absolute rth order discrete derivatives of the fitted function at the design points. For $r = 1$, the estimator reduces to total variation regularization which has received much attention in the statistics and image processing literature. In this paper, we study the performance of the trend filtering estimator for every $r \geq 1$, both in the constrained and penalized forms. Our main results show that in the strong sparsity setting when the underlying function is a (discrete) spline with few "knots," the risk (under the global squared error loss) of the trend filtering estimator (with an appropriate choice of the tuning parameter) achieves the *parametric* n^{-1}-rate, up to a logarithmic (multiplicative) factor. Our results therefore provide support for the use of trend filtering, for every $r \geq 1$, in the strong sparsity setting.

REFERENCES

[1] ARNOLD, T. B. and TIBSHIRANI, R. J. (2016). Efficient implementations of the generalized lasso dual path algorithm. *J. Comput. Graph. Statist.* **25** 1–27. MR3474034 https://doi.org/10.1080/10618600.2015.1008638

[2] BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. MR3782383 https://doi.org/10.1214/17-AOS1566

[3] 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

[4] BROCKMANN, M., GASSER, T. and HERRMANN, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *J. Amer. Statist. Assoc.* **88** 1302–1309. MR1245363

[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] 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

[7] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089 https://doi.org/10.1093/biomet/81.3.425

[8] 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

[9] FAN, Z. and GUAN, L. (2017). l_0-estimation of piecewise-constant signals on graphs. Preprint. Available at arXiv:1703.01421.

[10] FOYGEL, R. and MACKEY, L. (2014). Corrupted sensing: Novel guarantees for separating structured signals. *IEEE Trans. Inform. Theory* **60** 1223–1247. MR3164972 https://doi.org/10.1109/TIT.2013.2293654

MSC2010 subject classifications. 62G08, 62J05, 62J07.
Key words and phrases. Adaptive splines, discrete splines, fat shattering, higher order total variation regularization, metric entropy bounds, nonparametric function estimation, risk bounds, subdifferential, tangent cone.

[11] FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.* **19** 1–141. With discussion and a rejoinder by the author. MR1091842 https://doi.org/10.1214/aos/1176347963

[12] GAO, C., HAN, F. and ZHANG, C.-H. (2017). Minimax risk bounds for piecewise constant models. Preprint. Available at arXiv:1705.06386.

[13] GOLDENSHLUGER, A. and NEMIROVSKI, A. (1997). On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.* **6** 135–170. MR1466625

[14] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics. Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893

[15] GUNTUBOYINA, A., LIEU, D., CHATTERJEE, S. and SEN, B. (2020). Supplement to "Adaptive risk bounds in univariate total variation denoising and trend filtering." https://doi.org/10.1214/18-AOS1799SUPP.

[16] GUNTUBOYINA, A. and SEN, B. (2017). Nonparametric shape-restricted regression. Preprint. Available at arXiv:1709.05707.

[17] HARCHAOUI, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. MR2796565 https://doi.org/10.1198/jasa.2010.tm09181

[18] JOHNSTONE, I. M. (2015). *Gaussian Estimation: Sequence and Wavelet Models*. Available at http://statweb.stanford.edu/~imj/GE09-08-15.pdf.

[19] 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

[20] KUCZMA, M. (2009). *An Introduction to the Theory of Functional Equations and Inequalities: Cauchy's Equation and Jensen's Inequality*, 2nd ed. Birkhäuser, Basel. Edited and with a preface by Attila Gilányi. MR2467621 https://doi.org/10.1007/978-3-7643-8749-5

[21] LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. MR1447734 https://doi.org/10.1214/aos/1069362731

[22] LÉVY-LEDUC, C. and HARCHAOUI, Z. (2008). Catching change-points with lasso. In *Advances in Neural Information Processing Systems* 617–624.

[23] LIN, K., SHARPNACK, J., RINALDO, A. and TIBSHIRANI, R. J. (2016). Approximate recovery in change-point problems, from ℓ_2 estimation error rates. Preprint. Available at arXiv:1606.06746.

[24] 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

[25] MANGASARIAN, O. L. and SCHUMAKER, L. I. (1971). Discrete splines via mathematical programming. *SIAM J. Control* **9** 174–183. MR0304937

[26] MÜLLER, H.-G. and STADTMÜLLER, U. (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15** 182–201. MR0885731 https://doi.org/10.1214/aos/1176350260

[27] ORTELLI, F. and VAN DE GEER, S. (2018). On the total variation regularized estimator over the branched path graph. Preprint. Available at arXiv:1806.01009.

[28] OYMAK, S. and HASSIBI, B. (2016). Sharp MSE bounds for proximal denoising. *Found. Comput. Math.* **16** 965–1029. MR3529131 https://doi.org/10.1007/s10208-015-9278-4

[29] PINTORE, A., SPECKMAN, P. and HOLMES, C. C. (2006). Spatially adaptive smoothing splines. *Biometrika* **93** 113–125. MR2277744 https://doi.org/10.1093/biomet/93.1.113

[30] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inform. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799

[31] ROCKAFELLAR, R. T. (1970). *Convex Analysis. Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. MR0274683

[32] RUDELSON, M. and VERSHYNIN, R. (2006). Combinatorics of random processes and sections of convex bodies. *Ann. of Math.* (2) **164** 603–648. MR2247969 https://doi.org/10.4007/annals.2006.164.603

[33] 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

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

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

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

[37] VAN DE GEER, S. (2018). On tight bounds for the Lasso. Preprint. Available at arXiv:1804.00989.

[38] 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. MR3543511

[39] WANG, Y.-X., SMOLA, A. J. and TIBSHIRANI, R. J. (2014). The falling factorial basis and its statistical applications. In *ICML* 730–738.

[40] WINKLER, G. and LIEBSCHER, V. (2002). Smoothers for discontinuous signals. *J. Nonparametr. Stat.* **14** 203–222. MR1905594 https://doi.org/10.1080/10485250211388

[41] ZHOU, S. and SHEN, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *J. Amer. Statist. Assoc.* **96** 247–259. MR1952735 https://doi.org/10.1198/016214501750332820

The Annals of Statistics
2020, Vol. 48, No. 1, 230–250
https://doi.org/10.1214/18-AOS1800

SPECTRAL AND MATRIX FACTORIZATION METHODS FOR CONSISTENT COMMUNITY DETECTION IN MULTI-LAYER NETWORKS

BY SUBHADEEP PAUL[1] AND YUGUO CHEN[2]

[1]*Department of Statistics, The Ohio State University, paul.963@osu.edu*
[2]*Department of Statistics, University of Illinois, yuguo@illinois.edu*

We consider the problem of estimating a consensus community structure by combining information from multiple layers of a multi-layer network using methods based on the spectral clustering or a low-rank matrix factorization. As a general theme, these "intermediate fusion" methods involve obtaining a low column rank matrix by optimizing an objective function and then using the columns of the matrix for clustering. However, the theoretical properties of these methods remain largely unexplored. In the absence of statistical guarantees on the objective functions, it is difficult to determine if the algorithms optimizing the objectives will return good community structures. We investigate the consistency properties of the global optimizer of some of these objective functions under the multi-layer stochastic blockmodel. For this purpose, we derive several new asymptotic results showing consistency of the intermediate fusion techniques along with the spectral clustering of mean adjacency matrix under a high dimensional setup, where the number of nodes, the number of layers and the number of communities of the multi-layer graph grow. Our numerical study shows that the intermediate fusion techniques outperform late fusion methods, namely spectral clustering on aggregate spectral kernel and module allegiance matrix in sparse networks, while they outperform the spectral clustering of mean adjacency matrix in multi-layer networks that contain layers with both homophilic and heterophilic communities.

REFERENCES

[1] BARBILLON, P., DONNET, S., LAZEGA, E. and BAR-HEN, A. (2017). Stochastic block models for multiplex networks: An application to a multilevel network of researchers. *J. Roy. Statist. Soc. Ser. A* **180** 295–314. MR3600512 https://doi.org/10.1111/rssa.12193

[2] BASSETT, D. S., PORTER, M. A., WYMBS, N. F., GRAFTON, S. T., CARLSON, J. M. and MUCHA, P. J. (2013). Robust detection of dynamic community structure in networks. *Chaos* **23** 013142, 16. MR3389633 https://doi.org/10.1063/1.4790830

[3] BASSETT, D. S., WYMBS, N. F., PORTER, M. A., MUCHA, P. J., CARLSON, J. M. and GRAFTON, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. USA* **108** 7641–7646.

[4] BOCCALETTI, S., BIANCONI, G., CRIADO, R., DEL GENIO, C. I., GÓMEZ-GARDEÑES, J., ROMANCE, M., SENDIÑA-NADAL, I., WANG, Z. and ZANIN, M. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* **544** 1–122. MR3270140 https://doi.org/10.1016/j.physrep.2014.07.001

[5] BRAUN, U., SCHÄFER, A., WALTER, H., ERK, S., ROMANCZUK-SEIFERTH, N., HADDAD, L., SCHWEIGER, J. I., GRIMM, O., HEINZ, A. et al. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl. Acad. Sci. USA* **112** 11678–11683.

[6] CHEN, P.-Y. and HERO III, A. O. (2016). Multilayer spectral graph clustering via convex layer aggregation. Preprint. Available at arXiv:1609.07200.

[7] CHI, Y., SONG, X., ZHOU, D., HINO, K. and TSENG, B. L. (2007). Evolutionary spectral clustering by incorporating temporal smoothness. In *Proceedings of the* 13*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 153–162.

MSC2010 subject classifications. Primary 62F12, 62H30; secondary 90B15, 15A23.
Key words and phrases. Community detection, consistency, co-regularized spectral clustering, multi-layer networks, multi-layer stochastic block model, orthogonal linked matrix factorization.

[8] CHUNG, F. and RADCLIFFE, M. (2011). On the spectra of general random graphs. *Electron. J. Combin.* **18** Paper 215, 14. MR2853072

[9] DONG, X., FROSSARD, P., VANDERGHEYNST, P. and NEFEDOV, N. (2012). Clustering with multi-layer graphs: A spectral perspective. *IEEE Trans. Signal Process.* **60** 5820–5831. MR2990287 https://doi.org/10.1109/TSP.2012.2212886

[10] GHASEMIAN, A., ZHANG, P., CLAUSET, A., MOORE, C. and PEEL, L. (2016). Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Phys. Rev. X* **6** 031005.

[11] GREENE, D. and CUNNINGHAM, P. (2013). Producing a unified graph representation from multiple social network views. *ACM Web Sci.* **2** 129–233.

[12] HAN, Q., XU, K. S. and AIROLDI, E. M. (2015). Consistent estimation of dynamic and multi-layer block models. In *Proceedings of the* 32*nd International Conference on Machine Learning* 1511–1520.

[13] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[14] KIVELÄ, M., ARENAS, A., BARTHELEMY, M., GLEESON, J. P., MORENO, Y. and PORTER, M. A. (2014). Multilayer networks. *J. Complex Netw.* **2** 203–271.

[15] KUMAR, A., RAI, P. and DAUME, H. (2011). Co-regularized multi-view spectral clustering. In *Advances in Neural Information Processing Systems* 1413–1421.

[16] 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

[17] LIU, J., WANG, C., GAO, J. and HAN, J. (2013). Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the* 2013 *SIAM International Conference on Data Mining* 252–260.

[18] LONG, B., ZHANG, Z. M., WU, X. and YU, P. S. (2006). Spectral clustering for multi-type relational data. In *Proceedings of the* 23*rd International Conference on Machine Learning* 585–592.

[19] MUCHA, P. J., RICHARDSON, T., MACON, K., PORTER, M. A. and ONNELA, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328** 876–878. MR2662590 https://doi.org/10.1126/science.1184819

[20] NEWMAN, M. E. J. and GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* **69** 026113.

[21] 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.

[22] NICKEL, M., TRESP, V. and KRIEGEL, H. P. (2011). A three-way model for collective learning on multi-relational data. In *Proceedings of* 28*th International Conference on Machine Learning* 809–816.

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

[24] PAUL, S. and CHEN, Y. (2016). Null models and modularity based community detection in multi-layer networks. Preprint. Available at arXiv:1608.00623.

[25] PAUL, S. and CHEN, Y. (2020). Supplement to "Spectral and matrix factorization methods for consistent community detection in multi-layer networks." https://doi.org/10.1214/18-AOS1800SUPP.

[26] PEIXOTO, T. P. (2015). Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Phys. Rev. E* **92** 042807.

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

[28] 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

[29] SINGH, A. P. and GORDON, G. J. (2008). Relational learning via collective matrix factorization. In *Proceedings of the* 14*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 650–658.

[30] STANLEY, N., SHAI, S., TAYLOR, D. and MUCHA, P. J. (2016). Clustering network layers with the strata multilayer stochastic block model. *IEEE Trans. Netw. Sci. Eng.* **3** 95–105. MR3515211 https://doi.org/10.1109/TNSE.2016.2537545

[31] STEWART, G. W. and SUN, J. G. (1990). *Matrix Perturbation Theory*. *Computer Science and Scientific Computing*. Academic Press, Boston, MA. MR1061154

[32] SUN, S. (2013). A survey of multi-view machine learning. *Neural Comput. Appl.* **23** 2031–2038.

[33] TANG, L., WANG, X. and LIU, H. (2012). Community detection via heterogeneous interaction analysis. *Data Min. Knowl. Discov.* **25** 1–33. MR2911723 https://doi.org/10.1007/s10618-011-0231-0

[34] TANG, W., LU, Z. and DHILLON, I. S. (2009). Clustering with multiple graphs. In *Proceedings of the Ninth IEEE International Conference on Data Mining* 1016–1021.

[35] TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z

[36] VALLES-CATALA, T., MASSUCCI, F. A., GUIMERA, R. and SALES-PARDO, M. (2016). Multilayer stochastic block models reveal the multilayer structure of complex networks. *Phys. Rev. X* **6** 011036.

[37] WILSON, J. D., PALOWITCH, J., BHAMIDI, S. and NOBEL, A. B. (2017). Community extraction in multilayer networks with heterogeneous community structure. *J. Mach. Learn. Res.* **18** Paper No. 149, 49. MR3763783

[38] XU, C., TAO, D. and XU, C. (2013). A survey on multi-view learning. Preprint. Available at arXiv:1304.5634.

[39] ZHOU, D. and BURGES, C. J. (2007). Spectral clustering and transductive learning with multiple views. In *Proceedings of the* 24*th International Conference on Machine Learning* 1159–1166.

[40] ŽITNIK, M. and ZUPAN, B. (2015). Data fusion by matrix factorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 41–53. https://doi.org/10.1109/TPAMI.2014.2343973

The Annals of Statistics
2020, Vol. 48, No. 1, 251–273
https://doi.org/10.1214/18-AOS1801

STATISTICAL INFERENCE FOR MODEL PARAMETERS IN STOCHASTIC GRADIENT DESCENT

BY XI CHEN[1,*], JASON D. LEE[2], XIN T. TONG[3] AND YICHEN ZHANG[1,**]

[1]*Department of Technology, Operations, and Statistics, Stern School of Business, New York University,* [*]*xchen3@stern.nyu.edu;* [**]*yzhang@stern.nyu.edu*

[2]*Data Sciences and Operations, Marshall School of Business, University of Southern California, jasonlee@marshall.usc.edu*

[3]*Department of Mathematics, National University of Singapore, mattxin@nus.edu.sg*

The stochastic gradient descent (SGD) algorithm has been widely used in statistical estimation for large-scale data due to its computational and memory efficiency. While most existing works focus on the convergence of the objective function or the error of the obtained solution, we investigate the problem of statistical inference of true model parameters based on SGD when the population loss function is strongly convex and satisfies certain smoothness conditions.

Our main contributions are twofold. First, in the fixed dimension setup, we propose two consistent estimators of the asymptotic covariance of the average iterate from SGD: (1) a plug-in estimator, and (2) a batch-means estimator, which is computationally more efficient and only uses the iterates from SGD. Both proposed estimators allow us to construct asymptotically exact confidence intervals and hypothesis tests.

Second, for high-dimensional linear regression, using a variant of the SGD algorithm, we construct a debiased estimator of each regression coefficient that is asymptotically normal. This gives a one-pass algorithm for computing both the sparse regression coefficients and confidence intervals, which is computationally attractive and applicable to online data.

REFERENCES

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Stochastic optimization and sparse statistical recovery: Optimal algorithms for high dimensions. In *Proceedings of the Advances in Neural Information Processing Systems*.

BACH, F. and MOULINES, E. (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Proceedings of the Advances in Neural Information Processing Systems*.

BACH, F. and MOULINES, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$. In *Proceedings of the Advances in Neural Information Processing Systems*.

BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410

BÜHLMANN, P. and MANDOZZI, J. (2014). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Comput*. *Statist*. **29** 407–430. MR3261821 https://doi.org/10.1007/s00180-013-0436-3

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

BUJA, A., BERK, R., BROWN, L., GEORGE, E., TRASKIN, M., ZHANG, K. and ZHAO, L. (2013). A conspiracy of random X and model violation against classical inference in linear regression. Technical Report, Dept. Statistics, The Wharton School, Univ. Pennsylvania, Philadelphia, PA.

CHEN, X., LEE, J. D., TONG, X. T. and ZHANG, Y. (2020). Supplement to "Statistical inference for model parameters in stochastic gradient descent." https://doi.org/10.1214/18-AOS1801SUPP.

DAMERDJI, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Manage*. *Sci*. **37** 1424–1440.

MSC2010 subject classifications. Primary 62J10, 62M02; secondary 60K35.

Key words and phrases. Stochastic gradient descent, asymptotic variance, batch-means estimator, high-dimensional inference, time-inhomogeneous Markov chain.

FABIAN, V. (1968). On asymptotic normality in stochastic approximation. *Ann. Math. Stat.* **39** 1327–1332. MR0231429 https://doi.org/10.1214/aoms/1177698258

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

FISHMAN, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications. Springer Series in Operations Research*. Springer, New York. MR1392474 https://doi.org/10.1007/978-1-4757-2553-7

FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704 https://doi.org/10.1214/09-AOS735

GEYER, C. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7** 473–483.

GHADIMI, S. and LAN, G. (2012). Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework. *SIAM J. Optim.* **22** 1469–1492. MR3023780 https://doi.org/10.1137/110848864

GLYNN, P. W. and IGLEHART, D. L. (1990). Simulation output analysis using standardized time series. *Math. Oper. Res.* **15** 1–16. MR1038232 https://doi.org/10.1287/moor.15.1.1

GLYNN, P. W. and WHITT, W. (1991). Estimating the asymptotic variance with batch means. *Oper. Res. Lett.* **10** 431–435. MR1141337 https://doi.org/10.1016/0167-6377(91)90019-L

JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478 https://doi.org/10.1198/016214506000000492

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584 https://doi.org/10.1198/jasa.2009.tm08647

NEMIROVSKI, A., JUDITSKY, A., LAN, G. and SHAPIRO, A. (2008). Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19** 1574–1609. MR2486041 https://doi.org/10.1137/070704277

NESTEROV, YU. and VIAL, J.-PH. (2008). Confidence level solutions for stochastic programming. *Automatica J. IFAC* **44** 1559–1568. MR2531843 https://doi.org/10.1016/j.automatica.2008.01.017

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814 https://doi.org/10.1137/0330046

RAKHLIN, A., SHAMIR, O. and SRIDHARAN, K. (2012). Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the International Conference on Machine Learning*.

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

ROUX, N. L., SCHMIDT, M. and BACH, F. (2012). A stochastic gradient method with an exponential convergence rate for strongly-convex optimization with finite training sets. In *Proceedings of the Advances in Neural Information Processing Systems*.

RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins–Monro process. Technical Report, Dept. Operations Research and Industrial Engineering, Cornell Univ.

SREBRO, N. and TEWARI, A. (2010). Stochastic optimization for machine learning. Tutorial at *International Conference on Machine Learning*.

SULLIVAN, T. J. (2015). *Introduction to Uncertainty Quantification. Texts in Applied Mathematics* **63**. Springer, Cham. MR3364576 https://doi.org/10.1007/978-3-319-23395-6

TOULIS, P. and AIROLDI, E. M. (2017). Asymptotic and finite-sample properties of estimators based on stochastic gradients. *Ann. Statist.* **45** 1694–1727. MR3670193 https://doi.org/10.1214/16-AOS1506

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

XIAO, L. (2010). Dual averaging methods for regularized stochastic learning and online optimization. *J. Mach. Learn. Res.* **11** 2543–2596. MR2738777

XIAO, L. and ZHANG, T. (2014). A proximal stochastic gradient method with progressive variance reduction. *SIAM J. Optim.* **24** 2057–2075. MR3285905 https://doi.org/10.1137/140961791

ZHANG, T. (2004). Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the International Conference on Machine Learning*.

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

The Annals of Statistics
2020, Vol. 48, No. 1, 274–299
https://doi.org/10.1214/18-AOS1803

BOOTSTRAP CONFIDENCE REGIONS BASED ON M-ESTIMATORS UNDER NONSTANDARD CONDITIONS

BY STEPHEN M. S. LEE[1] AND PUYUDI YANG[2]

[1]*Department of Statistics and Actuarial Science, University of Hong Kong, smslee@hku.hk*
[2]*Department of Statistics, University of California, Davis, pydyang@ucdavis.edu*

Suppose that a confidence region is desired for a subvector θ of a multidimensional parameter $\xi = (\theta, \psi)$, based on an M-estimator $\hat{\xi}_n = (\hat{\theta}_n, \hat{\psi}_n)$ calculated from a random sample of size n. Under nonstandard conditions $\hat{\xi}_n$ often converges at a nonregular rate r_n, in which case consistent estimation of the distribution of $r_n(\hat{\theta}_n - \theta)$, a pivot commonly chosen for confidence region construction, is most conveniently effected by the m out of n bootstrap. The above choice of pivot has three drawbacks: (i) the shape of the region is either subjectively prescribed or controlled by a computationally intensive depth function; (ii) the region is not transformation equivariant; (iii) $\hat{\xi}_n$ may not be uniquely defined. To resolve the above difficulties, we propose a one-dimensional pivot derived from the criterion function, and prove that its distribution can be consistently estimated by the m out of n bootstrap, or by a modified version of the perturbation bootstrap. This leads to a new method for constructing confidence regions which are transformation equivariant and have shapes driven solely by the criterion function. A subsampling procedure is proposed for selecting m in practice. Empirical performance of the new method is illustrated with examples drawn from different nonstandard M-estimation settings. Extension of our theory to row-wise independent triangular arrays is also explored.

REFERENCES

[1] ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245 https://doi.org/10.1111/j.1468-0262.2005.00613.x

[2] ANDREWS, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405. MR1748009 https://doi.org/10.1111/1468-0262.00114

[3] BERAN, R. (1988). Balanced simultaneous confidence sets. *J. Amer. Statist. Assoc.* **83** 679–686. MR0963795

[4] BICKEL, P. J. and SAKOV, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statist. Sinica* **18** 967–985. MR2440400

[5] BOSE, A. and CHATTERJEE, S. (2001). Generalised bootstrap in non-regular M-estimation problems. *Statist. Probab. Lett.* **55** 319–328. MR1867535 https://doi.org/10.1016/S0167-7152(01)00161-4

[6] CHANG, C. and TODD OGDEN, R. (2009). Bootstrapping sums of independent but not identically distributed continuous processes with applications to functional data. *J. Multivariate Anal.* **100** 1291–1303. MR2508388 https://doi.org/10.1016/j.jmva.2008.11.008

[7] CHATTERJEE, A. and LAHIRI, S. N. (2013). Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. *Ann. Statist.* **41** 1232–1259. MR3113809 https://doi.org/10.1214/13-AOS1106

[8] CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414–436. MR2157808 https://doi.org/10.1214/009053604000000904

[9] CHENG, G. and HUANG, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.* **38** 2884–2915. MR2722459 https://doi.org/10.1214/10-AOS809

[10] CHEUNG, K. Y., LEE, S. M. S. and YOUNG, G. A. (2006). Stein confidence sets based on non-iterated and iterated parametric bootstraps. *Statist. Sinica* **16** 45–75. MR2256079

[11] COHEN, A. and SACKROWITZ, H. B. (2004). A discussion of some inference issues in order restricted models. *Canad. J. Statist.* **32** 199–205. MR2064402 https://doi.org/10.2307/3315943

MSC2010 subject classifications. 62G15, 62G09.
Key words and phrases. Confidence region, m out of n bootstrap, M-estimator, pivot, subsampling.

[12] DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. MR3713586 https://doi.org/10.1007/s11749-017-0554-2

[13] GHOSH, S. and POLANSKY, A. M. (2014). Smoothed and iterated bootstrap confidence regions for parameter vectors. *J. Multivariate Anal*. **132** 171–182. MR3266268 https://doi.org/10.1016/j.jmva.2014.08.003

[14] GÖTZE, F. and RAČKAUSKAS, A. (2001). Adaptive choice of bootstrap sample sizes. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 286–309. IMS, Beachwood, OH. MR1836566 https://doi.org/10.1214/lnms/1215090074

[15] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738

[16] HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist*. **37** 1079–1111. MR2509068 https://doi.org/10.1214/07-AOS555

[17] JIN, Z., YING, Z. and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88** 381–390. MR1844838 https://doi.org/10.1093/biomet/88.2.381

[18] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist*. **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[19] KOSOROK, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal*. **84** 299–318. MR1965224 https://doi.org/10.1016/S0047-259X(02)00040-4

[20] LAHIRI, S. N. (1992). On bootstrapping M-estimators. *Sankhyā Ser. A* **54** 157–170. MR1192092

[21] LAI, P. Y. and LEE, S. M. S. (2005). An overview of asymptotic properties of L_p regression under general classes of error distributions. *J. Amer. Statist. Assoc*. **100** 446–458. MR2160549 https://doi.org/10.1198/016214504000001385

[22] LEE, S. M. S. and YANG, P. (2019). Supplement to "Bootstrap confidence regions based on M-estimators under nonstandard conditions." https://doi.org/10.1214/18-AOS1803SUPP.

[23] LEE, S. M. S. and PUN, M. C. (2006). On m out of n bootstrapping for nonstandard M-estimation with nuisance parameters. *J. Amer. Statist. Assoc*. **101** 1185–1197. MR2328306 https://doi.org/10.1198/016214506000000014

[24] LÉGER, C. and MACGIBBON, B. (2006). On the bootstrap in cube root asymptotics. *Canad. J. Statist*. **34** 29–44. MR2267708 https://doi.org/10.1002/cjs.5550340104

[25] LI, Z., TAYLOR, J. M. G. and NAN, B. (2010). Construction of confidence intervals and regions for ordered binomial probabilities. *Amer. Statist*. **64** 291–298. MR2758560 https://doi.org/10.1198/tast.2010.09096

[26] LIU, H. and YU, B. (2013). Asymptotic properties of Lasso+mLS and Lasso+Ridge in sparse high-dimensional linear regression. *Electron. J. Stat*. **7** 3124–3169. MR3151764 https://doi.org/10.1214/14-EJS875

[27] MA, S. and KOSOROK, M. R. (2005). Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Anal*. **96** 190–217. MR2202406 https://doi.org/10.1016/j.jmva.2004.09.008

[28] MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. MR0436905 https://doi.org/10.1016/0304-4076(75)90032-9

[29] MEERSCHAERT, M. M. (1988). Regular variation in $\mathbf{R}^k$. *Proc. Amer. Math. Soc*. **102** 341–348. MR0920997 https://doi.org/10.2307/2045886

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

[31] PARK, Y., KALBFLEISCH, J. D. and TAYLOR, J. M. G. (2014). Confidence intervals under order restrictions. *Statist. Sinica* **24** 429–445. MR3183692

[32] PATRA, R. K., SEIJO, E. and SEN, B. (2018). A consistent bootstrap procedure for the maximum score estimator. *J. Econometrics* **205** 488–507. MR3813528 https://doi.org/10.1016/j.jeconom.2018.04.001

[33] POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

[34] ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc*. **79** 871–880. MR0770281

[35] SPOKOINY, V. and ZHILOVA, M. (2015). Bootstrap confidence sets under model misspecification. *Ann. Statist*. **43** 2653–2675. MR3405607 https://doi.org/10.1214/15-AOS1355

[36] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

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

[38] WEI, B. and LEE, S. M. S. (2012). Second-order accuracy of depth-based bootstrap confidence regions. *J. Multivariate Anal.* **105** 112–123. MR2877506 https://doi.org/10.1016/j.jmva.2011.08.016

[39] YEH, A. B. and SINGH, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *J. Roy. Statist. Soc. Ser. B* **59** 639–652. MR1452031 https://doi.org/10.1111/1467-9868.00088

[40] ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

[41] ZHOU, Q. and MIN, S. (2017). Uncertainty quantification under group sparsity. *Biometrika* **104** 613–632. MR3694586 https://doi.org/10.1093/biomet/asx037

The Annals of Statistics
2020, Vol. 48, No. 1, 300–323
https://doi.org/10.1214/18-AOS1804

SPARSE HIGH-DIMENSIONAL REGRESSION: EXACT SCALABLE ALGORITHMS AND PHASE TRANSITIONS

BY DIMITRIS BERTSIMAS[*] AND BART VAN PARYS[**]

Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, [*]dbertsim@mit.edu; [**]vanparys@mit.edu

We present a novel binary convex reformulation of the sparse regression problem that constitutes a new duality perspective. We devise a new cutting plane method and provide evidence that it can solve to provable optimality the sparse regression problem for sample sizes n and number of regressors p in the 100,000s, that is, two orders of magnitude better than the current state of the art, in seconds. The ability to solve the problem for very high dimensions allows us to observe new phase transition phenomena. Contrary to traditional complexity theory which suggests that the difficulty of a problem increases with problem size, the sparse regression problem has the property that as the number of samples n increases the problem becomes easier in that the solution recovers 100% of the true signal, and our approach solves the problem extremely fast (in fact faster than `Lasso`), while for small number of samples n, our approach takes a larger amount of time to solve the problem, but importantly the optimal solution provides a statistically more relevant regressor. We argue that our exact sparse regression approach presents a superior alternative over heuristic methods available at present.

REFERENCES

BARNHART, C., JOHNSON, E. L., NEMHAUSER, G. L., SAVELSBERGH, M. W. P. and VANCE, P. H. (1998). Branch-and-price: Column generation for solving huge integer programs. *Oper. Res.* **46** 316–329. MR1663050 https://doi.org/10.1287/opre.46.3.316

BERTSIMAS, D. and FERTIS, A. (2009). On the equivalence of robust optimization and regularization in statistics Technical report, MIT, Cambridge, MA.

BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

BERTSIMAS, D., PAUPHILET, J. and VAN PARYS, B. (2019). Sparse regression: Scalable algorithms and empirical performance. *Statist. Sci.* To appear.

BERTSIMAS, D. and VAN PARYS, B. (2019). Supplement to "Sparse high-dimensional regression: Exact scalable algorithms and phase transitions." https://doi.org/10.1214/18-AOS1804SUPP.

BIXBY, R. E. (2012). A brief history of linear and mixed-integer programming computation. *Doc. Math. Extra vol.: Optimization stories* 107–121. MR2991475

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

DONOHO, D. and STODDEN, V. (2006). Breakdown point of model selection when the number of variables exceeds the number of observations. In *International Joint Conference on Neural Networks* 1916–1921. IEEE, New York.

DONOHO, D. and TANNER, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4273–4293. MR2546388 https://doi.org/10.1098/rsta.2009.0152

DURAN, M. A. and GROSSMANN, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math. Program.* **36** 307–339. MR0866413 https://doi.org/10.1007/BF02592064

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

MSC2010 subject classifications. Primary 62J07; secondary 90C10.
Key words and phrases. Best subset selection, sparse regression, kernel learning, integer optimization, convex optimization.

FLETCHER, R. and LEYFFER, S. (1994). Solving mixed integer nonlinear programs by outer approximation. *Math. Program.* **66** 327–349. MR1297070 https://doi.org/10.1007/BF01581153

FRIEDMAN, J., HASTIE, T. J. and TIBSHIRANI, R. J. (2013). GLMNet: Lasso and elastic-net regularized generalized linear models. R package version 1.9–5.

FURNIVAL, G. M. and WILSON, R. W. (2000). Regressions by leaps and bounds. *Technometrics* **42** 69–79.

GAMARNIK, D. and ZADIK, I. (2017). High-Dimensional Regression with Binary Coefficients. Preprint. Available at arXiv:1701.04455.

HAGER, W. W. (1989). Updating the inverse of a matrix. *SIAM Rev.* **31** 221–239. MR0997457 https://doi.org/10.1137/1031049

HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity*: *The Lasso and Generalizations*. *Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. MR3616141

LAWLER, E. L. and WOOD, D. E. (1966). Branch-and-bound methods: A survey. *Oper. Res.* **14** 699–719. MR0202469 https://doi.org/10.1287/opre.14.4.699

MALLAT, S. G. and ZHANG, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **41** 3397–3415.

SCHÖLKOPF, B. and SMOLA, A. J. (2002). *Learning with Kernels*: *Support Vector Machines*, *Regularization*, *Optimization*, *and Beyond*. MIT press, Cambridge, MA.

SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. MR0097026

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

TIKHONOV, A. N. (1943). On the stability of inverse problems. *Dokl. Akad. Nauk SSSR* **39** 195–198.

VAPNIK, V. (1998). The support vector method of function estimation. In *Nonlinear Modeling* 55–85. Springer, Berlin.

WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

WANG, M., XU, W. and TANG, A. (2011). On the performance of sparse recovery via ℓ_p-minimization $(0 \leq p \leq 1)$. *IEEE Trans. Inform. Theory* **57** 7255–7278. MR2883654 https://doi.org/10.1109/TIT.2011.2159959

XU, H., CARAMANIS, C. and MANNOR, S. (2009). Robustness and regularization of support vector machines. *J. Mach. Learn. Res.* **10** 1485–1510. MR2534869

ZHANG, F., ed. (2006). *The Schur Complement and Its Applications*. *Numerical Methods and Algorithms* **4**. Springer, New York. MR2160825 https://doi.org/10.1007/b105056

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

ZHANG, Y., WAINRIGHT, M. and JORDAN, M. (2014). Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. *J. Mach. Learn. Res.* **35** 1–28.

ZHENG, L., MALEKI, A., WANG, X. and LONG, T. (2015). Does ℓ_p-minimization outperform ℓ_1-minimization? Available at arXiv:1501.03704.

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

The Annals of Statistics
2020, Vol. 48, No. 1, 324–345
https://doi.org/10.1214/18-AOS1805

TESTING FOR PRINCIPAL COMPONENT DIRECTIONS UNDER WEAK IDENTIFIABILITY

BY DAVY PAINDAVEINE[*], JULIEN REMY[**] AND THOMAS VERDEBOUT[†]

ECARES and Département de Mathématique, Université libre de Bruxelles, [*]*dpaindav@ulb.ac.be;* [**]*Julien.Remy@vub.ac.be;* [†]*tverdebo@ulb.ac.be*

We consider the problem of testing, on the basis of a p-variate Gaussian random sample, the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ against the alternative $\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0$, where $\boldsymbol{\theta}_1$ is the "first" eigenvector of the underlying covariance matrix and $\boldsymbol{\theta}_1^0$ is a fixed unit p-vector. In the classical setup where eigenvalues $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_p$ are fixed, the Anderson (*Ann. Math. Stat.* **34** (1963) 122–148) likelihood ratio test (LRT) and the Hallin, Paindaveine and Verdebout (*Ann. Statist.* **38** (2010) 3245–3299) Le Cam optimal test for this problem are asymptotically equivalent under the null hypothesis, hence also under sequences of contiguous alternatives. We show that this equivalence does not survive asymptotic scenarios where $\lambda_{n1}/\lambda_{n2} = 1 + O(r_n)$ with $r_n = O(1/\sqrt{n})$. For such scenarios, the Le Cam optimal test still asymptotically meets the nominal level constraint, whereas the LRT severely overrejects the null hypothesis. Consequently, the former test should be favored over the latter one whenever the two largest sample eigenvalues are close to each other. By relying on the Le Cam's asymptotic theory of statistical experiments, we study the non-null and optimality properties of the Le Cam optimal test in the aforementioned asymptotic scenarios and show that the null robustness of this test is not obtained at the expense of power. Our asymptotic investigation is extensive in the sense that it allows r_n to converge to zero at an arbitrary rate. While we restrict to single-spiked spectra of the form $\lambda_{n1} > \lambda_{n2} = \cdots = \lambda_{np}$ to make our results as striking as possible, we extend our results to the more general elliptical case. Finally, we present an illustrative real data example.

REFERENCES

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 https://doi.org/10.1214/aoms/1177704248

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

ATKINSON, A. C., RIANI, M. and CERIOLI, A. (2004). *Exploring Multivariate Data with the Forward Search. Springer Series in Statistics*. Springer, New York. MR2055967 https://doi.org/10.1007/978-0-387-21840-3

BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

BALI, J. L., BOENTE, G., TYLER, D. E. and WANG, J.-L. (2011). Robust functional principal components: A projection-pursuit approach. *Ann. Statist.* **39** 2852–2882. MR3012394 https://doi.org/10.1214/11-AOS923

BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

BOENTE, G. and FRAIMAN, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.* **48** 335–345. MR1771495 https://doi.org/10.1016/S0167-7152(00)00014-6

BURMAN, P. and POLONIK, W. (2009). Multivariate mode hunting: Data analytic tools with measures of significance. *J. Multivariate Anal.* **100** 1198–1218. MR2508381 https://doi.org/10.1016/j.jmva.2008.10.015

MSC2010 subject classifications. Primary 62F05, 62H25; secondary 62E20.

Key words and phrases. Le Cam's asymptotic theory of statistical experiments, local asymptotic normality, principal component analysis, spiked covariance matrices, weak identifiability.

CROUX, C. and HAESBROECK, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika* **87** 603–618. MR1789812 https://doi.org/10.1093/biomet/87.3.603

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

DUFOUR, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* **65** 1365–1387. MR1604305 https://doi.org/10.2307/2171740

DUFOUR, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *J. Econometrics* **133** 443–477. MR2252905 https://doi.org/10.1016/j.jeconom.2005.06.007

FLURY, B. (1988). *Common Principal Components and Related Multivariate Models*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0986245 https://doi.org/10.1007/978-3-642-01932-6_28

FLURY, B. and RIEDWYL, H. (1988). *Multivariate Statistics*: *A Practical Approach*. CRC Press, London.

FORCHINI, G. and HILLIER, G. (2003). Conditional inference for possibly unidentified structural equations. *Econometric Theory* **19** 707–743. MR2002577 https://doi.org/10.1017/S0266466603195011

FRITZ, H., GARCÍA-ESCUDERO, L. A. and MAYO-ISCAR, A. (2012). tclust: An R package for a trimming approach to cluster analysis. *J. Stat. Softw.* **47**.

GIROLAMI, M. (1999). *Self-Organizing Neural Networks*. *Independent Component Analysis and Blind Source Separation*. Springer, London.

HALLIN, M. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465 https://doi.org/10.1214/009053606000000731

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38** 3245–3299. MR2766852 https://doi.org/10.1214/10-AOS810

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2014). Efficient R-estimation of principal and common principal components. *J. Amer. Statist. Assoc.* **109** 1071–1083. MR3265681 https://doi.org/10.1080/01621459.2014.880057

HAN, F. and LIU, H. (2014). Scale-invariant sparse PCA on high-dimensional meta-elliptical data. *J. Amer. Statist. Assoc.* **109** 275–287. MR3180563 https://doi.org/10.1080/01621459.2013.844699

HÄRDLE, W. and SIMAR, L. (2007). *Applied Multivariate Statistical Analysis*, 2nd ed. Springer, Berlin. MR2367300

HE, R., HU, B.-G., ZHENG, W.-S. and KONG, X.-W. (2011). Robust principal component analysis based on maximum correntropy criterion. *IEEE Trans. Image Process.* **20** 1485–1494. MR2828599 https://doi.org/10.1109/TIP.2010.2103949

HUBERT, M., ROUSSEEUW, P. J. and VANDEN BRANDEN, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics* **47** 64–79. MR2135793 https://doi.org/10.1198/004017004000000563

JACKSON, J. E. (2005). *A User's Guide to Principal Components*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ.

JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11** 818–887. MR1458943 https://doi.org/10.1017/S0266466600009907

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

JOLICOEUR, P. (1984). Principal components, factor analysis, and multivariate allometry: A small-sample direction test. *Biometrics* **40** 685–690.

KOCH, I. (2014). *Analysis of Multivariate and High-Dimensional Data*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, New York. MR3154467

MAGNUS, J. R. and NEUDECKER, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Wiley, Chichester. MR1698873

MURRAY, P. M., BROWNE, R. P. and MCNICHOLAS, P. D. (2016). uskewFactors: Model-based clustering via mixtures of unrestricted skew-t sactor analyzer models. R package. Available at https://cran.r-project.org/web/packages/uskewFactors/index.html.

PAINDAVEINE, D., REMY, J. and VERDEBOUT, T. (2019). Supplement to "Testing for principal component directions under weak identifiability." https://doi.org/10.1214/18-AOS1805SUPP.

PAINDAVEINE, D. and VERDEBOUT, T. (2017). Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities. *Ann. Statist.* **45** 800–832. MR3650401 https://doi.org/10.1214/16-AOS1468

PÖTSCHER, B. M. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* **70** 1035–1065. MR1910411 https://doi.org/10.1111/1468-0262.00318

ROUSSAS, G. G. and BHATTACHARYA, D. (2011). Revisiting local asymptotic normality (LAN) and passing on to local asymptotic mixed normality (LAMN) and local asymptotic quadratic (LAQ) experiments. In *Advances in Directional and Linear Statistics* (M. T. Wells and A. Sengupta, eds.) 253–280. Physica-Verlag/Springer, Heidelberg. MR2767545 https://doi.org/10.1007/978-3-7908-2628-9_17

SALIBIÁN-BARRERA, M., VAN AELST, S. and WILLEMS, G. (2006). Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. *J. Amer. Statist. Assoc.* **101** 1198–1211. MR2328307 https://doi.org/10.1198/016214506000000096

SCHWARTZMAN, A., MASCARENHAS, W. F. and TAYLOR, J. E. (2008). Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann. Statist.* **36** 2886–2919. MR2485016 https://doi.org/10.1214/08-AOS628

SHINMURA, S. (2016). *New Theory of Discriminant Analysis After R. Fisher*. Springer, Singapore. MR3617625 https://doi.org/10.1007/978-981-10-2164-0

SYLVESTER, A. D., KRAMER, P. A. and JUNGERS, W. L. (2008). Modern humans are not (quite) isometric. *Amer. J. Phys. Anthropol.* **137** 371–383.

TYLER, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736. MR0619278

TYLER, D. E. (1983). A class of asymptotic tests for principal component vectors. *Ann. Statist.* **11** 1243–1250. MR0720269 https://doi.org/10.1214/aos/1176346337

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

The Annals of Statistics
2020, Vol. 48, No. 1, 346–373
https://doi.org/10.1214/19-AOS1809

THE MULTI-ARMED BANDIT PROBLEM: AN EFFICIENT NONPARAMETRIC SOLUTION

BY HOCK PENG CHAN

Department of Statistics and Applied Probability, National University of Singapore, stachp@nus.edu.sg

Lai and Robbins (*Adv. in Appl. Math.* **6** (1985) 4–22) and Lai (*Ann. Statist.* **15** (1987) 1091–1114) provided efficient parametric solutions to the multi-armed bandit problem, showing that arm allocation via upper confidence bounds (UCB) achieves minimum regret. These bounds are constructed from the Kullback–Leibler information of the reward distributions, estimated from specified parametric families. In recent years, there has been renewed interest in the multi-armed bandit problem due to new applications in machine learning algorithms and data analytics. Nonparametric arm allocation procedures like ϵ-greedy, Boltzmann exploration and BESA were studied, and modified versions of the UCB procedure were also analyzed under nonparametric settings. However, unlike UCB these nonparametric procedures are not efficient under general parametric settings. In this paper, we propose efficient nonparametric procedures.

REFERENCES

AGRAWAL, R. (1995). Sample mean based index policies with $O(\log n)$ regret for the multi-armed bandit problem. *Adv. in Appl. Probab.* **27** 1054–1078. MR1358906 https://doi.org/10.2307/1427934

AGRAWAL, R., TENEKETZIS, D. and ANANTHARAM, V. (1989). Asymptotically efficient adaptive allocation schemes for controlled Markov chains: Finite parameter space. *IEEE Trans. Automat. Control* **34** 1249–1259. MR1029375 https://doi.org/10.1109/9.40770

AUER, P., CESA-BIANCHI, N. and FISCHER, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **47** 235–256.

BARANSI, A., MAILLARD, O. A. and MANNOR, S. (2014). Sub-sampling for multi-armed bandits. In *Proceedings of the European Conference on Machine Learning* 13.

BERRY, D. (1972). A Bernoulli two-armed bandit. *Ann. Math. Stat.* **43** 871–897.

BERRY, D. A. and FRISTEDT, B. (1985). *Bandit Problems: Sequential Allocation of Experiments. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0813698 https://doi.org/10.1007/978-94-015-3711-7

BREZZI, M. and LAI, T. L. (2002). Optimal learning and experimentation in bandit problems. *J. Econom. Dynam. Control* **27** 87–108. MR1925627 https://doi.org/10.1016/S0165-1889(01)00028-8

BURNETAS, A. N. and KATEHAKIS, M. N. (1996). Optimal adaptive policies for sequential allocation problems. *Adv. in Appl. Math.* **17** 122–142. MR1390571 https://doi.org/10.1006/aama.1996.0007

BURTINI, G., LOEPPKY, J. and LAWRENCE, R. (2015). A survey of online experiment design with the stochastic multi-armed bandit. Available at arXiv:1510.00757.

CAPPÉ, O., GARIVIER, A., MAILLARD, O.-A., MUNOS, R. and STOLTZ, G. (2013). Kullback–Leibler upper confidence bounds for optimal sequential allocation. *Ann. Statist.* **41** 1516–1541. MR3113820 https://doi.org/10.1214/13-AOS1119

CHANG, F. and LAI, T. L. (1987). Optimal stopping and dynamic allocation. *Adv. in Appl. Probab.* **19** 829–853. MR0914595 https://doi.org/10.2307/1427104

FABIUS, J. and VAN ZWET, J. R. (1970). Some remarks on the two-armed bandit. *Ann. Math. Stat.* **41** 1906–1916.

GITTINS, J. C. (1979). Bandit processes and dynamic allocation indices. *J. Roy. Statist. Soc. Ser. B* **41** 148–177. MR0547241

GITTINS, J. C. and JONES, D. M. (1979). A dynamic allocation index for the discounted multi-armed bandit problem. *Biometrika* **66** 561–565.

MSC2010 subject classifications. 62L05.
Key words and phrases. Efficiency, KL-UCB, subsampling, Thompson sampling, UCB.

GRAVES, T. L. and LAI, T. L. (1997). Asymptotically efficient adaptive choice of control laws in controlled Markov chains. *SIAM J. Control Optim.* **35** 715–743. MR1444336 https://doi.org/10.1137/S0363012994275440

KAUFMANN, E. (2014). Analyse de stratégies bayésiennes et fréquentistes pour l'allocation séquentielle de ressources. Ph.D. thesis.

KAUFMANN, E., CAPPÉ, O. and GARIVIER, A. (2012). On Bayesian upper confidence bounds for bandit problems. *Proc. Mach. Learn. Res.* **22** 592–600.

KORDA, N., KAUFMANN, E. and MUNOS, R. (2013). Thompson sampling for 1-dimensional exponential family bandits. In *NIPS* 26 1448–1456.

KOTZ, S., KOZUBOWSKI, T. J. and PODGÓRSKI, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, Boston, MA. MR1935481 https://doi.org/10.1007/978-1-4612-0173-1

KULESHOV, V. and PRECUP, D. (2014). Algorithms for the multi-armed bandit problem. Available at arXiv:1402.6028.

LAI, T. L. (1987). Adaptive treatment allocation and the multi-armed bandit problem. *Ann. Statist.* **15** 1091–1114. MR0902248 https://doi.org/10.1214/aos/1176350495

LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22. MR0776826 https://doi.org/10.1016/0196-8858(85)90002-8

SHIVASWAMY, P. and JOACHIMS, T. (2012). Multi-armed bandit problems with history. *Proc. Mach. Learn. Res.* **22** 1046–1054.

SUTTON, B. and BARTO, A. (1998). *Reinforcement Learning, an Introduction*. MIT Press, Cambridge, MA.

TEKIN, C. and LIU, M. (2010). Online algorithms for the multi-armed bandit problem with Markovian rewards. In 48*th Annual Allerton Conference on Communication, Control and Computing* 1675–1682.

THATHACHER, V. and SASTRY, P. S. (1985). A class of rapidly converging algorithms for learning automata. *IEEE Trans. Syst. Man Cybern.* **16** 168–175.

YAKOWITZ, S. and LOWE, W. (1991). Nonparametric bandit methods. *Ann. Oper. Res.* **28** 297–312. MR1105179 https://doi.org/10.1007/BF02055587

The Annals of Statistics
2020, Vol. 48, No. 1, 374–396
https://doi.org/10.1214/19-AOS1810

CONCENTRATION AND CONSISTENCY RESULTS FOR CANONICAL AND CURVED EXPONENTIAL-FAMILY MODELS OF RANDOM GRAPHS

BY MICHAEL SCHWEINBERGER* AND JONATHAN STEWART**

*Department of Statistics, Rice University, *m.s@rice.edu; **jonathan.stewart@rice.edu*

Statistical inference for exponential-family models of random graphs with dependent edges is challenging. We stress the importance of additional structure and show that additional structure facilitates statistical inference. A simple example of a random graph with additional structure is a random graph with neighborhoods and local dependence within neighborhoods. We develop the first concentration and consistency results for maximum likelihood and M-estimators of a wide range of canonical and curved exponential-family models of random graphs with local dependence. All results are nonasymptotic and applicable to random graphs with finite populations of nodes, although asymptotic consistency results can be obtained as well. In addition, we show that additional structure can facilitate subgraph-to-graph estimation, and present concentration results for subgraph-to-graph estimators. As an application, we consider popular curved exponential-family models of random graphs, with local dependence induced by transitivity and parameter vectors whose dimensions depend on the number of nodes.

REFERENCES

[1] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR0489333

[2] BHAMIDI, S., BRESLER, G. and SLY, A. (2011). Mixing time of exponential random graphs. *Ann. Appl. Probab.* **21** 2146–2170. MR2895412 https://doi.org/10.1214/10-AAP740

[3] BHATTACHARYA, B. B. and MUKHERJEE, S. (2018). Inference in Ising models. *Bernoulli* **24** 493–525. MR3706767 https://doi.org/10.3150/16-BEJ886

[4] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[5] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. MR0882001

[6] BUTTS, C. T. (2011). Bernoulli graph bounds for general random graph models. *Sociol. Method.* **41** 299–345.

[7] BUTTS, C. T. and ALMQUIST, Z. W. (2015). A flexible parameterization for baseline mean degree in multiple-network ERGMs. *J. Math. Sociol.* **39** 163–167. MR3367715 https://doi.org/10.1080/0022250X.2014.967851

[8] CHATTERJEE, S. (2005). Concentration inequalities with exchangeable pairs. Ph.D. thesis, Dept. Statistics, Stanford Univ., Satnford, CA.

[9] CHATTERJEE, S. (2007). Estimation in spin glasses: A first step. *Ann. Statist.* **35** 1931–1946. MR2363958 https://doi.org/10.1214/009053607000000109

[10] CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155

[11] CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728

[12] CRANE, H. and DEMPSEY, W. (2015). A framework for statistical network modeling. Available at https://arxiv.org/abs/1509.08185.

MSC2010 subject classifications. Primary 05C80; secondary 62B05, 62F10, 91D30.

Key words and phrases. Curved exponential families, exponential families, exponential-family random graph models, M-estimators, multilevel networks, social networks.

[13] EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3** 1189–1242. MR0428531

[14] EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376. MR0471152

[15] FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518

[16] GEYER, C. J. (2009). Likelihood inference in exponential families and directions of recession. *Electron. J. Stat.* **3** 259–289. MR2495839 https://doi.org/10.1214/08-EJS349

[17] GODAMBE, V. P. and KALE, B. K. (1991). *Estimating Functions*. Oxford Univ. Press, Oxford.

[18] HANDCOCK, M. S. (2003). Statistical models for social networks: Inference and degeneracy. In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (R. Breiger, K. Carley and P. Pattison, eds.) 1–12. National Academies Press, Washington, DC.

[19] HANDCOCK, M. S. and GILE, K. J. (2010). Modeling social networks from sampled data. *Ann. Appl. Stat.* **4** 5–25. MR2758082 https://doi.org/10.1214/08-AOAS221

[20] HARRIS, J. K. (2013). *An Introduction to Exponential Random Graph Modeling*. Sage, Thousand Oaks, CA.

[21] HOLLAND, P. W. and LEINHARDT, S. (1976). Local structure in social networks. *Sociol. Method.* 1–45.

[22] HOLLWAY, J., LOMI, A., PALLOTTI, F. and STADTFELD, C. (2017). Multilevel social spaces: The network dynamics of organizational fields. *Network Science* **5** 187–212.

[23] HUNTER, D. R. (2007). Curved exponential family models for social networks. *Soc. Netw.* **29** 216–230.

[24] HUNTER, D. R., GOODREAU, S. M. and HANDCOCK, M. S. (2008). Goodness of fit of social network models. *J. Amer. Statist. Assoc.* **103** 248–258. MR2394635 https://doi.org/10.1198/016214507000000446

[25] HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264 https://doi.org/10.1198/106186006X133069

[26] HUNTER, D. R., KRIVITSKY, P. N. and SCHWEINBERGER, M. (2012). Computational statistical methods for social network models. *J. Comput. Graph. Statist.* **21** 856–882. MR3005801 https://doi.org/10.1080/10618600.2012.732921

[27] JANSON, S. and RUCIŃSKI, A. (2002). The infamous upper tail. *Random Structures Algorithms* **20** 317–342. MR1900611 https://doi.org/10.1002/rsa.10031

[28] JONASSON, J. (1999). The random triangle model. *J. Appl. Probab.* **36** 852–867. MR1737058 https://doi.org/10.1239/jap/1032374639

[29] KASS, R. E. and VOS, P. W. (1997). *Geometrical Foundations of Asymptotic Inference. Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. MR1461540 https://doi.org/10.1002/9781118165980

[30] KIM, J. H. and VU, V. H. (2004). Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms* **24** 166–174. MR2035874 https://doi.org/10.1002/rsa.10113

[31] KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models. Springer Series in Statistics*. Springer, New York. MR2724362 https://doi.org/10.1007/978-0-387-88146-1

[32] KONTOROVICH, L. and RAMANAN, K. (2008). Concentration inequalities for dependent random variables via the martingale method. *Ann. Probab.* **36** 2126–2158. MR2478678 https://doi.org/10.1214/07-AOP384

[33] KRIVITSKY, P. N. (2012). Exponential-family random graph models for valued networks. *Electron. J. Stat.* **6** 1100–1128. MR2988440 https://doi.org/10.1214/12-EJS696

[34] KRIVITSKY, P. N., HANDCOCK, M. S. and MORRIS, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Stat. Methodol.* **8** 319–339. MR2800354 https://doi.org/10.1016/j.stamet.2011.01.005

[35] KRIVITSKY, P. N. and KOLACZYK, E. D. (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statist. Sci.* **30** 184–198. MR3353102 https://doi.org/10.1214/14-STS502

[36] LAURITZEN, S., RINALDO, A. and SADEGHI, K. (2018). Random networks, graphical models and exchangeability. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 481–508. MR3798875 https://doi.org/10.1111/rssb.12266

[37] LAZEGA, E. and SNIJDERS, T. A. B., eds. (2016). *Multilevel Network Analysis for the Social Sciences*. Springer, Cham.

[38] LOMI, A., ROBINS, G. and TRANMER, M. (2016). Introduction to multilevel social networks. *Soc. Netw.* **44** 266–268.

[39] LUSHER, D., KOSKINEN, J. and ROBINS, G. (2013). *Exponential Random Graph Models for Social Networks*. Cambridge Univ. Press, Cambridge.

[40] MUKHERJEE, S. (2013). Consistent estimation in the two star exponential random graph model. Technical report. Dept. Statistics, Columbia Univ. New York. Available at arXiv:1310.4526v1.

[41] NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255 https://doi.org/10.1198/016214501753208735
[42] RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343 https://doi.org/10.1214/09-AOS691
[43] RINALDO, A., FIENBERG, S. E. and ZHOU, Y. (2009). On the geometry of discrete exponential families with application to exponential random graph models. *Electron. J. Stat.* **3** 446–484. MR2507456 https://doi.org/10.1214/08-EJS350
[44] RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804 https://doi.org/10.1214/12-AOS1078
[45] RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581
[46] SAMSON, P.-M. (2000). Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.* **28** 416–461. MR1756011 https://doi.org/10.1214/aop/1019160125
[47] SCHWEINBERGER, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106** 1361–1370. MR2896841 https://doi.org/10.1198/jasa.2011.tm10747
[48] SCHWEINBERGER, M. and HANDCOCK, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 647–676. MR3351449 https://doi.org/10.1111/rssb.12081
[49] SCHWEINBERGER, M. and LUNA, P. (2018). HERGM: Hierarchical exponential-family random graph models. *J. Stat. Softw.* **85** 1–39.
[50] SCHWEINBERGER, M. and STEWART, J. (2019). Supplement to "Concentration and consistency results for canonical and curved exponential-family models of random graphs." https://doi.org/10.1214/19-AOS1810SUPP.
[51] SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112 https://doi.org/10.1214/12-AOS1044
[52] SLAUGHTER, A. J. and KOEHLY, L. M. (2016). Multilevel models for social networks: Hierarchical Bayesian approaches to exponential random graph modeling. *Soc. Netw.* **44** 334–345.
[53] SNIJDERS, T. A. B., PATTISON, P. E., ROBINS, G. L. and HANDCOCK, M. S. (2006). New specifications for exponential random graph models. *Sociol. Method.* **36** 99–153.
[54] STEWART, J., SCHWEINBERGER, M., BOJANOWSKI, M. and MORRIS, M. (2019). Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Soc. Netw.* To appear.
[55] VU, V. H. (2002). Concentration of non-Lipschitz functions and applications. *Random Structures Algorithms* **20** 262–316. MR1900610 https://doi.org/10.1002/rsa.10032
[56] WANG, P., ROBINS, G., PATTISON, P. and LAZEGA, E. (2013). Exponential random graph models for multilevel networks. *Soc. Netw.* **35** 96–115.
[57] WASSERMAN, S. and PATTISON, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p. *Psychometrika* **61** 401–425. MR1424909 https://doi.org/10.1007/BF02294547
[58] XIANG, R. and NEVILLE, J. (2011). Relational learning with one network: An asymptotic analysis. In *Proceedings of the* 14*th International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 1–10.
[59] YAN, T., JIANG, B., FIENBERG, S. E. and LENG, C. (2019). Statistical inference in a directed network model with covariates. *J. Amer. Statist. Assoc.* **114** 857–868. MR3963186
[60] YAN, T., LENG, C. and ZHU, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *Ann. Statist.* **44** 31–57. MR3449761 https://doi.org/10.1214/15-AOS1343
[61] YANG, E., RAVIKUMAR, P., ALLEN, G. I. and LIU, Z. (2015). Graphical models via univariate exponential family distributions. *J. Mach. Learn. Res.* **16** 3813–3847. MR3450553
[62] ZAPPA, P. and LOMI, A. (2015). The analysis of multilevel networks in organizations: Models and empirical tests. *Organizational Research Methods* **18** 542–569.

The Annals of Statistics
2020, Vol. 48, No. 1, 397–412
https://doi.org/10.1214/19-AOS1812

THE NUMERICAL BOOTSTRAP

BY HAN HONG[1] AND JESSIE LI[2]

[1]*Department of Economics, Stanford University, hanhong@stanford.edu*
[2]*Department of Economics, University of California, Santa Cruz, jeqli@ucsc.edu*

This paper proposes a numerical bootstrap method that is consistent in many cases where the standard bootstrap is known to fail and where the m-out-of-n bootstrap and subsampling have been the most commonly used inference approaches. We provide asymptotic analysis under both fixed and drifting parameter sequences, and we compare the approximation error of the numerical bootstrap with that of the m-out-of-n bootstrap and subsampling. Finally, we discuss applications of the numerical bootstrap, such as constrained and unconstrained M-estimators converging at both regular and nonstandard rates, Laplace-type estimators, and test statistics for partially identified models.

REFERENCES

[1] ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245 https://doi.org/10.1111/j.1468-0262.2005.00613.x

[2] ANDREWS, D. W. K. and SOARES, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78** 119–157. MR2642858 https://doi.org/10.3982/ECTA7502

[3] ANDREWS, D. W. K. and SHI, X. (2013). Inference based on conditional moment inequalities. *Econometrica* **81** 609–666. MR3043344 https://doi.org/10.3982/ECTA9370

[4] BABU, G. J. and SINGH, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.* **17** 261–278. MR0813236 https://doi.org/10.1016/0047-259X(85)90084-3

[5] BERTAIL, P. (1997). Second-order properties of an extrapolated bootstrap without replacement under weak assumptions. *Bernoulli* **3** 149–179. MR1466305 https://doi.org/10.2307/3318585

[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] BONNANS, J. F. and SHAPIRO, A. (2000). *Perturbation Analysis of Optimization Problems. Springer Series in Operations Research*. Springer, New York. MR1756264 https://doi.org/10.1007/978-1-4612-1394-9

[8] BUGNI, F. A., CANAY, I. A. and SHI, X. (2015). Specification tests for partially identified models defined by moment inequalities. *J. Econometrics* **185** 259–282. MR3300346 https://doi.org/10.1016/j.jeconom.2014.10.013

[9] BUGNI, F. A., CANAY, I. A. and SHI, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quant. Econ.* **8** 1–38. MR3638598 https://doi.org/10.3982/QE490

[10] CATTANEO, M. D., JANSSON, M. and NAGASAWA, K. (2017). Bootstrap-Based Inference for Cube Root Consistent Estimators. Preprint. arXiv:1704.08066.

[11] CAVANAGH, C. L. (1987). Limiting behavior of estimators defined by optimization. Unpublished manuscript.

[12] CHAKRABORTY, B., LABER, E. B. and ZHAO, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics* **69** 714–723. MR3106599 https://doi.org/10.1111/biom.12052

[13] CHERNOZHUKOV, V. and HONG, H. (2003). An MCMC approach to classical estimation. *J. Econometrics* **115** 293–346. MR1984779 https://doi.org/10.1016/S0304-4076(03)00100-3

MSC2010 subject classifications. 62F40.
Key words and phrases. Bootstrap, numerical differentiation, directional differentiability.

[14] DAS, D. and LAHIRI, S. N. (2019). Second order correctness of perturbation bootstrap M-estimator of multiple linear regression parameter. *Bernoulli* **25** 654–682. MR3892332 https://doi.org/10.3150/17-bej1001

[15] DÜMBGEN, L. (1993). On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields* **95** 125–140. MR1207311 https://doi.org/10.1007/BF01197342

[16] FANG, Z.and SANTOS, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies* **86** 377–412. https://doi.org/10.1093/restud/rdy049

[17] GEYER, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179 https://doi.org/10.1214/aos/1176325768

[18] HONG, H. and LI, J. (2019). Supplement to "The numerical bootstrap." https://doi.org/10.1214/19-AOS1812SUPP.

[19] JUN, S. J., PINKSE, J. and WAN, Y. (2015). Classical Laplace estimation for $\sqrt[3]{n}$-consistent estimators: Improved convergence rates and rate-adaptive inference. *J. Econometrics* **187** 201–216. MR3347303 https://doi.org/10.1016/j.jeconom.2015.01.005

[20] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[21] KNIGHT, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. Unpublished manuscript 37.

[22] LABER, E. B. and MURPHY, S. A. (2011). Adaptive confidence intervals for the test error in classification. *J. Amer. Statist. Assoc.* **106** 904–913. MR2894746 https://doi.org/10.1198/jasa.2010.tm10053

[23] MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. MR0436905 https://doi.org/10.1016/0304-4076(75)90032-9

[24] MINNIER, J., TIAN, L. and CAI, T. (2011). A perturbation method for inference on regularized regression estimates. *J. Amer. Statist. Assoc.* **106** 1371–1382. MR2896842 https://doi.org/10.1198/jasa.2011.tm10382

[25] POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

[26] ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*. *Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402 https://doi.org/10.1007/978-1-4419-9076-1_11

[27] ROMANO, J. P. and SHAIKH, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *Ann. Statist.* **40** 2798–2822. MR3097960 https://doi.org/10.1214/12-AOS1051

[28] RÖMISCH, W. (2005). Delta method, infinite fimensional. In *Encyclopedia of Statistical Sciences* (S. Kotz et al., eds). Wiley, New York.

[29] SEIJO, E. and SEN, B. (2015). Bootstrapping Manski's Maximum Score Estimator. Preprint. arXiv:1105.1976.

[30] SHAO, J. (1994). Bootstrap sample size in nonregular cases. *Proc. Amer. Math. Soc.* **122** 1251–1262. MR1227529 https://doi.org/10.2307/2161196

[31] SHAPIRO, A. (1989). Asymptotic properties of statistical estimators in stochastic programming. *Ann. Statist.* **17** 841–858. MR0994271 https://doi.org/10.1214/aos/1176347146

[32] TEWES, J., POLITIS, D. N. and NORDMAN, D. J. (2019). Convolved subsampling estimation with applications to block bootstrap. *Ann. Statist.* **47** 468–496. MR3909939 https://doi.org/10.1214/18-AOS1695

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

[34] 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

The Annals of Statistics
2020, Vol. 48, No. 1, 413–439
https://doi.org/10.1214/19-AOS1814

CONSISTENT SELECTION OF THE NUMBER OF CHANGE-POINTS VIA SAMPLE-SPLITTING

BY CHANGLIANG ZOU[1,*], GUANGHUI WANG[1,**] AND RUNZE LI[2]

[1]*School of Statistics and Data Science, Nankai University, [*]nk.chlzou@gmail.com; [**]ghwang.nk@gmail.com*
[2]*Department of Statistics and The Methodology Center, Pennsylvania State University, rzli@psu.edu*

In multiple change-point analysis, one of the major challenges is to estimate the number of change-points. Most existing approaches attempt to minimize a Schwarz information criterion which balances a term quantifying model fit with a penalization term accounting for model complexity that increases with the number of change-points and limits overfitting. However, different penalization terms are required to adapt to different contexts of multiple change-point problems and the optimal penalization magnitude usually varies from the model and error distribution. We propose a data-driven selection criterion that is applicable to most kinds of popular change-point detection methods, including binary segmentation and optimal partitioning algorithms. The key idea is to select the number of change-points that minimizes the squared prediction error, which measures the fit of a specified model for a new sample. We develop a cross-validation estimation scheme based on an order-preserved sample-splitting strategy, and establish its asymptotic selection consistency under some mild conditions. Effectiveness of the proposed selection criterion is demonstrated on a variety of numerical experiments and real-data examples.

REFERENCES

ARLOT, S. and CELISSE, A. (2011). Segmentation of the mean of heteroscedastic data via cross-validation. *Stat. Comput.* **21** 613–632. MR2826696 https://doi.org/10.1007/s11222-010-9196-x

AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

AUGER, I. E. and LAWRENCE, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.* **51** 39–54. MR0978902 https://doi.org/10.1016/S0092-8240(89)80047-3

BAI, J. (1998). Estimation of multiple-regime regressions with least absolutes deviation. *J. Statist. Plann. Inference* **74** 103–134. MR1665123 https://doi.org/10.1016/S0378-3758(98)00082-2

BAI, Z., FUJIKOSHI, Y. and CHOI, K. P. (2017). Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis. *Ann. Statist.* **46** 1050–1076. MR3797996

BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121 https://doi.org/10.2307/2998540

BAI, J. and PERRON, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometrics* **18** 1–22.

BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031

BRAUN, J. V., BRAUN, R. K. and MÜLLER, H.-G. (2000). Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika* **87** 301–314. MR1782480 https://doi.org/10.1093/biomet/87.2.301

CAO, H. and WU, W. B. (2015). Changepoint estimation: Another look at multiple testing problems. *Biometrika* **102** 974–980. MR3431567 https://doi.org/10.1093/biomet/asv031

CHEN, J. and GUPTA, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *J. Amer. Statist. Assoc.* **92** 739–747. MR1467863 https://doi.org/10.2307/2965722

MSC2010 subject classifications. 62H12.
Key words and phrases. Cross-validation, dynamic programming, least-squares, model selection, multiple change-point model, selection consistency.

CHEN, J. and GUPTA, A. K. (2012). *Parametric Statistical Change Point Analysis*: *With Applications to Genetics*, *Medicine*, *and Finance*, 2nd ed. Birkhäuser/Springer, New York. MR3025631 https://doi.org/10.1007/978-0-8176-4801-5

DARLING, D. A. and ERDÖS, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math*. *J*. **23** 143–155. MR0074712

FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann*. *Statist*. **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

HANNART, A. and NAVEAU, P. (2012). An improved Bayesian information criterion for multiple change-point models. *Technometrics* **54** 256–268. MR2967976 https://doi.org/10.1080/00401706.2012.694780

HAO, N., NIU, Y. S. and ZHANG, H. (2013). Multiple change-point detection via a screening and ranking algorithm. *Statist*. *Sinica* **23** 1553–1572. MR3222810

HARCHAOUI, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J*. *Amer*. *Statist*. *Assoc*. **105** 1480–1493. MR2796565 https://doi.org/10.1198/jasa.2010.tm09181

HAWKINS, D. M. (2001). Fitting multiple change-point models to data. *Comput*. *Statist*. *Data Anal*. **37** 323–341. MR1856677 https://doi.org/10.1016/S0167-9473(00)00068-2

HAYNES, K., ECKLEY, I. A. and FEARNHEAD, P. (2017). Computationally efficient changepoint detection for a range of penalties. *J*. *Comput*. *Graph*. *Statist*. **26** 134–143. MR3610414 https://doi.org/10.1080/10618600.2015.1116445

HAYNES, K., FEARNHEAD, P. and ECKLEY, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Stat*. *Comput*. **27** 1293–1305. MR3647098 https://doi.org/10.1007/s11222-016-9687-5

HORVÁTH, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann*. *Statist*. **21** 671–680. MR1232511 https://doi.org/10.1214/aos/1176349143

JENG, X. J., CAI, T. T. and LI, H. (2010). Optimal sparse segment identification with application in copy number variation analysis. *J*. *Amer*. *Statist*. *Assoc*. **105** 1156–1166. MR2752611 https://doi.org/10.1198/jasa.2010.tm10083

KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J*. *Amer*. *Statist*. *Assoc*. **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann*. *Statist*. **17** 1217–1241. MR1015147 https://doi.org/10.1214/aos/1176347265

LAVIELLE, M. (2005). Using penalized contrasts for the change-point problem. *Signal Process*. **85** 1501–1510.

LEE, C.-B. (1996). Nonparametric multiple change-point estimators. *Statist*. *Probab*. *Lett*. **27** 295–304. MR1395582 https://doi.org/10.1016/0167-7152(95)00089-5

LOADER, C. R. (1996). Change point estimation using nonparametric regression. *Ann*. *Statist*. **24** 1667–1678. MR1416655 https://doi.org/10.1214/aos/1032298290

MATTESON, D. S. and JAMES, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J*. *Amer*. *Statist*. *Assoc*. **109** 334–345. MR3180567 https://doi.org/10.1080/01621459.2013.849605

MÜLLER, H.-G. and STADTMÜLLER, U. (1999). Discontinuous versus smooth regression. *Ann*. *Statist*. **27** 299–337. MR1701113 https://doi.org/10.1214/aos/1018031100

NISHII, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann*. *Statist*. **12** 758–765. MR0740928 https://doi.org/10.1214/aos/1176346522

NIU, Y. S., HAO, N. and ZHANG, H. (2016). Multiple change-point detection: A selective overview. *Statist*. *Sci*. **31** 611–623. MR3598742 https://doi.org/10.1214/16-STS587

NIU, Y. S. and ZHANG, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann*. *Appl*. *Stat*. **6** 1306–1326. MR3012531 https://doi.org/10.1214/12-AOAS539

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

SHAO, J. (1993). Linear model selection by cross-validation. *J*. *Amer*. *Statist*. *Assoc*. **88** 486–494. MR1224373

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

VENKATRAMAN, E. S. (1992). Consistency results in multiple change-point problems. Ph.D. thesis, Stanford Univ., ProQuest LLC, Ann Arbor, MI. MR2687536

WANG, G., ZOU, C. and YIN, G. (2018). Change-point detection in multinomial data with a large number of categories. *Ann*. *Statist*. **46** 2020–2044. MR3845009 https://doi.org/10.1214/17-AOS1610

WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **69** 391–410. MR2323759 https://doi.org/10.1111/j.1467-9868.2007.00594.x

YANG, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika* **92** 937–950. MR2234196 https://doi.org/10.1093/biomet/92.4.937

YANG, Y. (2007). Consistency of cross validation for comparing regression procedures. *Ann. Statist.* **35** 2450–2473. MR2382654 https://doi.org/10.1214/009053607000000514

YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statist. Probab. Lett.* **6** 181–189. MR0919373 https://doi.org/10.1016/0167-7152(88)90118-6

YAO, Y.-C. and AU, S. T. (1989). Least-squares estimation of a step function. *Sankhyā, Ser. A* **51** 370–381. MR1175613

ZHANG, P. (1993). Model selection via multifold cross validation. *Ann. Statist.* **21** 299–313. MR1212178 https://doi.org/10.1214/aos/1176349027

ZHANG, N. R. and SIEGMUND, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63** 22–32, 309. MR2345571 https://doi.org/10.1111/j.1541-0420.2006.00662.x

ZOU, C., WANG, G. and LI, R. (2020). Supplement to "Consistent selection of the number of change-points via sample-splitting." https://doi.org/10.1214/19-AOS1814SUPP.

ZOU, C., YIN, G., FENG, L. and WANG, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *Ann. Statist.* **42** 970–1002. MR3210993 https://doi.org/10.1214/14-AOS1210

The Annals of Statistics
2020, Vol. 48, No. 1, 440–463
https://doi.org/10.1214/19-AOS1815

UNIFORMLY VALID CONFIDENCE INTERVALS POST-MODEL-SELECTION

BY FRANÇOIS BACHOC[1], DAVID PREINERSTORFER[2] AND LUKAS STEINBERGER[3]

[1]*Institut de Mathématiques de Toulouse, Université Paul Sabatier, francois.bachoc@math.univ-toulouse.fr*
[2]*ECARES, Université libre de Bruxelles, david.preinerstorfer@ulb.ac.be*
[3]*Department of Mathematical Stochastics, University of Freiburg, lukas.steinberger@stochastik.uni-freiburg.de*

We suggest general methods to construct asymptotically uniformly valid confidence intervals post-model-selection. The constructions are based on principles recently proposed by Berk et al. (*Ann. Statist.* **41** (2013) 802–837). In particular, the candidate models used can be misspecified, the target of inference is model-specific, and coverage is guaranteed for *any* data-driven model selection procedure. After developing a general theory, we apply our methods to practically important situations where the candidate set of models, from which a working model is selected, consists of fixed design homoskedastic or heteroskedastic linear models, or of binary regression models with general link functions. In an extensive simulation study, we find that the proposed confidence intervals perform remarkably well, even when compared to existing methods that are tailored only for specific model selection procedures.

REFERENCES

ARNOLD, S. F. (1980). Asymptotic validity of F tests for the ordinary linear model and the multiple correlation model. *J. Amer. Statist. Assoc.* **75** 890–894. MR0600972

BACHOC, F., LEEB, H. and PÖTSCHER, B. M. (2019). Valid confidence intervals for post-model-selection predictors. *Ann. Statist.* **47** 1475–1504. MR3911119 https://doi.org/10.1214/18-AOS1721

BACHOC, F., PREINERSTORFER, D. and STEINBERGER, L. (2019). Supplement to "Uniformly valid confidence intervals post-model-selection." https://doi.org/10.1214/19-AOS1815SUPP.

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2011). Inference for high-dimensional sparse econometric models. In *Advances in Economics and Econometrics*. 10*th World Congress of the Econometric Society*, *Vol. III* 245–295.

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122 https://doi.org/10.1214/12-AOS1077

DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

EICKER, F. (1967). Limit theorems for regressions with unequal and dependent errors. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley*, *Calif.*, 1965/66) 59–82. Univ. California Press, Berkeley, CA. MR0214223

FAHRMEIR, L. (1990). Maximum likelihood estimation in misspecified generalized linear models. *Statistics* **21** 487–502. MR1087280 https://doi.org/10.1080/02331889008802259

FITHIAN, W., SUN, D. and TAYLOR, J. (2015). Optimal inference after model selection. Preprint. Available at arXiv:1410.2597.

GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge, MA. MR0062975

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

MSC2010 subject classifications. Primary 62F12, 62F25; secondary 62F35, 62J02.
Key words and phrases. Inference post-model-selection, uniform asymptotic inference, regression.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373

KABAILA, P. and LEEB, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *J. Amer. Statist. Assoc.* **101** 619–629. MR2256178 https://doi.org/10.1198/016214505000001140

KIVARANOVIC, D. and LEEB, H. (2018). Expected length of post-model-selection confidence intervals conditional on polyhedral constraints. Preprint. Available at arXiv:1803.01665.

KUBKOWSKI, M. and MIELNICZUK, J. (2017). Active sets of predictors for misspecified logistic regression. *Statistics* **51** 1023–1045. MR3698499 https://doi.org/10.1080/02331888.2017.1290096

LEE, J. D. and TAYLOR, J. E. (2014). Exact post model selection inference for marginal screening. In *Advances in Neural Information Processing Systems* 27 (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.) 136–144. Curran Associates, Red Hook, NY.

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

LEEB, H. and PÖTSCHER, B. M. (2003). The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory* **19** 100–142. MR1965844 https://doi.org/10.1017/S0266466603191050

LEEB, H. and PÖTSCHER, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory* **21** 21–59. MR2153856 https://doi.org/10.1017/S0266466605050036

LEEB, H. and PÖTSCHER, B. M. (2006). Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results. *Econometric Theory* **22** 69–97. MR2212693 https://doi.org/10.1017/S0266466606060038

LEEB, H. and PÖTSCHER, B. M. (2008). Model selection. In *Handbook of Financial Time Series* (T. G. Andersen, R. A. Davis, J.-P. Kreiß and T. Mikosch, eds.) 785–821. Springer, New York, NY.

LEEB, H., PÖTSCHER, B. M. and EWALD, K. (2015). On various confidence intervals post-model-selection. *Statist. Sci.* **30** 216–227. MR3353104 https://doi.org/10.1214/14-STS507

LV, J. and LIU, J. S. (2010). Model selection principles in misspecified models. Preprint. Available at arXiv:1005.5483v1.

LV, J. and LIU, J. S. (2014). Model selection principles in misspecified models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 141–167. MR3153937 https://doi.org/10.1111/rssb.12023

POLLAK, M. (1972). A note on infinitely divisible random vectors. *Ann. Math. Stat.* **43** 673–675.

PÖTSCHER, B. M. (2009). Confidence sets based on sparse estimators are necessarily large. *Sankhyā* **71** 1–18. MR2579644

RAIKOV, D. (1938). On a connection between the central limit-law of the theory of probability and the law of great numbers. *Izv. Ross. Akad. Nauk Ser. Mat.* **2** 323–338.

RINALDO, A., WASSERMAN, L., G'SELL, M., LEI, J. and TIBSHIRANI, R. (2016). Bootstrapping and sample splitting for high-dimensional, assumption-free inference. Preprint. Available at arXiv:1611.05401.

RUUD, P. A. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica* **51** 225–228. MR0694460 https://doi.org/10.2307/1912257

TAYLOR, J. and TIBSHIRANI, R. (2018). Post-selection inference for ℓ_1-penalized likelihood models. *Canad. J. Statist.* **46** 41–61. MR3767165 https://doi.org/10.1002/cjs.11313

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

TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2018). Uniform asymptotic inference and the bootstrap after model selection. *Ann. Statist.* **46** 1255–1287. MR3798003 https://doi.org/10.1214/17-AOS1584

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

WEDDERBURN, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* **63** 27–32. MR0408092 https://doi.org/10.1093/biomet/63.1.27

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

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163 https://doi.org/10.2307/1912526

ZHANG, K. (2017). Spherical cap packing asymptotics and rank-extreme detection. *IEEE Trans. Inform. Theory* **63** 4572–4584. MR3666977 https://doi.org/10.1109/TIT.2017.2700202

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

The Annals of Statistics
2020, Vol. 48, No. 1, 464–490
https://doi.org/10.1214/19-AOS1816

EFFICIENT ESTIMATION OF LINEAR FUNCTIONALS OF PRINCIPAL COMPONENTS

BY VLADIMIR KOLTCHINSKII[1], MATTHIAS LÖFFLER[2,*] AND RICHARD NICKL[2,**]

[1]*School of Mathematics, Georgia Institute of Technology, vladimir.koltchinskii@math.gatech.edu*
[2]*Statistical Laboratory, University of Cambridge, [*]m.loffler@statslab.cam.ac.uk; [**]r.nickl@statslab.cam.ac.uk*

We study principal component analysis (PCA) for mean zero i.i.d. Gaussian observations $X_1, \dots, X_n$ in a separable Hilbert space $\mathbb{H}$ with unknown covariance operator Σ. The complexity of the problem is characterized by its effective rank $\mathbf{r}(\Sigma) := \frac{\operatorname{tr}(\Sigma)}{\|\Sigma\|}$, where $\operatorname{tr}(\Sigma)$ denotes the trace of Σ and $\|\Sigma\|$ denotes its operator norm. We develop a method of bias reduction in the problem of estimation of linear functionals of eigenvectors of Σ. Under the assumption that $\mathbf{r}(\Sigma) = o(n)$, we establish the asymptotic normality and asymptotic properties of the risk of the resulting estimators and prove matching minimax lower bounds, showing their semiparametric optimality.

REFERENCES

[1] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 https://doi.org/10.1214/aoms/1177704248

[2] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

[3] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

[4] BLANCHARD, G., BOUSQUET, O. and ZWALD, L. (2007). Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** 259–294.

[5] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. MR3449395 https://doi.org/10.1007/s00440-015-0616-x

[6] CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

[7] 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

[8] DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934 https://doi.org/10.1016/0047-259X(82)90088-4

[9] EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. MR0716321

[10] FAN, J., RIGOLLET, P. and WANG, W. (2015). Estimation of functionals of sparse covariance matrices. *Ann. Statist.* **43** 2706–2737. MR3405609 https://doi.org/10.1214/15-AOS1357

[11] GAO, C. and ZHOU, H. H. (2015). Rate-optimal posterior contraction for sparse PCA. *Ann. Statist.* **43** 785–818. MR3325710 https://doi.org/10.1214/14-AOS1268

[12] GAO, C. and ZHOU, H. H. (2016). Bernstein–von Mises theorems for functionals of the covariance matrix. *Electron. J. Stat.* **10** 1751–1806. MR3522660 https://doi.org/10.1214/15-EJS1048

[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] 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

MSC2010 subject classifications. Primary 62H25; secondary 62E17.
Key words and phrases. Principal component analysis, spectral projections, asymptotic normality, semiparametric efficiency.

[15] JANKOVÁ, J. and VAN DE GEER, S. (2018). Semiparametric efficiency bounds for high-dimensional models. *Ann. Statist.* **46** 2336–2359. MR3845020 https://doi.org/10.1214/17-AOS1622

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

[17] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[18] KOLTCHINSKII, V. (2017). Asymptotically efficient estimation of smooth functionals of covariance operators. Available at https://arxiv.org/abs/1710.09072.

[19] KOLTCHINSKII, V., LÖFFLER, M. and NICKL, R. (2019). Supplement to "Efficient estimation of linear functionals of principal components." https://doi.org/10.1214/19-AOS1816SUPP.

[20] 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

[21] 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

[22] 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

[23] KOLTCHINSKII, V. and LOUNICI, K. (2017). New asymptotic results in principal component analysis. *Sankhya A* **79** 254–297. MR3707422 https://doi.org/10.1007/s13171-017-0106-6

[24] LILA, E., ASTON, J. A. D. and SANGALLI, L. M. (2016). Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *Ann. Appl. Stat.* **10** 1854–1879. MR3592040 https://doi.org/10.1214/16-AOAS975

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

[26] NAUMOV, A., SPOKOINY, V. and ULYANOV, V. (2018). Confidence sets for spectral projectors of covariance matrices. *Dokl. Math.* **98** 511–514.

[27] NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

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

[29] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

[30] REISS, M. and WAHL, M. (2016). Non-asymptotic upper bounds for the reconstruction error of PCA. Available at arXiv:1609.03779.

[31] REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286

[32] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

[33] 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

[34] 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

[35] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

[36] VU, V. Q. and LEI, J. (2013). Minimax sparse principal subspace estimation in high dimensions. *Ann. Statist.* **41** 2905–2947. MR3161452 https://doi.org/10.1214/13-AOS1151

[37] WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438 https://doi.org/10.1214/15-AOS1369

[38] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

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

The Annals of Statistics
2020, Vol. 48, No. 1, 491–513
https://doi.org/10.1214/19-AOS1819

OPTIMAL PREDICTION IN THE LINEARLY TRANSFORMED SPIKED MODEL

BY EDGAR DOBRIBAN[1], WILLIAM LEEB[2] AND AMIT SINGER[3]

[1]*Department of Statistics, The Wharton School, University of Pennsylvania, dobriban@wharton.upenn.edu*
[2]*School of Mathematics, University of Minnesota Twin Cities, wleeb@umn.edu*
[3]*Program in Applied and Computational Mathematics, Department of Mathematics, Princeton University, amits@math.princeton.edu*

We consider the *linearly transformed spiked model*, where the observations Y_i are noisy linear transforms of unobserved signals of interest X_i:

$$Y_i = A_i X_i + \varepsilon_i,$$

for $i = 1, \dots, n$. The transform matrices A_i are also observed. We model the unobserved signals (or regression coefficients) X_i as vectors lying on an unknown low-dimensional space. Given only Y_i and A_i how should we predict or recover their values?

The naive approach of performing regression for each observation separately is inaccurate due to the large noise level. Instead, we develop optimal methods for predicting X_i by "borrowing strength" across the different samples. Our linear empirical Bayes methods scale to large datasets and rely on weak moment assumptions.

We show that this model has wide-ranging applications in signal processing, deconvolution, cryo-electron microscopy, and missing data with noise. For missing data, we show in simulations that our methods are more robust to noise and to unequal sampling than well-known matrix completion methods.

REFERENCES

ANDÉN, J., KATSEVICH, E. and SINGER, A. (2015). Covariance estimation using conjugate gradient for 3D classification in cryo-EM. In *Biomedical Imaging* (*ISBI*), 2015 *IEEE* 12*th International Symposium on* 200–204. IEEE, New York.

ASPIRE (2017). Algorithms for single particle reconstruction. Available at http://spr.math.princeton.edu/.

BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl*. **1** 1150011, 21. MR2934717 https://doi.org/10.1142/S2010326311500110

BAI, X.-C., MCMULLAN, G. and SCHERES, S. H. (2015). How cryo-EM is revolutionizing structural biology. *Trends Biochem*. *Sci*. **40** 49–57.

BAI, Z. and SILVERSTEIN, J. W. (2009). *Spectral Analysis of Large Dimensional Random Matrices*. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal*. **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal*. **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal*. **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019

BHAMRE, T., ZHANG, T. and SINGER, A. (2016). Denoising and covariance estimation of single particle cryo-EM images. *J. Struct. Biol*. **195** 72–81.

MSC2010 subject classifications. Primary 62H25; secondary 62H15, 45B05.
Key words and phrases. Principal component analysis, random matrix theory, shrinkage, high dimensional, spiked model, missing data, matrix completion.

BLACKLEDGE, J. M. (2006). *Digital Signal Processing*: *Mathematical and Computational Methods*, *Software Development and Applications*. Elsevier, Amsterdam. MR2036815

BUJA, A. and EYUBOGLU, N. (1992). Remarks on parallel analysis. *Multivar. Behav. Res.* **27** 509–540.

CALLAWAY, E. (2015). The revolution will not be crystallized. *Nature* **525** 172.

CAMPISI, P. and EGIAZARIAN, K., eds. (2016). *Blind Image Deconvolution*: *Theory and Applications*. CRC Press, Boca Raton, FL. MR2404093 https://doi.org/10.1201/9781420007299

CANDÈS, E. J. 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. Inform. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061

CHEN, Y., BHOJANAPALLI, S., SANGHAVI, S. and WARD, R. (2015). Completing any low-rank matrix, provably. *J. Mach. Learn. Res.* **16** 2999–3034. MR3450532

DOBRIBAN, E. (2015). Efficient computation of limit spectra of sample covariance matrices. *Random Matrices Theory Appl.* **4** 1550019, 36. MR3418848 https://doi.org/10.1142/S2010326315500197

DOBRIBAN, E. (2017). Permutation methods for factor analysis and PCA. Preprint. Available at arXiv:1710.00479.

DOBRIBAN, E., LEEB, W. and SINGER, A. (2019). Supplement to "Optimal prediction in the linearly transformed spiked model." https://doi.org/10.1214/19-AOS1819SUPP.

DOBRIBAN, E. and OWEN, A. B. (2019). Deterministic parallel analysis: An improved method for selecting factors and principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 163–183. MR3904784

DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257

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

EFRON, B. (2012). *Large-Scale Inference*: *Empirical Bayes Methods for Estimation*, *Testing*, *and Prediction*. *Institute of Mathematical Statistics* (*IMS*) *Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

GAVISH, M. and DONOHO, D. L. (2017). Optimal shrinkage of singular values. *IEEE Trans. Inform. Theory* **63** 2137–2152. MR3626861 https://doi.org/10.1109/TIT.2017.2653801

GOLUB, G. H. and VAN LOAN, C. F. (2012). *Matrix Computations*, *Vol.* 3. *Johns Hopkins Studies in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR3024913

HACHEM, W., HARDY, A. and NAJIM, J. (2015). A survey on the eigenvalues local behavior of large complex correlated Wishart matrices. In *Modélisation Aléatoire et Statistique—Journées MAS* 2014. *ESAIM Proc. Surveys* **51** 150–174. EDP Sci., Les Ulis. MR3440796 https://doi.org/10.1051/proc/201551009

JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization. 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

JI, S. and YE, J. (2009). An accelerated gradient method for trace norm minimization. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 457–464. ACM, New York.

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

JOHNSTONE, I. M. and ONATSKI, A. (2015). Testing in high-dimensional spiked models. Preprint. Available at arXiv:1509.07269.

KAM, Z. (1980). The reconstruction of structure from electron micrographs of randomly oriented particles. *J. Theoret. Biol.* **82** 15–39.

KATSEVICH, E., KATSEVICH, A. and SINGER, A. (2015). Covariance matrix estimation for the cryo-EM heterogeneity problem. *SIAM J. Imaging Sci.* **8** 126–185. MR3302588 https://doi.org/10.1137/130935434

KESHAVAN, R. H. and MONTANARI, A. (2010). Regularization for matrix completion. In *Proceedings of International Symposium on Information Theory* 1503–1507. IEEE, New York.

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

KESHAVAN, R. H., OH, S. and MONTANARI, A. (2009). Matrix completion from a few entries. In 2009 *IEEE International Symposium on Information Theory* 324–328. IEEE, New York.

KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20** 282–303. MR3160583 https://doi.org/10.3150/12-BEJ486

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

KRITCHMAN, S. and NADLER, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemom. Intell. Lab. Syst.* **94** 19–32.

MALLAT, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way, with Contributions from Gabriel Peyré*. Elsevier/Academic Press, Amsterdam. MR2479996

MARCHENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* **114** 507–536. MR0208649

NADAKUDITI, R. R. (2014). OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Trans. Inform. Theory* **60** 3002–3018. MR3200641 https://doi.org/10.1109/TIT.2014.2311661

NADAKUDITI, R. R. and EDELMAN, A. (2008). Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Trans. Signal Process.* **56** 2625–2638. MR1500236 https://doi.org/10.1109/TSP.2008.917356

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

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348 https://doi.org/10.1214/10-AOS850

ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168** 244–258. MR2923766 https://doi.org/10.1016/j.jeconom.2012.01.034

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 https://doi.org/10.1214/13-AOS1100

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485 https://doi.org/10.1214/13-AOS1181

PASSEMIER, D. and YAO, J.-F. (2012). On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices Theory Appl.* **1** 1150002, 19. MR2930380 https://doi.org/10.1142/S201032631150002X

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

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342 https://doi.org/10.1214/10-AOS860

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (2009). *Variance Components. Wiley Series in Probability and Statistics* **391**. Wiley Interscience, Hoboken, NJ. MR2298115

SINGER, A. and WU, H.-T. (2013). Two-dimensional tomography from noisy projections taken at unknown random directions. *SIAM J. Imaging Sci.* **6** 136–175. MR3032950 https://doi.org/10.1137/090764657

SREBRO, N. and SALAKHUTDINOV, R. R. (2010). Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In *Advances in Neural Information Processing Systems* 2056–2064.

STEIN, E. M. and SHAKARCHI, R. (2011). *Fourier Analysis: An Introduction. Princeton Lectures in Analysis* **1**. Princeton Univ. Press, Princeton, NJ. MR1970295

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

ZHAO, Z., SHKOLNISKY, Y. and SINGER, A. (2016). Fast steerable principal component analysis. *IEEE Trans. Comput. Imag.* **2** 1–12. MR3472531 https://doi.org/10.1109/TCI.2016.2514700

The Annals of Statistics
2020, Vol. 48, No. 1, 514–538
https://doi.org/10.1214/19-AOS1820

AVERAGES OF UNLABELED NETWORKS: GEOMETRIC CHARACTERIZATION AND ASYMPTOTIC BEHAVIOR

BY ERIC D. KOLACZYK[1,*], LIZHEN LIN[2], STEVEN ROSENBERG[1,**], JACKSON WALTERS[1,†] AND JIE XU[1,‡]

[1]*Department of Mathematics and Statistics, Boston University, *kolaczyk@bu.edu; **sr@bu.edu; †jackwalt@bu.edu; ‡xujie@bu.edu*

[2]*Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, lizhen.lin@nd.edu*

It is becoming increasingly common to see large collections of network data objects, that is, data sets in which a network is viewed as a fundamental unit of observation. As a result, there is a pressing need to develop network-based analogues of even many of the most basic tools already standard for scalar and vector data. In this paper, our focus is on averages of unlabeled, undirected networks with edge weights. Specifically, we (i) characterize a certain notion of the space of all such networks, (ii) describe key topological and geometric properties of this space relevant to doing probability and statistics thereupon, and (iii) use these properties to establish the asymptotic behavior of a generalized notion of an empirical mean under sampling from a distribution supported on this space. Our results rely on a combination of tools from geometry, probability theory and statistical shape analysis. In particular, the lack of vertex labeling necessitates working with a quotient space modding out permutations of labels. This results in a nontrivial geometry for the space of unlabeled networks, which in turn is found to have important implications on the types of probabilistic and statistical results that may be obtained and the techniques needed to obtain them.

REFERENCES

[1] AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5

[2] ARROYO RELIÓN, J. D., KESSLER, D., LEVINA, E. and TAYLOR, S. F. (2017). Network classification with applications to brain connectomics. Preprint. Available at arXiv:1701.08140.

[3] AYDIN, B., PATAKI, G., WANG, H., BULLITT, E. and MARRON, J. S. (2009). A principal component analysis for trees. *Ann. Appl. Stat.* **3** 1597–1615. MR2752149 https://doi.org/10.1214/09-AOAS263

[4] BARDEN, D., LE, H. and OWEN, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab.* **18** no. 25, 25. MR3035753 https://doi.org/10.1214/EJP.v18-2201

[5] BHATTACHARYA, A. (2008). Statistical analysis on manifolds: A nonparametric approach for inference on shape spaces. *Sankhyā* **70** 223–266. MR2551814

[6] BHATTACHARYA, A. and BHATTACHARYA, R. (2012). *Nonparametric Inference on Manifolds: With Applications to Shape Spaces. Institute of Mathematical Statistics (IMS) Monographs* **2**. Cambridge Univ. Press, Cambridge. MR2934285 https://doi.org/10.1017/CBO9781139094764

[7] BHATTACHARYA, R. and LIN, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* **145** 413–428. MR3565392 https://doi.org/10.1090/proc/13216

[8] BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

[9] BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

[10] BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

MSC2010 subject classifications. Primary 62E20, 62G20; secondary 53C20.
Key words and phrases. Fundamental domain, Fréchet mean, graphs.

[11] BISWAL, B. B., MENNES, M., ZUO, X. N., GOHEL, S., KELLY, C., SMITH, S. M., BECKMANN, C. F., ADELSTEIN, J. S., BUCKNER, R. L. et al. (2010). Toward discovery science of human brain function. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **107** 4734–4739.

[12] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418 https://doi.org/10.1090/gsm/033

[13] CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J*. *Amer*. *Statist*. *Assoc*. **112** 1684–1696. MR3750891 https://doi.org/10.1080/01621459.2016.1228536

[14] 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

[15] COX, D. A. and KATZ, S. (1999). *Mirror Symmetry and Algebraic Geometry*. *Mathematical Surveys and Monographs* **68**. Amer. Math. Soc., Providence, RI. MR1677117 https://doi.org/10.1090/surv/068

[16] DO CARMO, M. P. (1992). *Riemannian Geometry*. *Mathematics*: *Theory & Applications*. Birkhäuser, Inc., Boston, MA. MR1138207 https://doi.org/10.1007/978-1-4757-2201-7

[17] DONALDSON, S. K. and KRONHEIMER, P. B. (1990). *The Geometry of Four-Manifolds*. *Oxford Mathematical Monographs*. Oxford University Press, New York. MR1079726

[18] DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann*. *Appl*. *Stat*. **3** 1102–1123. MR2750388 https://doi.org/10.1214/09-AOAS249

[19] DURANTE, D. and DUNSON, D. B. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika* **101** 883–898. MR3286923 https://doi.org/10.1093/biomet/asu040

[20] DURANTE, D., DUNSON, D. B. and VOGELSTEIN, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *J*. *Amer*. *Statist*. *Assoc*. **112** 1516–1530. MR3750873 https://doi.org/10.1080/01621459.2016.1219260

[21] GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. and KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann*. *Appl*. *Stat*. **11** 725–750. MR3693544 https://doi.org/10.1214/16-AOAS1015

[22] GJOKA, M., KURANT, M., BUTTS, C. T. and MARKOPOULOU, A. (2010). Walking in Facebook: A case study of unbiased sampling of OSNs. In *Infocom*, 2010 *Proceedings IEEE* 1–9. IEEE, New York.

[23] JACKSON, M. O. (2008). *Social and Economic Networks*. Princeton Univ. Press, Princeton, NJ. MR2435744

[24] JAIN, B. J. (2016). Statistical graph space analysis. *Pattern Recognit*. **60** 802–812.

[25] JAIN, B. J. (2016). On the geometry of graph spaces. *Discrete Appl*. *Math*. **214** 126–144. MR3546275 https://doi.org/10.1016/j.dam.2016.06.027

[26] KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc*. *Lond*. *Math*. *Soc*. (3) **61** 371–406. MR1063050 https://doi.org/10.1112/plms/s3-61.2.371

[27] KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data*: *Methods and Models*. *Springer Series in Statistics*. Springer, New York. MR2724362 https://doi.org/10.1007/978-0-387-88146-1

[28] KOLACZYK, E. D., LIN, L., ROSENBERG, S., WALTERS, J. and XU, J. (2019). Supplement to "Averages of unlabeled networks: Geometric characterization and asymptotic behavior." https://doi.org/10.1214/19-AOS1820SUPPA, https://doi.org/10.1214/19-AOS1820SUPPB, https://doi.org/10.1214/19-AOS1820SUPPC, https://doi.org/10.1214/19-AOS1820SUPPD, https://doi.org/10.1214/19-AOS1820SUPPE.

[29] LANGE, C. When is the underlying space of an orbifold a manifold with boundary? Preprint. Available at arXiv:1509.0679.

[30] LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035 https://doi.org/10.1090/coll/060

[31] MASON, D. M. and POLONIK, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann*. *Appl*. *Probab*. **19** 1108–1142. MR2537201 https://doi.org/10.1214/08-AAP569

[32] MCAULEY, J. and LESKOVEC, J. (2014). Discovering social circles in ego networks. *ACM Trans*. *Knowl*. *Discov*. *Data* **8** 4.

[33] MCCARTY, C. (2002). Structure in personal networks. *J*. *Soc*. *Struct*. **3** 20.

[34] MCDUFF, D. and SALAMON, D. (2012). *J-Holomorphic Curves and Symplectic Topology*, 2nd ed. *American Mathematical Society Colloquium Publications* **52**. Amer. Math. Soc., Providence, RI. MR2954391

[35] NEWMAN, M. E. J. (2010). *Networks*: *An Introduction*. Oxford Univ. Press, Oxford. MR2676073 https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

[36] TANG, R., KETCHA, M., VOGELSTEIN, J. T., PRIEBE, C. E. and SUSSMAN, D. L. (2016). Law of large graphs. Preprint. Available at arXiv:1609.01672.

[37] UGANDER, J., BACKSTROM, L. and KLEINBERG, J. (2013). Subgraph frequencies: Mapping the empirical and extremal geography of large graph collections. In *Proceedings of the* 22*nd International Conference on World Wide Web* 1307–1318. ACM, New York.

[38] VAKIL, R. (2003). The moduli space of curves and its tautological ring. *Notices Amer. Math. Soc.* **50** 647–658. MR1988577

[39] WANG, H. and MARRON, J. S. (2007). Object oriented data analysis: Sets of trees. *Ann. Statist.* **35** 1849–1873. MR2363955 https://doi.org/10.1214/009053607000000217

[40] ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory*, *Statistical Functions*, *Random Processes and of the Eightth European Meeting of Statisticians* **A** 591–602 (Tech. Univ. Prague, 1974). MR0501230

The Annals of Statistics
2020, Vol. 48, No. 1, 539–559
https://doi.org/10.1214/19-AOS1821

MARKOV EQUIVALENCE OF MARGINALIZED LOCAL INDEPENDENCE GRAPHS

BY SØREN WENGEL MOGENSEN* AND NIELS RICHARD HANSEN**

Department of Mathematical Sciences, University of Copenhagen, *swengel@math.ku.dk; **niels.r.hansen@math.ku.dk

Symmetric independence relations are often studied using graphical representations. Ancestral graphs or acyclic directed mixed graphs with m-separation provide classes of symmetric graphical independence models that are closed under marginalization. Asymmetric independence relations appear naturally for multivariate stochastic processes, for instance, in terms of local independence. However, no class of graphs representing such asymmetric independence relations, which is also closed under marginalization, has been developed. We develop the theory of directed mixed graphs with μ-separation and show that this provides a graphical independence model class which is closed under marginalization and which generalizes previously considered graphical representations of local independence.

Several graphs may encode the same set of independence relations and this means that in many cases only an equivalence class of graphs can be identified from observational data. For statistical applications, it is therefore pivotal to characterize graphs that induce the same independence relations. Our main result is that for directed mixed graphs with μ-separation each equivalence class contains a maximal element which can be constructed from the independence relations alone. Moreover, we introduce the directed mixed equivalence graph as the maximal graph with dashed and solid edges. This graph encodes all information about the edges that is identifiable from the independence relations, and furthermore it can be computed efficiently from the maximal graph.

REFERENCES

[1] AALEN, O. O. (1987). Dynamic modelling and causality. *Scand. Actuar. J.* **3–4** 177–190. MR0943579 https://doi.org/10.1016/j.rser.2011.04.029

[2] AALEN, O. O., BORGAN, Ø., KEIDING, N. and THORMANN, J. (1980). Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring. *Scand. J. Stat.* **7** 161–171. MR0605986

[3] AALEN, O. O., RØYSLAND, K., GRAN, J. M., KOUYOS, R. and LANGE, T. (2016). Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Stat. Methods Med. Res.* **25** 2294–2314. MR3553339 https://doi.org/10.1177/0962280213520436

[4] AALEN, O. O., RØYSLAND, K., GRAN, J. M. and LEDERGERBER, B. (2012). Causality, mediation and time: A dynamic viewpoint. *J. Roy. Statist. Soc. Ser. A* **175** 831–861. MR2993496 https://doi.org/10.1111/j.1467-985X.2011.01030.x

[5] ALI, R. A., RICHARDSON, T. S. and SPIRTES, P. (2009). Markov equivalence for ancestral graphs. *Ann. Statist.* **37** 2808–2837. MR2541448 https://doi.org/10.1214/08-AOS626

[6] COMMENGES, D. and GÉGOUT-PETIT, A. (2009). A general dynamical statistical model with causal interpretation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 719–736. MR2749916 https://doi.org/10.1111/j.1467-9868.2009.00703.x

[7] CONSTANTINOU, P. and DAWID, A. P. (2017). Extended conditional independence and applications in causal inference. *Ann. Statist.* **45** 2618–2653. MR3737904 https://doi.org/10.1214/16-AOS1537

[8] COX, D. R. and WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation. Monographs on Statistics and Applied Probability* **67**. CRC Press, London. MR1456990

MSC2010 subject classifications. Primary 62M99; secondary 62A99.

Key words and phrases. Directed mixed graphs, independence model, local independence, local independence graph, Markov equivalence, μ-separation.

[9] DANKS, D. and PLIS, S. (2013). Learning causal structure from undersampled time series. In *JMLR: Workshop and Conference Proceedings* (*NIPS Workshop on Causality*).
[10] DAWID, A. P. (2001). Separoids: A mathematical framework for conditional independence and irrelevance. *Ann. Math. Artif. Intell.* **32** 335–372. MR1859870 https://doi.org/10.1023/A:1016734104787
[11] DIDELEZ, V. (2000). *Graphical models for event history analysis based on local independence*. Dissertation, Univ. Dortmund. Logos Verlag Berlin, Berlin. MR1864064
[12] DIDELEZ, V. (2007). Graphical models for composable finite Markov processes. *Scand. J. Stat.* **34** 169–185. MR2325249 https://doi.org/10.1111/j.1467-9469.2006.00528.x
[13] DIDELEZ, V. (2008). Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 245–264. MR2412641 https://doi.org/10.1111/j.1467-9868.2007.00634.x
[14] EICHLER, M. (2012). Graphical modelling of multivariate time series. *Probab. Theory Related Fields* **153** 233–268. MR2925574 https://doi.org/10.1007/s00440-011-0345-8
[15] EICHLER, M. (2013). Causal inference with multiple time series: Principles and problems. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **371** 20110613, 17. MR3081173 https://doi.org/10.1098/rsta.2011.0613
[16] EICHLER, M. and DIDELEZ, V. (2007). Causal reasoning in graphical time series models. In *Proceedings of the* 23*rd Conference on Uncertainty in Artificial Intelligence* 109–116.
[17] EICHLER, M. and DIDELEZ, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16** 3–32. MR2575937 https://doi.org/10.1007/s10985-009-9143-3
[18] EVANS, R. J. (2016). Graphs for margins of Bayesian networks. *Scand. J. Stat.* **43** 625–648. MR3543314 https://doi.org/10.1111/sjos.12194
[19] EVANS, R. J. and RICHARDSON, T. S. (2014). Markovian acyclic directed mixed graphs for discrete data. *Ann. Statist.* **42** 1452–1482. MR3262457 https://doi.org/10.1214/14-AOS1206
[20] FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353. MR1096723
[21] GÉGOUT-PETIT, A. and COMMENGES, D. (2010). A general definition of influence between stochastic processes. *Lifetime Data Anal.* **16** 33–44. MR2575938 https://doi.org/10.1007/s10985-009-9131-7
[22] HYTTINEN, A., PLIS, S., JÄRVISALO, M., EBERHARDT, F. and DANKS, D. (2016). Causal discovery from subsampled time series data by constraint optimization. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models* **52** 216–227.
[23] JENSEN, A.-M. and SCHWEDER, T. (1986). The engine of fertility—Influenced by interbirth employment? Discussion paper, Central Bureau of Statistics, Oslo.
[24] KANDEL, D. (1975). Stages in adolescent involvement in drug use. *Science* **190** 912–914.
[25] 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
[26] LAURITZEN, S. and SADEGHI, K. (2018). Unifying Markov properties for graphical models. *Ann. Statist.* **46** 2251–2278. MR3845017 https://doi.org/10.1214/17-AOS1618
[27] LU, J. (2015). *On Causal Inference for Ordinal Outcomes*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Harvard Univ. MR3664965
[28] MEEK, C. (2014). Toward learning graphical and causal process models. In *Proceedings of the UAI* 2014 *Workshop Causal Inference*: *Learning and Prediction*.
[29] MOGENSEN, S. W. and HANSEN, N. R. (2020). Supplement to "Markov equivalence of marginalized local independence graphs." https://doi.org/10.1214/19-AOS1821SUPP.
[30] MOGENSEN, S. W., MALINSKY, D. and HANSEN, N. R. (2018). Causal learning for partially observed stochastic dynamical systems. In *Proceedings of the* 34*th Conference on Uncertainty in Artificial Intelligence*.
[31] PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161
[32] 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
[33] RICHARDSON, T. and SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015
[34] RICHARDSON, T. S., EVANS, R. J., ROBINS, J. M. and SHPITSER, I. (2017). Nested Markov properties for acyclic directed mixed graphs. https://arxiv.org/abs/1701.06686.
[35] RØYSLAND, K. (2012). Counterfactual analyses with graphical models based on local independence. *Ann. Statist.* **40** 2162–2194. MR3059080 https://doi.org/10.1214/12-AOS1031
[36] SADEGHI, K. (2013). Stable mixed graphs. *Bernoulli* **19** 2330–2358. MR3160556 https://doi.org/10.3150/12-BEJ454

[37] SCHWEDER, T. (1970). Composable Markov processes. *J. Appl. Probab.* **7** 400–410. MR0264755 https://doi.org/10.2307/3211973

[38] SOKOL, A. and HANSEN, N. R. (2014). Causal interpretation of stochastic differential equations. *Electron. J. Probab.* **19** no. 100, 24. MR3275852 https://doi.org/10.1214/ejp.v19-2891

[39] SPIRTES, P., RICHARDSON, T. S. and MEEK, C. (1997). The dimensionality of mixed ancestral graphs. Technical report. Philosophy Dept, Carnegie Mellon Univ., Pittsburgh, PA.

[40] VANYUKOV, M. M., TARTER, R. E., KIRILLOVA, G. P., KIRISCI, L., REYNOLDS, M. D., KREEK, M. J., CONWAY, K. P., MAHER, B. S., IACONO, W. G., BIERUT, L., NEALE, M. C., CLARK, D. B. and RIDENOUR, T. A. (2012). Common liability to addiction and "gateway hypothesis": Theoretical, empirical and evolutionary perspective. *Drug Alcohol Depend*. **123**.

[41] VERMA, T. and PEARL, J. (1991). Equivalence and synthesis of causal models Technical Report Univ. California, Los Angeles.

[42] VOLF, M. and STUDENÝ, M. (1999). A graphical characterization of the largest chain graphs. *Internat. J. Approx. Reason.* **20** 209–236. MR1685080 https://doi.org/10.1016/S0888-613X(99)00003-1

[43] WORLD HEALTH ORGANIZATION (2004). Neuroscience of psychoactive substance use and dependence Technical Report World Health Organization.

[44] XU, H., FARAJTABAR, M. and ZHA, H. (2016). Learning Granger causality for Hawkes processes. In *Proceedings of the* 33*rd International Conference on Machine Learning*.

[45] ZHAO, H., ZHENG, Z. and LIU, B. (2005). On the Markov equivalence of maximal ancestral graphs. *Sci. China Ser. A* **48** 548–562. MR2157690 https://doi.org/10.1360/04ys0023

The Annals of Statistics
2020, Vol. 48, No. 1, 560–583
https://doi.org/10.1214/19-AOS1823

ASYMPTOTIC GENEALOGIES OF INTERACTING PARTICLE SYSTEMS WITH AN APPLICATION TO SEQUENTIAL MONTE CARLO

BY JERE KOSKELA[1,*], PAUL A. JENKINS[2], ADAM M. JOHANSEN[1,**] AND DARIO SPANÒ[1,†]

[1]*Department of Statistics, University of Warwick,* [*]*j.koskela@warwick.ac.uk;* [**]*a.m.johansen@warwick.ac.uk;* [†]*d.spano@warwick.ac.uk*

[2]*Department of Statistics and Department of Computer Science, University of Warwick, p.jenkins@warwick.ac.uk*

We study weighted particle systems in which new generations are resampled from current particles with probabilities proportional to their weights. This covers a broad class of sequential Monte Carlo (SMC) methods, widely-used in applied statistics and cognate disciplines. We consider the genealogical tree embedded into such particle systems, and identify conditions, as well as an appropriate time-scaling, under which they converge to the Kingman n-coalescent in the infinite system size limit in the sense of finite-dimensional distributions. Thus, the tractable n-coalescent can be used to predict the shape and size of SMC genealogies, as we illustrate by characterising the limiting mean and variance of the tree height. SMC genealogies are known to be connected to algorithm performance, so that our results are likely to have applications in the design of new methods as well. Our conditions for convergence are strong, but we show by simulation that they do not appear to be necessary.

REFERENCES

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

CÉROU, F., DEL MORAL, P. and GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. MR2841068 https://doi.org/10.1214/10-AIHP358

CHAN, H. P. and LAI, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.* **41** 2877–2904. MR3161451 https://doi.org/10.1214/13-AOS1172

CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989 https://doi.org/10.1214/009053604000000698

CRISAN, D. and LYONS, T. (1997). Nonlinear filtering and measure-valued processes. *Probab. Theory Related Fields* **109** 217–244. MR1477650 https://doi.org/10.1007/s004400050131

DEL MORAL, P. (1996). Nonlinear filtering: Interacting particle solution. *Markov Process. Related Fields* **2** 555–579. MR1431187

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

DEL MORAL, P. (2013). *Mean Field Simulation for Monte Carlo Integration. Monographs on Statistics and Applied Probability* **126**. CRC Press, Boca Raton, FL. MR3060209

DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. MR2278333 https://doi.org/10.1111/j.1467-9868.2006.00553.x

DEL MORAL, P. and GARNIER, J. (2005). Genealogical particle analysis of rare events. *Ann. Appl. Probab.* **15** 2496–2534. MR2187302 https://doi.org/10.1214/105051605000000566

DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. MR1819122 https://doi.org/10.1016/S0246-0203(00)01064-5

MSC2010 subject classifications. Primary 60E15; secondary 60G99, 62E20.

Key words and phrases. Coalescent, finite-dimensional distributions, genealogy, interacting particle system, sequential Monte Carlo.

DEL MORAL, P. and MICLO, L. (2001). Genealogies and increasing propagation of chaos for Feynman–Kac and genetic models. *Ann. Appl. Probab.* **11** 1166–1198. MR1878294

DEL MORAL, P. and MICLO, L. (2007). Strong propagations of chaos in Moran's type particle interpretations of Feynman–Kac measures. *Stoch. Anal. Appl.* **25** 519–575. MR2321897 https://doi.org/10.1080/07362990701282666

DEL MORAL, P., MICLO, L., PATRAS, F. and RUBENTHALER, S. (2010). The convergence to equilibrium of neutral genetic models. *Stoch. Anal. Appl.* **28** 123–143. MR2597983 https://doi.org/10.1080/07362990903415833

DOUC, R., CAPPÉ, O. and MOULINES, E. (2005). Comparison of resampling schemes for particle filtering. In *Proceedings of the* 4*th International Symposium on Image and Signal Processing and Analysis* 64–69.

DOUCET, A. and JOHANSEN, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In *The Oxford Handbook of Nonlinear Filtering* 656–704. Oxford Univ. Press, Oxford. MR2884612

GORDON, N. J., SALMOND, S. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140** 107–113.

JACOB, P. E., MURRAY, L. M. and RUBENTHALER, S. (2015). Path storage in the particle filter. *Stat. Comput.* **25** 487–496. MR3306720 https://doi.org/10.1007/s11222-013-9445-x

KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034 https://doi.org/10.1016/0304-4149(82)90011-4

KONG, A., LIU, J. S. and WONG, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.* **89** 278–288.

KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077 https://doi.org/10.1214/009053605000000426

LEE, A. and WHITELEY, N. (2018). Variance estimation in the particle filter. *Biometrika* **105** 609–625. MR3842888 https://doi.org/10.1093/biomet/asy028

LIU, J. S. and CHEN, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.* **90** 567–576. MR3363399 https://doi.org/10.1080/01621459.1995.10476549

MÖHLE, M. (1998). Robustness results for the coalescent. *J. Appl. Probab.* **35** 438–447. MR1641829 https://doi.org/10.1239/jap/1032192859

MÖHLE, M. (1999). Weak convergence to the coalescent in neutral population models. *J. Appl. Probab.* **36** 446–460. MR1724816 https://doi.org/10.1017/s0021900200017241

MOSIMANN, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* **49** 65–82. MR0143299 https://doi.org/10.1093/biomet/49.1-2.65

OLSSON, J. and DOUC, R. (2019). Numerically stable online estimation of variance in particle filters. *Bernoulli* **25** 1504–1535. MR3920380 https://doi.org/10.3150/18-bej1028

WAKELEY, J. (2009). *Coalescent Theory: An Introduction*. Roberts & Co., Greenwood Village, CO.

The Annals of Statistics
2020, Vol. 48, No. 1, 584–606
https://doi.org/10.1214/19-AOS1829

ALMOST SURE UNIQUENESS OF A GLOBAL MINIMUM WITHOUT CONVEXITY

BY GREGORY COX

Department of Economics, Columbia University, gfc2106@gmail.com

This paper establishes the argmin of a random objective function to be unique almost surely. This paper first formulates a general result that proves almost sure uniqueness without convexity of the objective function. The general result is then applied to a variety of applications in statistics. Four applications are discussed, including uniqueness of M-estimators, both classical likelihood and penalized likelihood estimators, and two applications of the argmin theorem, threshold regression and weak identification.

REFERENCES

[1] ANDREWS, D. W. K. and CHENG, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* **80** 2153–2211. MR3013721 https://doi.org/10.3982/ECTA9456

[2] ARCONES, M. A. (1992). On the arg max of a Gaussian process. *Statist. Probab. Lett.* **15** 373–374. MR1193895 https://doi.org/10.1016/0167-7152(92)90156-Y

[3] CHENG, K. F. and CHEN, C. H. (1988). Estimation of the Weibull parameters with grouped data. *Comm. Statist. Theory Methods* **17** 325–341. MR0939654 https://doi.org/10.1080/03610928808829626

[4] CHENG, X. (2015). Robust inference in nonlinear models with mixed identification strength. *J. Econometrics* **189** 207–228. MR3397361 https://doi.org/10.1016/j.jeconom.2015.07.003

[5] COPAS, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62** 701–704. MR0388627 https://doi.org/10.1093/biomet/62.3.701

[6] COX, G. (2019). Weak identification with bounds in a class of minimum distance models. Unpublished manuscript.

[7] DEMIDENKO, E. (2008). Criteria for unconstrained global optimization. *J. Optim. Theory Appl.* **136** 375–395. MR2377513 https://doi.org/10.1007/s10957-007-9298-6

[8] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

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

[10] FAN, J., XUE, L. and ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *Ann. Statist.* **42** 819–849. MR3210988 https://doi.org/10.1214/13-AOS1198

[11] FERGER, D. (1999). On the uniqueness of maximizers of Markov–Gaussian processes. *Statist. Probab. Lett.* **45** 71–77. MR1718353 https://doi.org/10.1016/S0167-7152(99)00044-9

[12] GUILLEMIN, V. and POLLACK, A. (1974). *Differential Topology*. Prentice-Hall, Inc., Englewood Cliffs, NJ. MR0348781

[13] HAN, S. and MCCLOSKEY, A. (2018). Estimation and inference with a (nearly) singular Jacobian. *Quant. Econ.* To appear.

[14] HILL, D. L., SAUNDERS, R. and LAUD, P. W. (1980). Maximum likelihood estimation for mixtures. *Canad. J. Statist.* **8** 87–93. MR0595718 https://doi.org/10.2307/3314673

[15] HILLIER, G. and ARMSTRONG, M. (1999). The density of the maximum likelihood estimator. *Econometrica* **67** 1459–1470.

[16] HUANG, J., HOROWITZ, J. L. and MA, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808 https://doi.org/10.1214/009053607000000875

[17] JEWELL, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479–484. MR0653523

[18] KARLIN, S. (1968). *Total Positivity, Vol. I.* Stanford Univ. Press, Stanford, Calif. MR0230102

MSC2010 subject classifications. 60G17.

Key words and phrases. Global optimization, nonconvex optimization, M-estimation, mixture model, argmax theorem, threshold regression, weak identification.

[19] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[20] KIM, Y., CHOI, H. and OH, H.-S. (2008). Smoothly clipped absolute deviation on high dimensions. *J. Amer. Statist. Assoc.* **103** 1665–1673. MR2510294 https://doi.org/10.1198/016214508000001066

[21] LIFSHITS, M. (1982). On the absolute continuity of the distributions of functionals of stochastic processes. *Theory Probab. Appl.* **27** 600–607.

[22] LINDSAY, B., CLOGG, C. C. and GREGO, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Amer. Statist. Assoc.* **86** 96–107. MR1137102

[23] LINDSAY, B. G. (1983). The geometry of mixture likelihoods: A general theory. *Ann. Statist.* **11** 86–94. MR0684866 https://doi.org/10.1214/aos/1176346059

[24] LINDSAY, B. G. (1983). The geometry of mixture likelihoods, part II: the exponential family. *Ann. Statist.* **11** 783–792. MR0707929 https://doi.org/10.1214/aos/1176346245

[25] LINDSAY, B. G. and ROEDER, K. (1993). Uniqueness of estimation and identifiability in mixture models. *Canad. J. Statist.* **21** 139–147. MR1234757 https://doi.org/10.2307/3315807

[26] LOH, P.-L. and WAINWRIGHT, M. J. (2017). Support recovery without incoherence: A case for nonconvex regularization. *Ann. Statist.* **45** 2455–2482. MR3737898 https://doi.org/10.1214/16-AOS1530

[27] LÓPEZ, S. and PIMENTEL, L. (2016). On the location of the maximum of a process: Lévy, Gaussian, and multidimensional cases. Preprint. Available at arXiv:1611.02334.

[28] MÄKELÄINEN, T., SCHMIDT, K. and STYAN, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *Ann. Statist.* **9** 758–767. MR0619279

[29] MALLET, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika* **73** 645–656. MR0897856 https://doi.org/10.1093/biomet/73.3.645

[30] MALLIK, A., BANERJEE, M. and SEN, B. (2013). Asymptotics for p-value based threshold estimation in regression settings. *Electron. J. Stat.* **7** 2477–2515. MR3117104 https://doi.org/10.1214/13-EJS845

[31] MASCARENHAS, W. F. (2010). A Mountain Pass Lemma and its implications regarding the uniqueness of constrained minimizers. *Optimization* **60** 1121–1159. MR2860297 https://doi.org/10.1080/02331934.2010.527973

[32] MÜLLER, H.-G. and SONG, K.-S. (1996). A set-indexed process in a two-region image. *Stochastic Process. Appl.* **62** 87–101. MR1388764 https://doi.org/10.1016/0304-4149(95)00094-1

[33] OLSEN, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the Tobit model. *Econometrica* **46** 1211–1215. MR0508692 https://doi.org/10.2307/1911445

[34] ORME, C. (1989). On the uniqueness of the maximum likelihood estimator in truncated regression models. *Econometric Rev.* **8** 217–222. MR1029878 https://doi.org/10.1080/07474938908800171

[35] ORME, C. D. and RUUD, P. A. (2002). On the uniqueness of the maximum likelihood estimator. *Econom. Lett.* **75** 209–217. MR1889393 https://doi.org/10.1016/S0165-1765(01)00600-0

[36] PIMENTEL, L. P. R. (2014). On the location of the maximum of a continuous stochastic process. *J. Appl. Probab.* **51** 152–161. MR3189448 https://doi.org/10.1239/jap/1395771420

[37] ROŚ, B., BIJMA, F., DE MUNCK, J. C. and DE GUNST, M. C. M. (2016). Existence and uniqueness of the maximum likelihood estimator for models with a Kronecker product covariance structure. *J. Multivariate Anal.* **143** 345–361. MR3431436 https://doi.org/10.1016/j.jmva.2015.05.019

[38] SEREGIN, A. (2010). Uniqueness of the maximum likelihood estimator for k-monotone densities. *Proc. Amer. Math. Soc.* **138** 4511–4515. MR2680075 https://doi.org/10.1090/S0002-9939-2010-10496-3

[39] SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200–1209. MR0428582

[40] 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

[41] STOCK, J. H. and WRIGHT, J. H. (2000). GMM with weak identification. *Econometrica* **68** 1055–1096. MR1779144 https://doi.org/10.1111/1468-0262.00151

[42] TARONE, R. E. and GRUENHAGE, G. (1975). A note on the uniqueness of roots of the likelihood equations for vector-valued parameters. *J. Amer. Statist. Assoc.* **70** 903–904. MR0391368

[43] VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence*. Springer, Berlin.

[44] WALL, C. T. C. (2016). *Differential Topology*. *Cambridge Studies in Advanced Mathematics* **156**. Cambridge Univ. Press, Cambridge. MR3558600 https://doi.org/10.1017/CBO9781316597835

[45] WANG, W. and BICE, D. (1997). A model with mixed binary responses and censored observations. *Comm. Statist. Theory Methods* **26** 921–941.

[46] WOOD, G. R. (1999). Binomial mixtures: Geometric estimation of the mixing distribution. *Ann. Statist.* **27** 1706–1721. MR1742506 https://doi.org/10.1214/aos/1017939148

[47] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

[48] ZHANG, C.-H. and ZHANG, T. (2012). A general theory of concave regularization for high-dimensional sparse estimation problems. *Statist. Sci.* **27** 576–593. MR3025135 https://doi.org/10.1214/12-STS399

The Annals of Statistics
2020, Vol. 48, No. 1, 607–627
https://doi.org/10.1214/19-AOS1870

PENALIZED GENERALIZED EMPIRICAL LIKELIHOOD WITH A DIVERGING NUMBER OF GENERAL ESTIMATING EQUATIONS FOR CENSORED DATA

BY NIANSHENG TANG[1], XIAODONG YAN[2] AND XINGQIU ZHAO[3]

[1]*School of Mathematics and Statistics, Yunnan University, nstang@ynu.edu.cn*
[2]*School of Economics, Shandong University, yanxiaodong@sdu.edu.cn*
[3]*Department of Applied Mathematics, Hong Kong Polytechnic University, xingqiu.zhao@polyu.edu.hk*

This article considers simultaneous variable selection and parameter estimation as well as hypothesis testing in censored survival models where a parametric likelihood is not available. For the problem, we utilize certain growing dimensional general estimating equations and propose a penalized generalized empirical likelihood, where the general estimating equations are constructed based on the semiparametric efficiency bound of estimation with given moment conditions. The proposed penalized generalized empirical likelihood estimators enjoy the oracle properties, and the estimator of any fixed dimensional vector of nonzero parameters achieves the semiparametric efficiency bound asymptotically. Furthermore, we show that the penalized generalized empirical likelihood ratio test statistic has an asymptotic central chi-square distribution. The conditions of local and restricted global optimality of weighted penalized generalized empirical likelihood estimators are also discussed. We present a two-layer iterative algorithm for efficient implementation, and investigate its convergence property. The performance of the proposed methods is demonstrated by extensive simulation studies, and a real data example is provided for illustration.

REFERENCES

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

ANTONIADIS, A., FRYZLEWICZ, P. and LETUÉ, F. (2010). The Dantzig selector in Cox's proportional hazards model. *Scand. J. Stat.* **37** 531–552. MR2779635 https://doi.org/10.1111/j.1467-9469.2009.00685.x

BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

CAO, Y., HUANG, J., LIU, Y. and ZHAO, X. (2016). Sieve estimation of Cox models with latent structures. *Biometrics* **72** 1086–1097. MR3591593 https://doi.org/10.1111/biom.12529

CHANG, J., CHEN, S. X. and CHEN, X. (2015). High dimensional generalized empirical likelihood for moment restrictions with dependent data. *J. Econometrics* **185** 283–304. MR3300347 https://doi.org/10.1016/j.jeconom.2014.10.011

CHANG, J., TANG, C. Y. and WU, T. T. (2017). A new scope of penalized empirical likelihood with high-dimensional estimating equations. Preprint. Available at arXiv:1704.00566.

CHEN, S. X., PENG, L. and QIN, Y.-L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96** 711–722. MR2538767 https://doi.org/10.1093/biomet/asp037

CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18** 415–447. MR2566404 https://doi.org/10.1007/s11749-009-0159-5

COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

DABROWSKA, D. M. (1989). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *Ann. Statist.* **17** 1157–1167. MR1015143 https://doi.org/10.1214/aos/1176347261

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

MSC2010 subject classifications. Primary 62N02, 62N03; secondary 62F12, 62F05.
Key words and phrases. Censored survival data, penalized generalized empirical likelihood, penalized generalized empirical likelihood ratio test, oracle property, semiparametric efficiency.

FAN, J. and LI, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Ann. Statist.* **30** 74–99. MR1892656 https://doi.org/10.1214/aos/1015362185

FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory* **57** 5467–5484. MR2849368 https://doi.org/10.1109/TIT.2011.2158486

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

FANG, K.-T., LI, G., LU, X. and QIN, H. (2013). An empirical likelihood method for semiparametric linear regression with right censored data. *Comput. Math. Methods Med.* Art. ID 469373. MR3032121 https://doi.org/10.1155/2013/469373

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics.* Wiley, New York. MR1100924

GRAMBSCH, P. M., THERNEAU, T. M. and FLEMING, T. R. (1995). Diagnostic plots to reveal functional form for covariates in multiplicative intensity models. *Biometrics* **51** 1469–1482.

HANSEN, L. P., HEATON, J. and YARON, A. (1996). Finite-sample properties of some alternative GMM estimators. *J. Bus. Econom. Statist.* **14** 262–280.

HE, X., WANG, L. and HONG, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41** 342–369. MR3059421 https://doi.org/10.1214/13-AOS1087

HE, S., LIANG, W., SHEN, J. and YANG, G. (2016). Empirical likelihood for right censored lifetime data. *J. Amer. Statist. Assoc.* **111** 646–655. MR3538694 https://doi.org/10.1080/01621459.2015.1024058

JIN, Z., LIN, D. Y., WEI, L. J. and YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90** 341–353. MR1986651 https://doi.org/10.1093/biomet/90.2.341

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data. Wiley Series in Probability and Mathematical Statistics.* Wiley, New York. MR0570114

KITAMURA, Y. and STUTZER, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **65** 861–874. MR1458431 https://doi.org/10.2307/2171942

LENG, C. and MA, S. (2007). Path consistent model selection in additive risk model via Lasso. *Stat. Med.* **26** 3753–3770. MR2395831 https://doi.org/10.1002/sim.2834

LENG, C. and TANG, C. Y. (2012). Penalized empirical likelihood and growing dimensional general estimating equations. *Biometrika* **99** 703–716. MR2966779 https://doi.org/10.1093/biomet/ass014

LI, G. and WANG, Q.-H. (2003). Empirical likelihood regression analysis for right censored data. *Statist. Sinica* **13** 51–68. MR1963919

LIN, D. Y. (2003). Regression analysis of incomplete medical cost data. *Stat. Med.* **15** 1181–1200.

LIN, W. and LV, J. (2013). High-dimensional sparse additive hazards regression. *J. Amer. Statist. Assoc.* **108** 247–264. MR3174617 https://doi.org/10.1080/01621459.2012.746068

LIN, D. Y. and YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81** 61–71. MR1279656 https://doi.org/10.1093/biomet/81.1.61

LV, J. and FAN, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.* **37** 3498–3528. MR2549567 https://doi.org/10.1214/09-AOS683

MARTINUSSEN, T. and SCHEIKE, T. H. (2009). Covariate selection for the semiparametric additive risk model. *Scand. J. Stat.* **36** 602–619. MR2572578 https://doi.org/10.1111/j.1467-9469.2009.00650.x

NEWEY, W. (1990). Semiparametric efficiency bounds. *J. Appl. Econometrics* **5** 99–135.

NEWEY, W. K. and SMITH, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72** 219–255. MR2031017 https://doi.org/10.1111/j.1468-0262.2004.00482.x

OWEN, A. B. (2001). *Empirical Likelihood.* CRC Press, New York.

PARENTE, P. M. D. C. and SMITH, R. J. (2011). GEL methods for nonsmooth moment indicators. *Econometric Theory* **27** 74–113. MR2771012 https://doi.org/10.1017/S0266466610000137

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085 https://doi.org/10.1214/aos/1176325370

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328. MR1041395 https://doi.org/10.1214/aos/1176347502

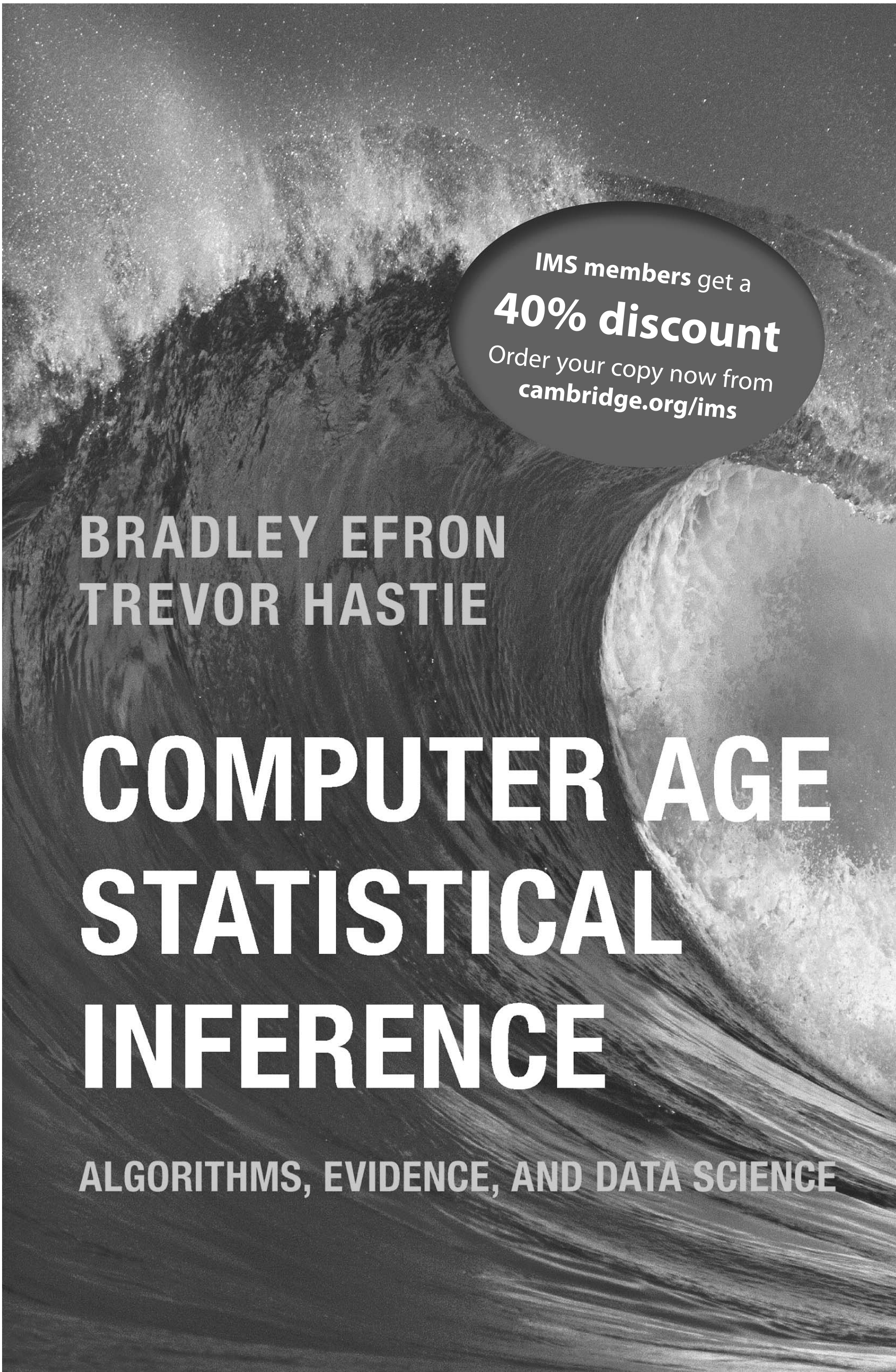
SUN, Q., ZHU, R., WANG, T. and ZENG, D. (2019). Counting process-based dimension reduction methods for censored outcomes. *Biometrika* **106** 181–196. MR3912390 https://doi.org/10.1093/biomet/asy064

TANG, C. Y. and LENG, C. (2010). Penalized high-dimensional empirical likelihood. *Biometrika* **97** 905–919. MR2746160 https://doi.org/10.1093/biomet/asq057

TANG, N., YAN, X. and ZHAO, X. (2019). Supplement to "Penalized generalized empirical likelihood with a diverging number of general estimating equations for censored data." https://doi.org/10.1214/19-AOS1870SUPP.

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

TIBSHIRANI, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395.
TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 303–328.
WU, T. T., LI, G. and TANG, C. (2015). Empirical likelihood for censored linear regression and variable selection. *Scand. J. Stat.* **42** 798–812. MR3391693 https://doi.org/10.1111/sjos.12137
ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729
ZHANG, H. H. and LU, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika* **94** 691–703. MR2410017 https://doi.org/10.1093/biomet/asm037
ZHOU, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika* **92** 492–498. MR2201374 https://doi.org/10.1093/biomet/92.2.492
ZHOU, M. and LI, G. (2008). Empirical likelihood analysis of the Buckley–James estimator. *J. Multivariate Anal.* **99** 649–664. MR2406076 https://doi.org/10.1016/j.jmva.2007.02.007
ZHOU, X. H., QIN, G. S., LIN, H. Z. and LI, G. (2006). Inferences in censored cost regression models with empirical likelihood. *Statist. Sinica* **16** 1213–1232. MR2327487
ZHU, R., ZHANG, J., ZHAO, R., XU, P., ZHOU, W. and ZHANG, X. (2018). orthoDr: Semiparametric dimension reduction via orthogonality constrained optimization. Available at arXiv:1811.11733.
ZOU, H. (2008). A note on path-based variable selection in the penalized proportional hazards model. *Biometrika* **95** 241–247. MR2409726 https://doi.org/10.1093/biomet/asm083


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