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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 45, No. 1—February 2017

The Annals of Statistics
2017, Vol. 45, No. 1, 1–38
DOI: 10.1214/15-AOS1414

TENSOR DECOMPOSITIONS AND SPARSE LOG-LINEAR MODELS

BY JAMES E. JOHNDROW[*,1], ANIRBAN BHATTACHARYA[†,2] AND DAVID B. DUNSON[*,1]

Duke University[*] and Texas A&M University[†]

Contingency table analysis routinely relies on log-linear models, with latent structure analysis providing a common alternative. Latent structure models lead to a reduced rank tensor factorization of the probability mass function for multivariate categorical data, while log-linear models achieve dimensionality reduction through sparsity. Little is known about the relationship between these notions of dimensionality reduction in the two paradigms. We derive several results relating the support of a log-linear model to nonnegative ranks of the associated probability tensor. Motivated by these findings, we propose a new collapsed Tucker class of tensor decompositions, which bridge existing PARAFAC and Tucker decompositions, providing a more flexible framework for parsimoniously characterizing multivariate categorical data. Taking a Bayesian approach to inference, we illustrate empirical advantages of the new decompositions.

REFERENCES

[1] AGRESTI, A. (2002). *Categorical Data Analysis*, 2nd ed. Wiley, New York. MR1914507
[2] ANDERSON, T. W. (1954). On estimation of parameters in latent structure analysis. *Psychometrika* **19** 1–10. MR0075492
[3] BHATTACHARYA, A. and DUNSON, D. B. (2012). Simplex factor models for multivariate unordered categorical data. *J. Amer. Statist. Assoc.* **107** 362–377. MR2949366
[4] BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (2007). *Discrete Multivariate Analysis*: *Theory and Practice*. Springer, New York. MR2344876
[5] COHEN, J. E. and ROTHBLUM, U. G. (1993). Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra Appl.* **190** 149–168. MR1230356
[6] DAHINDEN, C., KALISCH, M. and BÜHLMANN, P. (2010). Decomposition and model selection for large contingency tables. *Biom. J.* **52** 233–252. MR2756875
[7] DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539. MR0568718
[8] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. MR1241267
[9] DELLAPORTAS, P. and FORSTER, J. J. (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika* **86** 615–633. MR1723782
[10] DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** 1253–1278 (electronic). MR1780272

MSC2010 subject classifications. 62F15.
Key words and phrases. Bayesian, categorical data, contingency table, latent class analysis, graphical model, high-dimensional, low rank, Parafac, Tucker, sparsity.

[11] De Lathauwer, L. De Moor, B. and Vandewalle, J. (2000). On the best rank-1 and rank-$(r_1, r_2, \ldots, r_n)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21** 1324–1342.
[12] Dobra, A., Hans, C., Jones, B., Nevins, J. R., Yao, G. and West, M. (2004). Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.* **90** 196–212. MR2064941
[13] Dobra, A. and Lenkoski, A. (2011). Copula Gaussian graphical models and their application to modeling functional disability data. *Ann. Appl. Stat.* **5** 969–993. MR2840183
[14] Dobra, A. and Massam, H. (2010). The mode oriented stochastic search (MOSS) algorithm for log-linear models with conjugate priors. *Stat. Methodol.* **7** 240–253. MR2643600
[15] Dunson, D. B. and Xing, C. (2009). Nonparametric Bayes modeling of multivariate categorical data. *J. Amer. Statist. Assoc.* **104** 1042–1051. MR2562004
[16] Fienberg, S. E., Hersh, P., Rinaldo, A. and Zhou, Y. (2010). Maximum likelihood estimation in latent class models for contingency table data. In *Algebraic and Geometric Methods in Statistics* 27–62. Cambridge Univ. Press, Cambridge. MR2642657
[17] Fienberg, S. E. and Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *J. Statist. Plann. Inference* **137** 3430–3445. MR2363267
[18] Garcia, L. D., Stillman, M. and Sturmfels, B. (2005). Algebraic geometry of Bayesian networks. *J. Symbolic Comput.* **39** 331–355. MR2168286
[19] Geiger, D., Heckerman, D., King, H. and Meek, C. (2001). Stratified exponential families: Graphical models and model selection. *Ann. Statist.* **29** 505–529. MR1863967
[20] Gibson, W. A. (1955). An extension of Anderson's solution for the latent structure equations. *Psychometrika* **20** 69–73. MR0075493
[21] Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61** 215–231. MR0370936
[22] Gregory, D. A. and Pullman, N. J. (1983). Semiring rank: Boolean rank and nonnegative rank factorizations. *J. Comb. Inf. Syst. Sci.* **8** 223–233. MR0783759
[23] Haberman, S. J. (1974). Log-linear models for frequency tables derived by indirect observation: Maximum likelihood equations. *Ann. Statist.* **2** 911–924. MR0458687
[24] Harshman, R. A. (1970). Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis. UCLA Working Papers in Phonetics **16** 1–84.
[25] Hu, J., Joshi, A. and Johnson, V. E. (2009). Log-linear models for gene association. *J. Amer. Statist. Assoc.* **104** 597–607. MR2751441
[26] Humphreys, K. and Titterington, D. M. (2003). Variational approximations for categorical causal modeling with latent variables. *Psychometrika* **68** 391–412. MR2272386
[27] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729
[28] Johndrow, J. E., Battacharya, A. and Dunson, D. B. (2016). Supplement to "Tensor decompositions and sparse log-linear models." DOI:10.1214/15-AOS1414SUPP.
[29] Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056
[30] Kunihama, T. and Dunson, D. B. (2013). Bayesian modeling of temporal dependence in large sparse contingency tables. *J. Amer. Statist. Assoc.* **108** 1324–1338. MR3174711
[31] Lauritzen, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Oxford Univ. Press, New York. MR1419991
[32] Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*. Houghton, Mifflin, New York.
[33] Letac, G. and Massam, H. (2012). Bayes factors and the geometry of discrete hierarchical loglinear models. *Ann. Statist.* **40** 861–890. MR2985936

[34] LIM, L.-H. and COMON, P. (2009). Nonnegative approximations of nonnegative tensors. *J. Chemom*. **23** 432–441.

[35] MADANSKY, A. (1960). Determinantal methods in latent class analysis. *Psychometrika* **25** 183–197. MR0112763

[36] MASSAM, H., LIU, J. and DOBRA, A. (2009). A conjugate prior for discrete hierarchical log-linear models. *Ann. Statist*. **37** 3431–3467. MR2549565

[37] NARDI, Y. and RINALDO, A. (2012). The log-linear group-lasso estimator and its asymptotic properties. *Bernoulli* **18** 945–974. MR2948908

[38] ROTH, V. and FISCHER, B. (2008). The group-lasso for generalized linear models: Uniqueness of solutions and efficient algorithms. In *Proceedings of the* 25*th International Conference on Machine Learning* 848–855. ACM, New York.

[39] RUSAKOV, D. and GEIGER, D. (2002). Asymptotic model selection for naive Bayesian networks. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence* 438–455. Morgan Kaufmann, San Francisco, CA.

[40] SETTIMI, R. and SMITH, J. Q. (1998). On the geometry of Bayesian graphical models with hidden variables. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* 472–479. Morgan Kaufmann, San Francisco, CA.

[41] SMITH, J. Q. and CROFT, J. (2003). Bayesian networks for discrete multivariate data: An algebraic approach to inference. *J. Multivariate Anal*. **84** 387–402. MR1965229

[42] STOUFFER, S. A., GUTTMAN, L., SUCHMAN, E. A., LAZARSFELD, P. F., STAR, S. A. and CLAUSEN, J. A. (1950). Measurement and prediction. Princeton Univ. Press, Princeton, NJ.

[43] TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395

[44] ZHOU, J., BHATTACHARYA, A., HERRING, A. H. and DUNSON, D. B. (2015). Bayesian factorizations of big sparse tensors. *J. Amer. Statist. Assoc*. **110** 1562–1576. MR3449055

The Annals of Statistics
2017, Vol. 45, No. 1, 39–76
DOI: 10.1214/16-AOS1434

A LAVA ATTACK ON THE RECOVERY OF SUMS OF DENSE AND SPARSE SIGNALS

BY VICTOR CHERNOZHUKOV, CHRISTIAN HANSEN[1] AND YUAN LIAO

Massachusetts Institute of Technology, University of Chicago and University of Maryland

Common high-dimensional methods for prediction rely on having either a sparse signal model, a model in which most parameters are zero and there are a small number of nonzero parameters that are large in magnitude, or a dense signal model, a model with no large parameters and very many small nonzero parameters. We consider a generalization of these two basic models, termed here a "sparse + dense" model, in which the signal is given by the sum of a sparse signal and a dense signal. Such a structure poses problems for traditional sparse estimators, such as the lasso, and for traditional dense estimation methods, such as ridge estimation. We propose a new penalization-based method, called lava, which is computationally efficient. With suitable choices of penalty parameters, the proposed method strictly dominates both lasso and ridge. We derive analytic expressions for the finite-sample risk function of the lava estimator in the Gaussian sequence model. We also provide a deviation bound for the prediction risk in the Gaussian regression model with fixed design. In both cases, we provide Stein's unbiased estimator for lava's prediction risk. A simulation example compares the performance of lava to lasso, ridge and elastic net in a regression example using data-dependent penalty parameters and illustrates lava's improved performance relative to these benchmarks.

REFERENCES

[1] BANERJEE, A., CHEN, S., FAZAYELI, F. and SIVAKUMAR, V. (2015). Estimation with norm regularization. Available at arXiv:1505.02294 [stat.ML].

[2] BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163

[3] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806. MR2860324

[4] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2014). Pivotal estimation via square-root Lasso in nonparametric regression. *Ann. Statist.* **42** 757–788. MR3210986

[5] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

[6] BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2007). Sparsity oracle inequalities for the Lasso. *Electron. J. Stat.* **1** 169–194. MR2312149

[7] BUNEA, F., TSYBAKOV, A. B., WEGKAMP, M. H. and BARBU, A. (2010). Spades and mixture models. *Ann. Statist.* **38** 2525–2558. MR2676897

MSC2010 subject classifications. Primary 62J07; secondary 62J05.

Key words and phrases. High-dimensional models, penalization, shrinkage, nonsparse signal recovery.

[8] CAI, T. T., LIANG, T. and RAKHLIN, A. (2016). Geometric inference for general high-dimensional linear inverse problems. *Ann. Statist.* **44** 1536–1563. MR3519932

[9] CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

[10] CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** Art. 11, 37. MR2811000

[11] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A. and WILLSKY, A. S. (2012). The convex geometry of linear inverse problems. *Found. Comput. Math.* **12** 805–849. MR2989474

[12] CHANDRASEKARAN, V., SANGHAVI, S., PARRILO, P. A. and WILLSKY, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* **21** 572–596. MR2817479

[13] CHEN, Y. and DALALYAN, A. (2012). Fused sparsity and robust estimation for linear models with unknown variance. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*.

[14] CHERNOZHUKOV, V., HANSEN, C. and LIAO, Y. (2016). Supplement to "A lava attack on the recovery of sums of dense and sparse signals." DOI:10.1214/16-AOS1434SUPP.

[15] DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. MR1379464

[16] DOSSAL, C., KACHOUR, M., FADILI, M. J., PEYRÉ, G. and CHESNEAU, C. (2013). The degrees of freedom of the Lasso for general design matrix. *Statist. Sinica* **23** 809–828. MR3086657

[17] EFRON, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation. *J. Amer. Statist. Assoc.* **99** 619–642. MR2090899

[18] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166

[19] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

[20] FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653

[21] 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

[22] FRANK, L. E. and FRIEDMAN, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109–135.

[23] HIRANO, K. and PORTER, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica* **80** 1769–1790. MR2977437

[24] HORN, R. A. and JOHNSON, C. R. (2012). *Matrix Analysis*. Cambridge University Press, Cambridge.

[25] HSU, D., KAKADE, S. M. and ZHANG, T. (2014). Random design analysis of ridge regression. *Found. Comput. Math.* **14** 569–600. MR3201956

[26] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101. MR0161415

[27] JALALI, A., RAVIKUMAR, P. and SANGHAVI, S. (2013). A dirty model for multiple sparse regression. *IEEE Trans. Inform. Theory* **59** 7947–7968. MR3142275

[28] KLOPP, O., LOUNICI, K. and TSYBAKOV, A. B. (2014). Robust matrix completion. Preprint. Available at arXiv:1412.8132.

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

[30] MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.* **37** 246–270. MR2488351

[31] MEYER, M. and WOODROOFE, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist.* **28** 1083–1104. MR1810920

[32] NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. MR3025133

[33] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. i* 197–206. Univ. California Press, Berkeley and Los Angeles. MR0084922

[34] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

[35] SUN, T. and ZHANG, C. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166

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

[37] TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 91–108. MR2136641

[38] TIBSHIRANI, R. J. and TAYLOR, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.* **40** 1198–1232. MR2985948

[39] 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

[40] YANG, E. and RAVIKUMAR, P. K. (2013). Dirty statistical models. In *Advances in Neural Information Processing Systems* 26 (C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Weinberger, eds.) 611–619. Curran Associates, Red Hook, New York. Available at http://papers.nips.cc/paper/5092-dirty-statistical-models.pdf.

[41] 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

[42] ZHANG, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

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

[44] 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

The Annals of Statistics
2017, Vol. 45, No. 1, 77–120
DOI: 10.1214/16-AOS1435

STATISTICAL GUARANTEES FOR THE EM ALGORITHM: FROM POPULATION TO SAMPLE-BASED ANALYSIS[1]

BY SIVARAMAN BALAKRISHNAN[*,†], MARTIN J. WAINWRIGHT[†] AND BIN YU[†]

University of California, Berkeley[*] and Carnegie Mellon University[†]

The EM algorithm is a widely used tool in maximum-likelihood estimation in incomplete data problems. Existing theoretical work has focused on conditions under which the iterates or likelihood values converge, and the associated rates of convergence. Such guarantees do not distinguish whether the ultimate fixed point is a near global optimum or a bad local optimum of the sample likelihood, nor do they relate the obtained fixed point to the global optima of the idealized population likelihood (obtained in the limit of infinite data). This paper develops a theoretical framework for quantifying when and how quickly EM-type iterates converge to a small neighborhood of a given global optimum of the population likelihood. For correctly specified models, such a characterization yields rigorous guarantees on the performance of certain two-stage estimators in which a suitable initial pilot estimator is refined with iterations of the EM algorithm. Our analysis is divided into two parts: a treatment of the EM and first-order EM algorithms at the population level, followed by results that apply to these algorithms on a finite set of samples. Our conditions allow for a characterization of the region of convergence of EM-type iterates to a given population fixed point, that is, the region of the parameter space over which convergence is guaranteed to a point within a small neighborhood of the specified population fixed point. We verify our conditions and give tight characterizations of the region of convergence for three canonical problems of interest: symmetric mixture of two Gaussians, symmetric mixture of two regressions and linear regression with covariates missing completely at random.

REFERENCES

[1] ANANDKUMAR, A., GE, R., HSU, D., KAKADE, S. M. and TELGARSKY, M. (2012). Tensor decompositions for learning latent variable models. Preprint. Available at arXiv:1210.7559.

[2] ANANDKUMAR, A., JAIN, P., NETRAPALLI, P. and TANDON, R. (2013). Learning sparsely used overcomplete dictionaries via alternating minimization. Technical report, Microsoft Research, Redmond, OR.

[3] BALAKRISHNAN, S., WAINWRIGHT, M. J. and YU, B. (2014). Supplement to "Statistical guarantees for the EM algorithm: From population to sample-based analysis." DOI:10.1214/16-AOS1435SUPP.

MSC2010 subject classifications. Primary 62F10, 60K35; secondary 90C30.

Key words and phrases. EM algorithm, first-order EM algorithm, nonconvex optimization, maximum likelihood estimation.

[4] BALAN, R., CASAZZA, P. and EDIDIN, D. (2006). On signal reconstruction without phase. *Appl. Comput. Harmon. Anal.* **20** 345–356. MR2224902

[5] BAUM, L. E., PETRIE, T., SOULES, G. and WEISS, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* **41** 164–171. MR0287613

[6] BEALE, E. M. L. and LITTLE, R. J. A. (1975). Missing values in multivariate analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **37** 129–145. MR0373113

[7] BERTSEKAS, D. P. (1995). *Nonlinear Programming*. Athena Scientific, Belmont, CA.

[8] BUBECK, S. (2014). Theory of convex optimization for machine learning. Unpublished manuscript.

[9] BULDYGIN, V. V. and KOZACHENKO, YU. V. (2000). *Metric Characterization of Random Variables and Random Processes*. *Translations of Mathematical Monographs* **188**. Amer. Math. Soc., Providence, RI. Translated from the 1998 Russian original by V. Zaiats. MR1743716

[10] CANDÈS, E. J., STROHMER, T. and VORONINSKI, V. (2013). PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.* **66** 1241–1274. MR3069958

[11] CELEUX, G., CHAUVEAU, D. and DIEBOLT, J. (1995). On stochastic versions of the EM algorithm. Technical Report, No. 2514, INRIA.

[12] CELEUX, G. and GOVAERT, G. (1992). A classification EM algorithm for clustering and two stochastic versions. *Comput. Statist. Data Anal.* **14** 315–332. MR1192205

[13] CHAGANTY, A. T. and LIANG, P. (2013). Spectral experts for estimating mixtures of linear regressions. Unpublished manuscript.

[14] CHEN, Y., YI, X. and CARAMANIS, C. (2013). A convex formulation for mixed regression: Near optimal rates in the face of noise. Unpublished manuscript.

[15] CHRÉTIEN, S. and HERO, A. O. (2008). On EM algorithms and their proximal generalizations. *ESAIM Probab. Stat.* **12** 308–326. MR2404033

[16] DASGUPTA, S. and SCHULMAN, L. (2007). A probabilistic analysis of EM for mixtures of separated, spherical Gaussians. *J. Mach. Learn. Res.* **8** 203–226. MR2320668

[17] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **39** 1–38. MR0501537

[18] HARTLEY, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics* **14** 174–194.

[19] HEALY, M. and WESTMACOTT, M. (1956). Missing values in experiments analysed on automatic computers. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **5** 203–206.

[20] HERO, A. O. and FESSLER, J. A. (1995). Convergence in norm for alternating expectation-maximization (EM) type algorithms. *Statist. Sinica* **5** 41–54. MR1329288

[21] HSU, D. and KAKADE, S. M. (2012). Learning Gaussian mixture models: Moment methods and spectral decompositions. Preprint. Available at arXiv:1206.5766.

[22] ITURRIA, S. J., CARROLL, R. J. and FIRTH, D. (1999). Polynomial regression and estimating functions in the presence of multiplicative measurement error. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 547–561. MR1707860

[23] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. *Lecture Notes in Math.* **2033**. Springer, Heidelberg. MR2829871

[24] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. Springer, Berlin. MR1102015

[25] LIU, C. and RUBIN, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81** 633–648. MR1326414

[26] LOH, P. and WAINWRIGHT, M. J. (2012). Corrupted and missing predictors: Minimax bounds for high-dimensional linear regression. In *ISIT* 2601–2605. IEEE, Piscataway Township, NJ.
[27] LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **44** 226–233. MR0676213
[28] MA, J. and XU, L. (2005). Asymptotic convergence properties of the EM algorithm with respect to the overlap in the mixture. *Neurocomputing* **68** 105–129.
[29] MCLACHLAN, G. and KRISHNAN, T. (2007). *The EM Algorithm and Extensions*. Wiley, New York.
[30] MEILIJSON, I. (1989). A fast improvement to the EM algorithm on its own terms. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **51** 127–138. MR0984999
[31] MENG, X. (1994). On the rate of convergence of the ECM algorithm. *Ann. Statist.* **22** 326–339. MR1272086
[32] MENG, X. and RUBIN, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80** 267–278. MR1243503
[33] MENG, X. and RUBIN, D. B. (1994). On the global and componentwise rates of convergence of the EM algorithm. *Linear Algebra Appl.* **199** 413–425. MR1274429
[34] NEAL, R. M. and HINTON, G. E. (1999). A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models* (M. I. Jordan, ed.) 355–368. MIT Press, Cambridge, MA.
[35] NESTEROV, Y. (2004). *Introductory Lectures on Convex Optimization*: *A Basic Course*. *Applied Optimization* **87**. Kluwer Academic, Boston, MA. MR2142598
[36] NETRAPALLI, P., JAIN, P. and SANGHAVI, S. (2013). Phase retrieval using alternating minimization. In *Neural Information Processing Systems* Curran Associates, Inc., Red Hook, New York 2796–2804.
[37] ORCHARD, T. and WOODBURY, M. A. (1972). A missing information principle: Theory and applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Theory of Statistics* **1** 697–715. Univ. California Press, Berkeley, CA. MR0400516
[38] PEARSON, K. (1894). *Contributions to the Mathematical Theory of Evolution*. Harrison and Sons, London.
[39] REDNER, R. A. and WALKER, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26** 195–239. MR0738930
[40] RUBIN, D. B. (1974). Characterizing the estimation of parameters in incomplete-data problems. *J. Amer. Statist. Assoc.* **69** 467–474.
[41] SUNDBERG, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scand. J. Stat.* **1** 49–58. MR0381110
[42] TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550. MR0898357
[43] TSENG, P. (2004). An analysis of the EM algorithm and entropy-like proximal point methods. *Math. Oper. Res.* **29** 27–44. MR2065712
[44] VAN DYK, D. A. and MENG, X. L. (2000). Algorithms based on data augmentation: A graphical representation and comparison. In *Computing Science and Statistics*: *Proceedings of the* 31*st Symposium on the Interface* 230–239. Interface Foundation of North America, Fairfax Station, VA.
[45] VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.
[46] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671
[47] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

[48] WANG, Z., GU, Q., NING, Y. and LIU, H. (2014). High dimensional expectation-maximization algorithm: Statistical optimization and asymptotic normality. Unpublished manuscript.
[49] WEI, G. and TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *J. Amer. Statist. Assoc.* **85** 699–704.
[50] WU, C.-F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11** 95–103. MR0684867
[51] XU, L. and JORDAN, M. I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Comput.* **8** 129–151.
[52] XU, Q. and YOU, J. (2007). Covariate selection for linear errors-in-variables regression models. *Comm. Statist. Theory Methods* **36** 375–386. MR2391878
[53] YI, X., CARAMANIS, C. and SANGHAVI, S. (2013). Alternating minimization for mixed linear regression. Unpublished manuscript.

The Annals of Statistics
2017, Vol. 45, No. 1, 121–157
DOI: 10.1214/16-AOS1437

NORMAL APPROXIMATION AND CONCENTRATION OF SPECTRAL PROJECTORS OF SAMPLE COVARIANCE

BY VLADIMIR KOLTCHINSKII[1] AND KARIM LOUNICI[2]

Georgia Institute of Technology

Let $X, X_1, \dots, X_n$ be i.i.d. Gaussian random variables in a separable Hilbert space $\mathbb{H}$ with zero mean and covariance operator $\Sigma = \mathbb{E}(X \otimes X)$, and let $\hat{\Sigma} := n^{-1} \sum_{j=1}^n (X_j \otimes X_j)$ be the sample (empirical) covariance operator based on $(X_1, \dots, X_n)$. Denote by P_r the spectral projector of Σ corresponding to its rth eigenvalue μ_r and by $\hat{P}_r$ the empirical counterpart of P_r. The main goal of the paper is to obtain tight bounds on

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2}{\mathrm{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)} \le x \right\} - \Phi(x) \right|,$$

where $\|\cdot\|_2$ denotes the Hilbert–Schmidt norm and Φ is the standard normal distribution function. Such accuracy of normal approximation of the distribution of squared Hilbert–Schmidt error is characterized in terms of so-called effective rank of Σ defined as $\mathbf{r}(\Sigma) = \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|_\infty}$, where $\mathrm{tr}(\Sigma)$ is the trace of Σ and $\|\Sigma\|_\infty$ is its operator norm, as well as another parameter characterizing the size of $\mathrm{Var}(\|\hat{P}_r - P_r\|_2^2)$. Other results include nonasymptotic bounds and asymptotic representations for the mean squared Hilbert–Schmidt norm error $\mathbb{E}\|\hat{P}_r - P_r\|_2^2$ and the variance $\mathrm{Var}(\|\hat{P}_r - P_r\|_2^2)$, and concentration inequalities for $\|\hat{P}_r - P_r\|_2^2$ around its expectation.

REFERENCES

[1] BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803

[2] BUNEA, F. and XIAO, L. (2015). On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli* **21** 1200–1230. MR3338661

[3] CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458

[4] 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

[5] DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277. MR0885466

[6] EICHELSBACHER, P. and THÄLE, C. (2014). New Berry–Esseen bounds for non-linear functionals of Poisson random measures. *Electron. J. Probab.* **19** no. 102, 25. MR3275854

MSC2010 subject classifications. 62H12.

Key words and phrases. Sample covariance, spectral projectors, effective rank, principal component analysis, concentration inequalities, normal approximation, perturbation theory.

[7] HAEUSLER, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *Ann. Probab.* **16** 275–299. MR0920271

[8] HALL, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096

[9] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[10] 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

[11] KATO, T. (1980). *Perturbation Theory for Linear Operators*. Springer, New York.

[12] KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. MR1946423

[13] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymtotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. H. Poincaré Probab. Statist.* **52** 1976–2013.

[14] KOLTCHINSKII, V. and LOUNICI, K. (2016). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*. To appear. Available at arXiv:1405.2468.

[15] KOLTCHINSKII, V. I. (1998). Asymptotics of spectral projections of some random matrices approximating integral operators. In *High Dimensional Probability* (*Oberwolfach*, 1996). *Progress in Probability* **43** 191–227. Birkhäuser, Basel. MR1652327

[16] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[17] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. Springer, Berlin. MR1102015

[18] LOUNICI, K. (2013). Sparse principal component analysis with missing observations. In *High Dimensional Probability VI. Progress in Probability* **66** 327–356. Birkhäuser, Basel. MR3443508

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

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

[21] VU, V. and LEI, J. (2012). Minimax rates of estimation for sparse PCA in high dimensions. *J. Mach. Learn. Res.* **22** 1278–1286.

The Annals of Statistics
2017, Vol. 45, No. 1, 158–195
DOI: 10.1214/16-AOS1448

A GENERAL THEORY OF HYPOTHESIS TESTS AND CONFIDENCE REGIONS FOR SPARSE HIGH DIMENSIONAL MODELS[1]

BY YANG NING AND HAN LIU

Cornell University and Princeton University

We consider the problem of uncertainty assessment for low dimensional components in high dimensional models. Specifically, we propose a novel decorrelated score function to handle the impact of high dimensional nuisance parameters. We consider both hypothesis tests and confidence regions for generic penalized M-estimators. Unlike most existing inferential methods which are tailored for individual models, our method provides a general framework for high dimensional inference and is applicable to a wide variety of applications. In particular, we apply this general framework to study five illustrative examples: linear regression, logistic regression, Poisson regression, Gaussian graphical model and additive hazards model. For hypothesis testing, we develop general theorems to characterize the limiting distributions of the decorrelated score test statistic under both null hypothesis and local alternatives. These results provide asymptotic guarantees on the type I errors and local powers. For confidence region construction, we show that the decorrelated score function can be used to construct point estimators that are asymptotically normal and semiparametrically efficient. We further generalize this framework to handle the settings of misspecified models. Thorough numerical results are provided to back up the developed theory.

REFERENCES

[1] BACH, F. (2010). Self-concordant analysis for logistic regression. *Electron. J. Stat.* **4** 384–414. MR2645490

[2] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root Lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806. MR2860324

[3] BELLONI, A., CHERNOZHUKOV, V. and WEI, Y. (2013). Honest confidence regions for logistic regression with a large number of controls. Preprint. Available at arXiv:1304.3969.

[4] BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434. MR0386168

[5] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

[6] CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

[7] COX, D. R. and HINKLEY, D. V. (1979). *Theoretical Statistics*. CRC Press, Boca Raton, FL.

[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

MSC2010 subject classifications. Primary 62E20, 62F03; secondary 62F25.

Key words and phrases. High dimensional inference, nuisance parameter, sparsity, hypothesis test, confidence interval, score function, model misspecification.

[9] FANG, E. X., NING, Y. and LIU, H. (2014). Testing and confidence intervals for high dimensional proportional hazards model. Preprint. Available at arXiv:1412.5158.
[10] GODAMBE, V. P. and KALE, B. K. (1991). Estimating functions: An overview. In *Estimating Functions*. *Oxford Statist. Sci. Ser.* **7** 3–20. Oxford Univ. Press, New York. MR1163993
[11] GODFREY, L. G. (1991). *Misspecification Tests in Econometrics*: *The Lagrange Multiplier Principle and Other Approaches*. *Econometric Society Monographs* **16**. Cambridge Univ. Press, Cambridge. Reprint of the 1988 original. MR1113260
[12] JANKOVA, J. and VAN DE GEER, S. (2013). Confidence intervals for high-dimensional inverse covariance estimation. Preprint. Available at arXiv:1403.6752.
[13] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152
[14] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inform. Theory* **60** 6522–6554. MR3265038
[15] KNIGHT, K. and FU, W. (2000). Asymptotics for Lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787
[16] LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2013). Exact inference after model selection via the Lasso. Preprint. Available at arXiv:1311.6238.
[17] LIN, D. Y. and YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81** 61–71. MR1279656
[18] LIN, W. and LV, J. (2013). High-dimensional sparse additive hazards regression. *J. Amer. Statist. Assoc.* **108** 247–264. MR3174617
[19] LINDSAY, B. (1982). Conditional score functions: Some optimality results. *Biometrika* **69** 503–512. MR0695197
[20] LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453
[21] LIU, W. and LUO, X. (2012). High-dimensional sparse precision matrix estimation via sparse column inverse operator. Preprint. Available at arXiv:1203.3896.
[22] LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the Lasso. *Ann. Statist.* **42** 413–468. MR3210970
[23] MARTINUSSEN, T. and SCHEIKE, T. H. (2009). Covariate selection for the semiparametric additive risk model. *Scand. J. Stat.* **36** 602–619. MR2572578
[24] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Ann. Statist.* **34** 1436–1462. MR2278363
[25] MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523
[26] MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584
[27] NING, Y. and LIU, H. (2016). Supplement to "A general theory of hypothesis tests and confidence regions for sparse high dimensional models." DOI:10.1214/16-AOS1448SUPP.
[28] PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309. MR0760690
[29] PORTNOY, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. MR0811499
[30] 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
[31] SMALL, C. G. and MCLEISH, D. L. (2011). *Hilbert Space Methods in Probability and Statistical Inference* **920**. John Wiley & Sons, New York.
[32] SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166

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

[34] 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

[35] VAN DER VAART, A. W. (2000). *Asymptotic Statistics* **3**. Cambridge Univ. Press, Cambridge.

[36] VOORMAN, A., SHOJAIE, A. and WITTEN, D. (2014). Inference in high dimensions with the penalized score test. Preprint. Available at arXiv:1401.2678.

[37] WANG, L., KIM, Y. and LI, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *Ann. Statist.* **41** 2505–2536. MR3127873

[38] WANG, Z., LIU, H. and ZHANG, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *Ann. Statist.* **42** 2164–2201. MR3269977

[39] WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. MR2543689

[40] WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163

[41] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

[42] 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

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

[44] ZHONG, P.-S., HU, T. and LI, J. (2015). Tests for coefficients in high-dimensional additive hazard models. *Scand. J. Stat.* **42** 649–664. MR3391684

The Annals of Statistics
2017, Vol. 45, No. 1, 196–222
DOI: 10.1214/16-AOS1449

A BAYESIAN APPROACH FOR ENVELOPE MODELS

BY KSHITIJ KHARE[1], SUBHADIP PAL AND ZHIHUA SU[2]

University of Florida

The envelope model is a new paradigm to address estimation and prediction in multivariate analysis. Using sufficient dimension reduction techniques, it has the potential to achieve substantial efficiency gains compared to standard models. This model was first introduced by [Statist. Sinica **20** (2010) 927–960] for multivariate linear regression, and has since been adapted to many other contexts. However, a Bayesian approach for analyzing envelope models has not yet been investigated in the literature. In this paper, we develop a comprehensive Bayesian framework for estimation and model selection in envelope models in the context of multivariate linear regression. Our framework has the following attractive features. First, we use the matrix Bingham distribution to construct a prior on the orthogonal basis matrix of the envelope subspace. This prior respects the manifold structure of the envelope model, and can directly incorporate prior information about the envelope subspace through the specification of hyperparamaters. This feature has potential applications in the broader Bayesian sufficient dimension reduction area. Second, sampling from the resulting posterior distribution can be achieved by using a block Gibbs sampler with standard associated conditionals. This in turn facilitates computationally efficient estimation and model selection. Third, unlike the current frequentist approach, our approach can accommodate situations where the sample size is smaller than the number of responses. Lastly, the Bayesian approach inherently offers comprehensive uncertainty characterization through the posterior distribution. We illustrate the utility of our approach on simulated and real datasets.

REFERENCES

[1] ASMUSSEN, S. and GLYNN, P. W. (2011). A new proof of convergence of MCMC via the ergodic theorem. *Statist. Probab. Lett.* **81** 1482–1485. MR2818658

[2] BINGHAM, C. (1974). An antipodally symmetric distribution on the sphere. *Ann. Statist.* **2** 1201–1225. MR0397988

[3] CONWAY, J. B. (1990). *A Course in Functional Analysis*, 2nd ed. *Graduate Texts in Mathematics* **96**. Springer, New York. MR1070713

[4] 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

[5] COOK, R. D., LI, B. and CHIAROMONTE, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica* **20** 927–960. MR2729839

[6] COOK, R. D. and SU, Z. (2013). Scaled envelopes: Scale-invariant and efficient estimation in multivariate linear regression. *Biometrika* **100** 939–954. MR3142342

MSC2010 subject classifications. Primary 62F15, 62F30; secondary 60J20, 62H12.

Key words and phrases. Sufficient dimension reduction, envelope model, Gibbs sampling, Stiefel manifold, matrix Bingham distribution.

[7] GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2027492

[8] HOFF, P. D. (2009). Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. *J. Comput. Graph. Statist.* **18** 438–456. MR2749840

[9] KHARE, K., PAL, S. and SU, Z. (2016). Supplement to "A Bayesian approach for envelope models." DOI:10.1214/16-AOS1449SUPP.

[10] MAO, K., LIANG, F. and MUKHERJEE, S. (2010). Supervised dimension reduction using Bayesian mixture modeling. In *International Conference on Artificial Intelligence and Statistics* 501–508.

[11] REICH, B. J., BONDELL, H. D. and LI, L. (2011). Sufficient dimension reduction via Bayesian mixture modeling. *Biometrics* **67** 886–895. MR2829263

[12] ROSSI, P. E., ALLENBY, G. M. and MCCULLOCH, R. (2005). *Bayesian Statistics and Marketing*. Wiley, Chichester. MR2193403

[13] SAMEH, A. H. and WISNIEWSKI, J. A. (1982). A trace minimization algorithm for the generalized eigenvalue problem. *SIAM J. Numer. Anal.* **19** 1243–1259. MR0679663

[14] SCHOTT, J. R. (2013). On the likelihood ratio test for envelope models in multivariate linear regression. *Biometrika* **100** 531–537. MR3068454

[15] SU, Z. and COOK, D. (2012). Inner envelopes: Efficient estimation in multivariate linear regression. *Biometrika* **99** 687–702. MR2966778

[16] SU, Z. and COOK, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98** 133–146. MR2804215

[17] SU, Z. and COOK, R. D. (2013). Estimation of multivariate means with heteroscedastic errors using envelope models. *Statist. Sinica* **23** 213–230. MR3076165

[18] TOKDAR, S. T., ZHU, Y. M. and GHOSH, J. K. (2010). Bayesian density regression with logistic Gaussian process and subspace projection. *Bayesian Anal.* **5** 319–344. MR2719655

The Annals of Statistics
2017, Vol. 45, No. 1, 223–256
DOI: 10.1214/16-AOS1450

MONGE–KANTOROVICH DEPTH, QUANTILES, RANKS AND SIGNS

BY VICTOR CHERNOZHUKOV[*,1], ALFRED GALICHON[†,2], MARC HALLIN[‡,3] AND MARC HENRY[§,4]

Massachusetts Institute of Technology[*], New York University[†], Université libre de Bruxelles[‡] and Pennsylvania State University[§]

We propose new concepts of statistical depth, multivariate quantiles, vector quantiles and ranks, ranks and signs, based on canonical transportation maps between a distribution of interest on $\mathbb{R}^d$ and a reference distribution on the d-dimensional unit ball. The new depth concept, called *Monge–Kantorovich depth*, specializes to halfspace depth for $d = 1$ and in the case of spherical distributions, but for more general distributions, differs from the latter in the ability for its contours to account for non-convex features of the distribution of interest. We propose empirical counterparts to the population versions of those Monge–Kantorovich depth contours, quantiles, ranks, signs and vector quantiles and ranks, and show their consistency by establishing a uniform convergence property for empirical (forward and reverse) transport maps, which is the main theoretical result of this paper.

REFERENCES

[1] AGOSTINELLI, C. and ROMANAZZI, M. (2011). Local depth. *J. Statist. Plann. Inference* **141** 817–830. MR2732952

[2] AURENHAMMER, F., HOFFMANN, F. and ARONOV, B. (1998). Minkowski-type theorems and mean-square clustering. *Algorithmica* **20** 61–76.

[3] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. MR1738163

[4] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809

[5] CARLIER, G., CHERNOZHUKOV, V. and GALICHON, A. (2016). Vector quantile regression: An optimal transport approach. *Ann. Statist.* **44** 1165–1192. MR3485957

[6] CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. MR1395753

[7] CHEN, Y., DANG, X., PENG, H. and BART, H. L. J. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Trans. Pattern Anal. Mach. Intell.* **31** 288–305.

[8] CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2016). Supplement to "Monge–Kantorovich depth, quantiles, ranks and signs." DOI:10.1214/16-AOS1450SUPP.

[9] CUESTA-ALBERTOS, J. A. and NIETO-REYES, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.* **52** 4979–4988. MR2526207

MSC2010 subject classifications. 62M15, 62G35.

Key words and phrases. Statistical depth, vector quantiles, vector ranks, multivariate signs, empirical transport maps, uniform convergence of empirical transport.

[10] DECURNINGE, A. (2014). Multivariate quantiles and multivariate L-moments. Preprint. Available at arXiv:1409.6013.
[11] DENEEN, L. and SHUTE, G. (1988). Polygonizations of point sets in the plane. *Discrete Comput. Geom.* **3** 77–87. MR0918181
[12] DOKSUM, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2** 267–277. MR0356350
[13] DOKSUM, K. A. and SIEVERS, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika* **63** 421–434. MR0443210
[14] DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Qualifying paper, Harvard Univ.
[15] DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. MR1193313
[16] DUTTA, S., GHOSH, A. K. and CHAUDHURI, P. (2011). Some intriguing properties of Tukey's halfspace depth. *Bernoulli* **17** 1420–1434. MR2854779
[17] EDELSBRUNNER, H., KIRKPATRICK, D. G. and SEIDEL, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory* **29** 551–559. MR0713690
[18] EKELAND, I., GALICHON, A. and HENRY, M. (2012). Comonotonic measures of multivariate risks. *Math. Finance* **22** 109–132. MR2881882
[19] GALICHON, A. and HENRY, M. (2012). Dual theory of choice with multivariate risks. *J. Econom. Theory* **147** 1501–1516. MR2942620
[20] GHOSH, A. K. and CHAUDHURI, P. (2005). On maximum depth and related classifiers. *Scand. J. Statist.* **32** 327–350. MR2188677
[21] GRÜNBAUM, B. (1994). Hamiltonian polygons and polyhedra. *Geombinatorics* **3** 83–89. MR1326479
[22] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York. MR0229351
[23] HALLIN, M. and PAINDAVEINE, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170
[24] HALLIN, M. and PAINDAVEINE, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32** 2642–2678. MR2153998
[25] HALLIN, M. and PAINDAVEINE, D. (2005). Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. *J. Multivariate Anal.* **93** 122–163. MR2119768
[26] 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
[27] HALLIN, M. and PAINDAVEINE, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.* **36** 1261–1298. MR2418657
[28] HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670
[29] HALLIN, M. and WERKER, B. J. M. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165. MR1963675
[30] HASSAIRI, A. and REGAIEG, O. (2008). On the Tukey depth of a continuous probability distribution. *Statist. Probab. Lett.* **78** 2308–2313. MR2462666
[31] HLUBINKA, D., KOTÍK, L. and VENCÁLEK, O. (2010). Weighted halfspace depth. *Kybernetika* (*Prague*) **46** 125–148. MR2666899
[32] JUDD, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, MA. MR1646331
[33] KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[34] KOLTCHINSKII, V. and DUDLEY, R. (1992). On spatial quantiles. Unpublished manuscript.
[35] KOSHEVOY, G. and MOSLER, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* **25** 1998–2017. MR1474078
[36] KOSHEVOY, G. A. (2002). The Tukey depth characterizes the atomic measure. *J. Multivariate Anal.* **83** 360–364. MR1945958
[37] LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400
[38] LIU, R. Y. (1992). Data depth and multivariate rank tests. In *L_1-Statistical Analysis and Related Methods* (*Neuchâtel*, 1992) (Y. Dodge, ed.) 279–294. North-Holland, Amsterdam. MR1214839
[39] LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. MR1724033
[40] LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260. MR1212489
[41] MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Natl. Acad. Sci., India* **12** 49–55.
[42] MCCANN, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323. MR1369395
[43] MOSLER, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach. Lecture Notes in Statistics* **165**. Springer, Berlin. MR1913862
[44] MÖTTÖNEN, J. and OJA, H. (1995). Multivariate sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213.
[45] OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332. MR0721446
[46] PAINDAVEINE, D. and ŠIMAN, M. (2012). Computing multiple-output regression quantile regions. *Comput. Statist. Data Anal.* **56** 840–853. MR2888729
[47] PAINDAVEINE, D. and VAN BEVER, G. (2013). From depth to local depth. *J. Amer. Statist. Assoc.* **108** 1105–1119.
[48] POLITIS, D., ROMANO, J. P. and WOLF, M. (1999). Weak convergence of dependent empirical measures with application to subsampling in function spaces. *J. Statist. Plann. Inference* **79** 179–190. MR1700758
[49] ROCKAFELLAR, R. T. (1997). *Convex Analysis. Princeton Landmarks in Mathematics*. Princeton Univ. Press, Princeton, NJ. MR1451876
[50] ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **317**. Springer, Berlin. MR1491362
[51] SERFLING, R. (2002). Quantile functions for multivariate analysis: Approaches and applications. *Stat. Neerl.* **56** 214–232. MR1916321
[52] SINGH, K. (1991). Majority depth. Unpublished manuscript.
[53] STAHEL, W. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen. Ph.D. thesis, Univ. Zürich.
[54] TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (*Vancouver, B. C.*, 1974), *Vol.* 2 523–531. Canad. Math. Congress, Montreal, Que. MR0426989
[55] VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. Mathematisch Centrum, Amsterdam.
[56] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence*. Springer, New York.
[57] VARDI, Y. and ZHANG, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426 (electronic). MR1740461

[58] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

[59] VILLANI, C. (2008). Stability of a 4th-order curvature condition arising in optimal transport theory. *J. Funct. Anal.* **255** 2683–2708. MR2473273

[60] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454

[61] ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. MR2012822

[62] ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005

The Annals of Statistics
2017, Vol. 45, No. 1, 257–288
DOI: 10.1214/16-AOS1452

IDENTIFYING THE NUMBER OF FACTORS FROM SINGULAR VALUES OF A LARGE SAMPLE AUTO-COVARIANCE MATRIX[1]

BY ZENG LI, QINWEN WANG AND JIANFENG YAO

University of Hong Kong

Identifying the number of factors in a high-dimensional factor model has attracted much attention in recent years and a general solution to the problem is still lacking. A promising ratio estimator based on singular values of lagged sample auto-covariance matrices has been recently proposed in the literature with a reasonably good performance under some specific assumption on the strength of the factors. Inspired by this ratio estimator and as a first main contribution, this paper proposes a complete theory of such sample singular values for both the factor part and the noise part under the large-dimensional scheme where the dimension and the sample size proportionally grow to infinity. In particular, we provide an exact description of the phase transition phenomenon that determines whether a factor is strong enough to be detected with the observed sample singular values. Based on these findings and as a second main contribution of the paper, we propose a new estimator of the number of factors which is strongly consistent for the detection of all significant factors (which are the only theoretically detectable ones). In particular, factors are assumed to have the minimum strength above the phase transition boundary which is of the order of a constant; they are thus not required to grow to infinity together with the dimension (as assumed in most of the existing papers on high-dimensional factor models). Empirical Monte-Carlo study as well as the analysis of stock returns data attest a very good performance of the proposed estimator. In all the tested cases, the new estimator largely outperforms the existing estimator using the same ratios of singular values.

REFERENCES

ALESSI, L., BARIGOZZI, M. and CAPASSO, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statist. Probab. Lett.* **80** 1806–1813. MR2734245

BAI, J. and LI, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. MR3014313

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

BAI, J. and NG, S. (2007). Determining the number of primitive shocks in factor models. *J. Bus. Econom. Statist.* **25** 52–60. MR2338870

BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053

MSC2010 subject classifications. Primary 62M10, 62H25; secondary 15B52.

Key words and phrases. High-dimensional factor model, high-dimensional time series, large sample auto-covariance matrices, spiked population model, number of factors, phase transition, random matrices.

BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. MR2835249

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

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2004). The generalized dynamic factor model consistency and rates. *J. Econometrics* **119** 231–255. MR2057100

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100** 830–840. MR2201012

GEWEKE, J. (1977). The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models* (D. J. Aigner and A. S. Goldberger, eds.). North-Holland, Amsterdam.

HALLIN, M. and LIŠKA, R. (2007). Determining the number of factors in the general dynamic factor model. *J. Amer. Statist. Assoc.* **102** 603–617. MR2325115

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663

LI, Z., PAN, G. and YAO, J. (2015). On singular value distribution of large-dimensional autocovariance matrices. *J. Multivariate Anal.* **137** 119–140. MR3332802

LI, Z., WANG, Q. W. and YAO, J. (2016). Supplement to "Identifying the number of factors from singular values of a large sample auto-covariance matrix." DOI:10.1214/16-AOS1452SUPP.

ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* **92** 1004–1016.

ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168** 244–258. MR2923766

ONATSKI, A. (2015). Asymptotic analysis of the squared estimation error in misspecified factor models. *J. Econometrics* **186** 388–406. MR3343793

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

SARGENT, T. J. and SIMS, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In *New Methods in Business Cycle Research*, *Vol.* 1 45–109. Federal Reserve Bank of Minneapolis, Minneapolis.

STOCK, J. H. and WATSON, M. W. (2011). Dynamic factor models. In *The Oxford Handbook of Economic Forecasting* 35–59. Oxford Univ. Press, Oxford. MR3204189

WANG, Q. W. and YAO, J. (2016). Moment approach for singular values distribution of a large auto-covariance matrix. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1641–1666. MR3573290

The Annals of Statistics
2017, Vol. 45, No. 1, 289–315
DOI: 10.1214/16-AOS1453

CONSISTENCY OF SPECTRAL HYPERGRAPH PARTITIONING UNDER PLANTED PARTITION MODEL

BY DEBARGHYA GHOSHDASTIDAR AND AMBEDKAR DUKKIPATI

Indian Institute of Science

Hypergraph partitioning lies at the heart of a number of problems in machine learning and network sciences. Many algorithms for hypergraph partitioning have been proposed that extend standard approaches for graph partitioning to the case of hypergraphs. However, theoretical aspects of such methods have seldom received attention in the literature as compared to the extensive studies on the guarantees of graph partitioning. For instance, consistency results of spectral graph partitioning under the stochastic block model are well known. In this paper, we present a planted partition model for sparse random nonuniform hypergraphs that generalizes the stochastic block model. We derive an error bound for a spectral hypergraph partitioning algorithm under this model using matrix concentration inequalities. To the best of our knowledge, this is the first consistency result related to partitioning nonuniform hypergraphs.

REFERENCES

ACHLIOPTAS, D. and COJA-OGHLAN, A. (2008). Algorithmic barriers from phase transitions. In *Proceedings of* 49*th Annual Symposium on Foundations of Computer Science*.

AGARWAL, S., BRANSON, K. and BELONGIE, S. (2006). Higher order learning with graphs. In *Proceedings of the International Conference on Machine Learning* (*Pittsburgh*, *Pennsylvania*, 2006) 17–24. ACM, New York.

ALON, N. and KAHALE, N. (1997). A spectral technique for coloring random 3-colorable graphs. *SIAM J. Comput.* **26** 1733–1748. MR1484153

ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. In *Proceedings of the Ninth Annual ACM–SIAM Symposium on Discrete Algorithms* (*San Francisco*, *CA*, 1998) 594–598. ACM, New York. MR1642973

ALPERT, C. J. (1998). The ISPD98 circuit benchmark suite. In *ISPD'*98: *Proceedings of the* 1998 *International Symposium on Physical Design* 80–85. ACM, New York.

AMINI, A. A. and LEVINA, E. (2014). On semi-definite relaxations for the block model. Available at arXiv:1406.5647.

ANDRITSOS, P., TSAPARAS, P., MILLER, R. J. and SEVCIK, K. C. (2004). LIMBO: Scalable clustering of categorical data. In *International Conference on Extending Database Technology* (*Heraklion*, *Crete*, *Greece*, 2004) 123–146. Springer, Berlin, Heidelberg.

BERGE, C. (1984). *Hypergraphs*: *Combinatorics of Finite Sets*. Elsevier, Amsterdam.

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

BOLLA, M. (1993). Spectra, Euclidean representations and clusterings of hypergraphs. *Discrete Math.* **117** 19–39. MR1226129

MSC2010 subject classifications. Primary 62H30, 62F12; secondary 05C65, 05C80.
Key words and phrases. Hypergraph, stochastic block model, spectral algorithm.

CHEN, H. and FRIEZE, A. (1996). Coloring bipartite hypergraphs. In *Integer Programming and Combinatorial Optimization* (*Vancouver*, *BC*, 1996). *Lecture Notes in Computer Science* **1084** 345–358. Springer, Berlin. MR1441812

CHEN, Y., SANGHAVI, S. and XU, H. (2014). Improved graph clustering. *IEEE Trans. Inform. Theory* **60** 6440–6455. MR3265033

CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253

CHUNG, F. R. K. (1993). The Laplacian of a hypergraph. In *Expanding Graphs* (*Princeton*, *NJ*, 1992). *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **10** 21–36. Amer. Math. Soc., Providence, RI. MR1235565

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

COOPER, J. and DUTLE, A. (2012). Spectra of uniform hypergraphs. *Linear Algebra Appl.* **436** 3268–3292. MR2900714

DARLING, R. W. R. and NORRIS, J. R. (2005). Structure of large random hypergraphs. *Ann. Appl. Probab.* **15** 125–152. MR2115039

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** 066106.

FIEDLER, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Math. J.* **23** (98) 298–305. MR0318007

FRIEDMAN, J., KAHN, J. and SZEMEREDI, E. (1989). On the second eigenvalue of random regular graphs. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*.

GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2015). Achieving optimal misclassification proportion in stochastic block model. Available at arXiv:1505.03772.

GHOSHAL, G., ZLATIĆ, V., CALDARELLI, G. and NEWMAN, M. E. J. (2009). Random hypergraphs and their applications. *Phys. Rev. E* **79** 066118. MR2551286

GHOSHDASTIDAR, D. and DUKKIPATI, A. (2014). Consistency of spectral partitioning of uniform hypergraphs under planted partition model. In *Advances in Neural Information Processing Systems* (*Montréal*, *Canada*, 2014) 397–405. Curran Associates, Inc., Red Hook, NY.

GHOSHDASTIDAR, D. and DUKKIPATI, A. (2015). A provable generalized tensor spectral method for uniform hypergraph partitioning. In *Proceedings of the International Conference on Machine Learning* (*Lille*, *France*, 2015). 400–409.

GHOSHDASTIDAR, D. and DUKKIPATI, A. (2016). Supplement to "Consistency of spectral hypergraph partitioning under planted partition model." DOI:10.1214/16-AOS1453SUPP.

GIBSON, D., KLEINBERG, J. and RAGHAVAN, P. (2000). Clustering categorical data: An approach based on dynamical systems. *VLDB J.* **8** 222–236.

GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826 (electronic). MR1908073

GOVINDU, V. M. (2005). A tensor decomposition for geometric grouping and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* (*San Diego*, *CA*, *USA*, 2005) 1150–1157. IEEE Computer Society, Washington, DC.

GUIMERA, R. and AMARAL, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* **433** 895–900.

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088

HU, S. and QI, L. (2012). Algebraic connectivity of an even uniform hypergraph. *J. Comb. Optim.* **24** 564–579. MR2974300

KARYPIS, G. and KUMAR, V. (2000). Multilevel k-way hypergraph partitioning. *VLSI Des.* **11** 285–300.

KERNIGHAN, B. W. and LIN, S. (1970). An efficient heuristic procedure for partitioning graphs. *Bell Syst. Tech. J.* **49** 291–307.

KRZAKALA, F., MOORE, C., MOSSEL, E., NEEMAN, J., SLY, A., ZDEBOROVÁ, L. and ZHANG, P. (2013). Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA* **110** 20935–20940. MR3174850

KUMAR, A., SABHARWAL, Y. and SEN, S. (2004). A simple linear time $(1+\varepsilon)$-approximation algorithm for geometric k-means clustering in any dimensions. In *Proceedings of the Annual Symposium on Foundations of Computer Science* (*Rome*, *Italy*, 2004) 454–462. IEEE Computer Society, Washington, DC.

LE, C. M., LEVINA, E. and VERSHYNIN, R. (2015). Sparse random graphs: Regularization and concentration of the Laplacian. Available at arXiv:1502.03049.

LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605

LEI, J. and ZHU, L. (2014). A generic sample splitting approach for refined community recovery in stochastic block models. Available at arXiv:1411.1469.

LICHMAN, M. (2013). UCI machine learning repository. Available at http://archive.ics.uci.edu/ml.

LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inform. Theory* **28** 129–137. MR0651807

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

MICHOEL, T. and NACHTERGAELE, B. (2012). Alignment and integration of complex networks by hypergraph-based spectral clustering. *Phys. Rev. E* **86** 056111.

NG, A., JORDAN, M. and WEISS, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems* (*Vancouver*, *British Columbia*, *Canada*) 849–856. MIT Press, Cambridge, MA.

OSTROVSKY, R., RABANI, Y., SCHULMAN, L. J. and SWAMY, C. (2012). The effectiveness of Lloyd-type methods for the k-means problem. *J. ACM* **59** Art. 28. MR3008400

PANAGIOTOU, K. and COJA-OGHLAN, A. (2012). Catching the k-NAESAT threshold. In *ACM Symposium on Theory of Computing*.

RODRÍGUEZ, J. A. (2002). On the Laplacian eigenvalues and metric parameters of hypergraphs. *Linear Multilinear Algebra* **50** 1–14. MR1890984

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856

SCHMIDT-PRUZAN, J. and SHAMIR, E. (1985). Component structure in the evolution of random hypergraphs. *Combinatorica* **5** 81–94. MR0803242

SCHWEIKERT, G. and KERNIGHAN, B. W. (1979). A proper model for the partitioning of electrical circuits. In *Proceedings of the* 9*th Design Automation Workshop* 57–62. ACM, New York, NY.

STASI, D., SADEGHI, K., RINALDO, A., PETROVIC, S. and FIENBERG, S. (2014). β models for random hypergraphs with a given degree sequence. In *Proceedings of COMPSTAT* 2014—21*st International Conference on Computational Statistics* 593–600. Internat. Statist. Inst., The Hague. MR3372442

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

TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459

VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416. MR2409803

VU, V. (2014). A simple SVD algorithm for finding hidden partitions. Available at arXiv:1404.3918.

WASSERMAN, S. (1994). *Social Network Analysis*: *Methods and Applications*. Cambridge Univ. Press, Cambridge.

ZHANG, Y., LEVINA, E. and ZHU, J. (2014). Detecting overlapping communities in networks with spectral methods. Available at arXiv:1412.3432v2.

ZHOU, D., HUANG, J. and SCHÖLKOPF, B. (2007). Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems* (*Vancouver, British Columbia*, *Canada*) 1601–1608. MIT Press, Cambridge, MA.

The Annals of Statistics
2017, Vol. 45, No. 1, 316–354
DOI: 10.1214/16-AOS1454

ORACLE INEQUALITIES FOR NETWORK MODELS AND SPARSE GRAPHON ESTIMATION

BY OLGA KLOPP[*,1], ALEXANDRE B. TSYBAKOV[†,2] AND NICOLAS VERZELEN[‡,3]

Université Paris Ouest[*], ENSAE, UMR CNRS 9194[†] and INRA, UMR 729, MISTEA[‡]

Inhomogeneous random graph models encompass many network models such as stochastic block models and latent position models. We consider the problem of statistical estimation of the matrix of connection probabilities based on the observations of the adjacency matrix of the network. Taking the stochastic block model as an approximation, we construct estimators of network connection probabilities—the ordinary block constant least squares estimator, and its restricted version. We show that they satisfy oracle inequalities with respect to the block constant oracle. As a consequence, we derive optimal rates of estimation of the probability matrix. Our results cover the important setting of sparse networks. Another consequence consists in establishing upper bounds on the minimax risks for graphon estimation in the L_2 norm when the probability matrix is sampled according to a graphon model. These bounds include an additional term accounting for the "agnostic" error induced by the variability of the latent unobserved variables of the graphon model. In this setting, the optimal rates are influenced not only by the bias and variance components as in usual nonparametric problems but also include the third component, which is the agnostic error. The results shed light on the differences between estimation under the empirical loss (the probability matrix estimation) and under the integrated loss (the graphon estimation).

REFERENCES

[1] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII*—1983. *Lecture Notes in Math.* **1117** 1–198. Springer, Berlin. MR0883646

[2] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

[3] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868

[4] BORGS, C., CHAYES, J. and SMITH, A. (2015). Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems* 1369–1377. Nips Foundation. Available at http://books.nips.cc.

[5] CHAN, S. H. and AIROLDI, E. M. (2014). A consistent histogram estimator for exchangeable graph models. In *Proceedings of the* 31*st International Conference on Machine Learning* 208–216. Omnipress, Madison, WI.

MSC2010 subject classifications. Primary 62G05; secondary 60C05.

Key words and phrases. Inhomogeneous random graph, networks, oracle inequality, sparse graphon, sparsity, stochastic block model.

[6] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604

[7] CHOI, D. (2015). Co-clustering of nonsmooth graphons. Preprint. Available at arXiv:1507.06352.

[8] CHVÁTAL, V. (1983). *Linear Programming*. W. H. Freeman and Company, New York. MR0717219

[9] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

[10] GAO, C., LU, Y. and ZHOU, H. H. (2014). Rate-optimal graphon estimation. Preprint. Available at arXiv:1410.5837.

[11] LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035

[12] LOVASZ, L. and SZEGEDY, B. (2004). Limits of dense graph sequences. ArXiv Mathematics E-prints.

[13] PENROSE, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198

[14] TANG, M., SUSSMAN, D. L. and PRIEBE, C. E. (2013). Universally consistent vertex classification for latent positions graphs. *Ann. Statist.* **41** 1406–1430. MR3113816

[15] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. Revised and extended from the 2004 French original, translated by Vladimir Zaiats. MR2724359

[16] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. Preprint. Available at arXiv:1309.5936.

[17] XU, J., MASSOULIÉ, L. and LELARGE, M. (2014). Edge label inference in generalized stochastic block models: From spectral theory to impossibility results. Preprint. Available at arXiv:1406.6897.

The Annals of Statistics
2017, Vol. 45, No. 1, 355–385
DOI: 10.1214/16-AOS1455

APPROXIMATE GROUP CONTEXT TREE

BY ALEXANDRE BELLONI AND ROBERTO I. OLIVEIRA[1]

The Fuqua School of Business
Duke University and IMPA

We study a variable length Markov chain model associated with a group of stationary processes that share the same context tree but each process has potentially different conditional probabilities. We propose a new model selection and estimation method which is computationally efficient. We develop oracle and adaptivity inequalities, as well as model selection properties, that hold under continuity of the transition probabilities and polynomial β-mixing. In particular, model misspecification is allowed.

These results are applied to interesting families of processes. For Markov processes, we obtain uniform rate of convergence for the estimation error of transition probabilities as well as perfect model selection results. For chains of infinite order with complete connections, we obtain explicit uniform rates of convergence on the estimation of conditional probabilities, which have an explicit dependence on the processes' continuity rates. Similar guarantees are also derived for renewal processes.

Our results are shown to be applicable to discrete stochastic dynamic programming problems and to dynamic discrete choice models. We also apply our estimator to a linguistic study, based on recent work by Galves et al. [*Ann. Appl. Stat.* **6** (2012) 186–209], of the rhythmic differences between Brazilian and European Portuguese.

REFERENCES

[1] AGUIRREGABIRIA, V. and MIRA, P. (2010). Dynamic discrete choice structural models: A survey. *J. Econometrics* **156** 38–67. MR2609919

[2] ARELLANO, M. and HONORÉ, B. H. (2001). Panel data models: Some recent developments. *Handb. Econom.* **5** 3229–3296.

[3] BEJERANO, G. (2004). Algorithms for variable length Markov chain modeling. *Bioinformatics* **20** 788–789.

[4] BELLONI, A. and OLIVEIRA, R. I. (2016). Supplement to "Approximate group context tree." DOI:10.1214/16-AOS1455SUPP.

[5] BERTSEKAS, D. P. (1987). *Dynamic Programming*: *Deterministic and Stochastic Models*. Prentice Hall, Englewood Cliffs, NJ. MR0896902

[6] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

[7] BROWNING, M. and CARRO, J. M. (2010). Heterogeneity in dynamic discrete choice models. *Econom. J.* **13** 1–39. MR2656540

MSC2010 subject classifications. Primary 62M05, 62M09, 62G05; secondary 62P20, 60J10.

Key words and phrases. Categorical time series, group context tree, dynamic discrete choice models, dynamic programming, model selection, VLMC.

[8] BROWNING, M. and CARRO, J. M. (2014). Dynamic binary outcome models with maximal heterogeneity. *J. Econometrics* **178** 805–823. MR3144684

[9] BÜHLMANN, P. (1999). Efficient and adaptive post-model-selection estimators. *J. Statist. Plann. Inference* **79** 1–9. MR1704215

[10] BÜHLMANN, P. (2000). Model selection for variable length Markov chains and tuning the context algorithm. *Ann. Inst. Statist. Math.* **52** 287–315. MR1763564

[11] BÜHLMANN, P. and WYNER, A. J. (1999). Variable length Markov chains. *Ann. Statist.* **27** 480–513. MR1714720

[12] CHERNOZHUKOV, V., FERNANDEZ-VAL, I., HAHN, J. and NEWEY, W. (2009). Identification and estimation of marginal effects in nonlinear panel models. Available at arXiv:0904.1990.

[13] CSISZÁR, I. and SHIELDS, P. C. (1996). Redundancy rates for renewal and other processes. *IEEE Trans. Inform. Theory* **42** 2065–2072.

[14] CSISZÁR, I. and TALATA, Z. (2006). Context tree estimation for not necessarily finite memory processes, via BIC and MDL. *IEEE Trans. Inform. Theory* **52** 1007–1016. MR2238067

[15] FARIAS, V. F., MOALLEMI, C. C., VAN ROY, B. and WEISSMAN, T. (2010). Universal reinforcement learning. *IEEE Trans. Inform. Theory* **56** 2441–2454. MR2729794

[16] FERRARI, F. and WYNER, A. (2003). Estimation of general stationary processes by variable length Markov chains. *Scand. J. Statist.* **30** 459–480. MR2002222

[17] GALVES, A., GALVES, C., GARCÍA, J. E., GARCIA, N. L. and LEONARDI, F. (2012). Context tree selection and linguistic rhythm retrieval from written texts. *Ann. Appl. Stat.* **6** 186–209. MR2951534

[18] GARIVIER, A. (2006). Redundancy of the context-tree weighting method on renewal and Markov renewal processes. *IEEE Trans. Inform. Theory* **52** 5579–5586. MR2300720

[19] GARIVIER, A. and LEONARDI, F. (2011). Context tree selection: A unifying view. *Stochastic Process. Appl.* **121** 2488–2506. MR2832411

[20] LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.* **35** 459–470. MR1091202

[21] LOUNICI, K., PONTIL, M., TSYBAKOV, A. B. and VAN DE GEER, S. (2010). Taking advantage of sparsity in multi-task learning. In *Proc. Computational Learning Theory Conference* (*COLT* 2009).

[22] OBOZINSKI, G., WAINWRIGHT, M. J. and JORDAN, M. I. (2011). Support union recovery in high-dimensional multivariate regression. *Ann. Statist.* **39** 1–47. MR2797839

[23] PUTERMAN, M. L. (1994). *Markov Decision Processes*: *Discrete Stochastic Dynamic Programming*. Wiley, New York. MR1270015

[24] RISSANEN, J. (1983). A universal data compression system. *IEEE Trans. Inform. Theory* **29** 656–664. MR0730903

[25] ROSS, S. (1983). *Introduction to Stochastic Dynamic Programming*. Academic Press, New York. MR0749232

[26] TALATA, Z. and DUNCAN, T. (2009). Unrestricted bic context tree estimation for not necessarily finite memory processes. In 2009 *IEEE International Symposium on Information Theory* 724–728.

[27] VERT, J.-P. (2001). Adaptive context trees and text clustering. *IEEE Trans. Inform. Theory* **47** 1884–1901. MR1842525

[28] WILLEMS, F. M. J., SHTARKOV, Y. M. and TJALKENS, T. J. (1995). The context-tree weighting method: Basic properties. *IEEE Trans. Inform. Theory* **41** 653–664.

[29] 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

The Annals of Statistics
2017, Vol. 45, No. 1, 386–414
DOI: 10.1214/16-AOS1456

FLEXIBLE RESULTS FOR QUADRATIC FORMS WITH APPLICATIONS TO VARIANCE COMPONENTS ESTIMATION

BY LEE H. DICKER[1] AND MURAT A. ERDOGDU

Rutgers University and Stanford University

We derive convenient uniform concentration bounds and finite sample multivariate normal approximation results for quadratic forms, then describe some applications involving variance components estimation in linear random-effects models. Random-effects models and variance components estimation are classical topics in statistics, with a corresponding well-established asymptotic theory. However, our finite sample results for quadratic forms provide additional flexibility for easily analyzing random-effects models in nonstandard settings, which are becoming more important in modern applications (e.g., genomics). For instance, in addition to deriving novel non-asymptotic bounds for variance components estimators in classical linear random-effects models, we provide a concentration bound for variance components estimators in linear models with correlated random-effects and discuss an application involving sparse random-effects models. Our general concentration bound is a uniform version of the Hanson–Wright inequality. The main normal approximation result in the paper is derived using Reinert and Röllin [*Ann. Probab.* (2009) **37** 2150–2173] embedding technique for Stein's method of exchangeable pairs.

REFERENCES

ADAMCZAK, R. (2015). A note on the Hanson–Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.* **20** no. 72, 13. MR3407216

BAI, Z. D., MIAO, B. and YAO, J.-F. (2003). Convergence rates of spectral distributions of large sample covariance matrices. *SIAM J. Matrix Anal. Appl.* **25** 105–127. MR2002902

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

CHATTERJEE, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610. MR2435859

DE LOS CAMPOS, G., SORENSEN, D. and GIANOLA, D. (2015). Genomic heritability: What is it? *PLoS Genet.* **11** e1005048.

DEMIDENKO, E. (2013). *Mixed Models*: *Theory and Applications with R*, 2nd ed. Wiley, Hoboken, NJ. MR3235905

DICKER, L. H. and ERDOGDU, M. A. (2016a). Maximum likelihood for variance estimation in high-dimensional linear models. In *Proceedings of the* 19*th International Conference on Artificial Intelligence and Statistics* 159–167. JMLR Workshop & Conference Proceedings.

DICKER, L. H. and ERDOGDU, M. A. (2016b). Supplement to "Flexible results for quadratic forms with applications to variance components estimation." DOI:10.1214/16-AOS1456SUPP.

MSC2010 subject classifications. Primary 62F99; secondary 62E17, 62F12.

Key words and phrases. Hanson–Wright inequality, random-effects models, model misspecification, Stein's method, uniform concentration bounds.

GOLAN, D. and ROSSET, S. (2011). Accurate estimation of heritability in genome wide studies using random effects models. *Bioinformatics* **27** i317–i323.

GÖTZE, F. and TIKHOMIROV, A. N. (1999). Asymptotic distribution of quadratic forms. *Ann. Probab.* **27** 1072–1098. MR1699003

GÖTZE, F. and TIKHOMIROV, A. (2002). Asymptotic distribution of quadratic forms and applications. *J. Theoret. Probab.* **15** 423–475. MR1898815

HALL, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096

HANSON, D. L. and WRIGHT, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083. MR0279864

HARTLEY, H. O. and RAO, J. N. K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54** 93–108. MR0216684

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–340. MR0451550

HENDERSON, C. R. (1950). Abstract: Estimation of genetic parameters. *Ann. Math. Stat.* **21** 309–310.

HSU, D., KAKADE, S. M. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17** no. 52, 6. MR2994877

JIANG, J. (1996). REML estimation: Asymptotic behavior and related topics. *Ann. Statist.* **24** 255–286. MR1389890

JIANG, J. (1998). Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Statist. Sinica* **8** 861–885. MR1651513

JIANG, J., LI, C., PAUL, D., YANG, C. and ZHAO, H. (2016). On high-dimensional genome-wide association study and misspecified mixed model analysis. *Ann. Statist.* **44** 2127–2160. MR3546446

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York. MR1639875

REINERT, G. and RÖLLIN, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. MR2573554

RICHARDSON, A. M. and WELSH, A. H. (1994). Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Aust. J. Stat.* **36** 31–43. MR1309503

RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. MR3125258

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York. MR1190470

SEVASTYANOV, B. A. (1961). A class of limit distributions for quadratic forms of normal stochastic variables. *Theory Probab. Appl.* **6** 337–340.

SPEED, D., HEMANI, G., JOHNSON, M. R. and BALDING, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91** 1011–1021.

VAN DE GEER, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

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

WHITTLE, P. (1964). On the convergence to normality of quadratic forms in independent variables. *Teor. Veroyatn. Primen.* **9** 113–118. MR0161429

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. and VISSCHER, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42** 565–569.

YANG, J., ZAITLEN, N. A., GODDARD, M. E., VISSCHER, P. M. and PRICE, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46** 100–106.

ZAITLEN, N. A. and KRAFT, P. (2012). Heritability in the genome-wide association era. *Hum. Genet.* **131** 1655–1664.

The Annals of Statistics
2017, Vol. 45, No. 1, 415–460
DOI: 10.1214/16-AOS1463

EXTREME EIGENVALUES OF LARGE-DIMENSIONAL SPIKED FISHER MATRICES WITH APPLICATION

BY QINWEN WANG AND JIANFENG YAO

University of Hong Kong

Consider two p-variate populations, not necessarily Gaussian, with covariance matrices Σ_1 and Σ_2, respectively. Let S_1 and S_2 be the corresponding sample covariance matrices with degrees of freedom m and n. When the difference Δ between Σ_1 and Σ_2 is of small rank compared to p, m and n, the Fisher matrix $S := S_2^{-1}S_1$ is called a *spiked Fisher matrix*. When p, m and n grow to infinity proportionally, we establish a phase transition for the extreme eigenvalues of the Fisher matrix: a displacement formula showing that when the eigenvalues of Δ (*spikes*) are above (or under) a critical value, the associated extreme eigenvalues of S will converge to some point outside the support of the global limit (LSD) of other eigenvalues (become outliers); otherwise, they will converge to the edge points of the LSD. Furthermore, we derive central limit theorems for those outlier eigenvalues of S. The limiting distributions are found to be Gaussian if and only if the corresponding population spike eigenvalues in Δ are *simple*. Two applications are introduced. The first application uses the largest eigenvalue of the Fisher matrix to test the equality between two high-dimensional covariance matrices, and explicit power function is found under the spiked alternative. The second application is in the field of signal detection, where an estimator for the number of signals is proposed while the covariance structure of the noise is arbitrary.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York. MR0771294

BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053

BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686

BAI, Z. D., YIN, Y. Q. and KRISHNAIAH, P. R. (1987). On limiting empirical distribution function of the eigenvalues of a multivariate F matrix. *Theory Probab. Appl.* **32** 490–500.

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

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

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

MSC2010 subject classifications. Primary 62H12; secondary 60F05.

Key words and phrases. Large-dimensional Fisher matrices, spiked Fisher matrix, spiked population model, extreme eigenvalue, phase transition, central limit theorem, signal detection, high-dimensional data analysis.

BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. MR2835249

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

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618

CAPITAINE, M. (2013). Additive/multiplicative free subordination property and limiting eigenvectors of spiked additive deformations of Wigner matrices and spiked sample covariance matrices. *J. Theory Probab.* **26** 595–648. MR3090543

CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158

DHARMAWANSA, P., JOHNSTONE, I. M. and ONATSKI, A. (2014). Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios. Preprint. Available at arXiv:1411.3875.

FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807

HAN, X., PAN, G. and ZHANG, B. (2016). The Tracy–Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44** 1564–1592. MR3519933

HU, J. and BAI, Z. (2014). Strong representation of weak convergence. *Sci. China Math.* **57** 2399–2406. MR3266500

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

KARGIN, V. (2015). On estimation in the reduced-rank regression with a large number of responses and predictors. *J. Multivariate Anal.* **140** 377–394. MR3372575

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.

LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

NADLER, B. (2010). Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator. *IEEE Trans. Signal Process.* **58** 2746–2756. MR2789420

ONATSKI, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070

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

PASSEMIER, D. and YAO, J. (2014). Estimation of the number of spikes, possibly equal, in the high-dimensional case. *J. Multivariate Anal.* **127** 173–183. MR3188885

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

PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787

PIZZO, A., RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 64–94. MR3060148

RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices II: Delocalized perturbations. *Random Matrices Theory Appl.* **2** 1250015, 36. MR3039820

SHI, D. (2013). Asymptotic joint distribution of extreme sample eigenvalues and eigenvectors in the spiked population model. Preprint. Available at arXiv:1304.6113.

SILVERSTEIN, J. W. (1985). The limiting eigenvalue distribution of a multivariate F matrix. *SIAM J. Math. Anal.* **16** 641–646. MR0783987

SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **1** 261–290.

WACHTER, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.* **8** 937–957. MR0585695

WANG, Q., SU, Z. and YAO, J. (2014). Joint CLT for several random sesquilinear forms with applications to large-dimensional spiked population models. *Electron. J. Probab.* **19** 1–28. MR3275855

ZHENG, S. R., BAI, Z. D. and YAO, J. F. (2013). CLT for linear spectral statistics of random matrix $S^{-1}T$. Preprint. Available at arXiv:1305.1376.