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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 46, No. 2—April 2018

THE ANNALS OF STATISTICS Vol. 46, No. 2, pp. 457–923 April 2018

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 46, Number 2, April 2018. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

The Annals of Statistics
2018, Vol. 46, No. 2, 457–468
https://doi.org/10.1214/17-AOS1555

STRONG ORTHOGONAL ARRAYS OF STRENGTH TWO PLUS

BY YUANZHEN HE, CHING-SHUI CHENG AND BOXIN TANG[1]

Beijing Normal University, Academia Sinica and Simon Fraser University

Strong orthogonal arrays were recently introduced and studied in He and Tang [*Biometrika* **100** (2013) 254–260] as a class of space-filling designs for computer experiments. To enjoy the benefits of better space-filling properties, when compared to ordinary orthogonal arrays, strong orthogonal arrays need to have strength three or higher, which may require run sizes that are too large for experimenters to afford. To address this problem, we introduce a new class of arrays, called strong orthogonal arrays of strength two plus. These arrays, while being more economical than strong orthogonal arrays of strength three, still enjoy the better two-dimensional space-filling property of the latter. Among the many results we have obtained on the characterizations and constructions of strong orthogonal arrays of strength two plus, worth special mention is their intimate connection with second-order saturated designs.

REFERENCES

BLOCK, R. M. and MEE, R. W. (2003). Secord order saturated resolution IV designs. *J. Stat. Theory Appl.* **2** 96–112. MR2040435

CHEN, J., SUN, D. X. and WU, C. F. J. (1993). A catalogue of two-level and three-level fractional factorial designs with small runs. *Int. Stat. Rev.* **61** 131–145.

CHENG, C. S. (2014). *Theory of Factorial Design: Single- and Multi-Stratum Experiments*. CRC Press, Boca Raton.

DEY, A. and MUKERJEE, R. (1999). *Fractional Factorial Plans*. Wiley, New York. MR1679441

FANG, K.-T., LI, R. and SUDJIANTO, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, Boca Raton. MR2223960

GABIDULIN, E. M., DAVYDOV, A. A. and TOMBAK, L. M. (1991). Linear codes with covering radius 2 and other new covering codes. *IEEE Trans. Inform. Theory* **37** 219–224.

GEORGIOU, S. D. and EFTHIMIOU, I. (2014). Some classes of orthogonal Latin hypercube designs. *Statist. Sinica* **24** 101–120.

HE, Y. and TANG, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100** 254–260. MR3034340

HE, Y. and TANG, B. (2014). A characterization of strong orthogonal arrays of strength three. *Ann. Statist.* **42** 1347–1360. MR3226159

HEDAYAT, A. S., SLOANE, N. J. A. and STUFKEN, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York.

MCKAY, M. D., BECKMAN, R. J. and CONOVER, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** 239–245. MR0533252

MSC2010 subject classifications. Primary 62K15; secondary 05B15.

Key words and phrases. Complementary design, computer experiment, Latin hypercube, second-order saturated design, space-filling design.

NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia.

OWEN, A. B. (1992). Orthogonal arrays for computer experiments, integration, and visualization. *Statist*. *Sinica* **2** 439–452.

SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York. MR2160708

SUN, F., LIU, M.-Q. and LIN, D. K. J. (2009). Construction of orthogonal Latin hypercube designs. *Biometrika* **96** 971–974. MR2564504

TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc*. **88** 1392–1397. MR1245375

XU, H., PHOA, F. K. H. and WONG, W. K. (2009). Recent developments in nonregular fractional factorial designs. *Stat. Surv.* **3** 18–46. MR2520978

The Annals of Statistics
2018, Vol. 46, No. 2, 469–499
https://doi.org/10.1214/17-AOS1556

STATISTICAL INFERENCE FOR SPATIAL STATISTICS DEFINED IN THE FOURIER DOMAIN

BY SUHASINI SUBBA RAO[1]

Texas A&M University

A class of Fourier based statistics for irregular spaced spatial data is introduced. Examples include the Whittle likelihood, a parametric estimator of the covariance function based on the L_2-contrast function and a simple nonparametric estimator of the spatial autocovariance which is a nonnegative function. The Fourier based statistic is a quadratic form of a discrete Fourier-type transform of the spatial data. Evaluation of the statistic is computationally tractable, requiring $O(nb)$ operations, where b are the number of Fourier frequencies used in the definition of the statistic and n is the sample size. The asymptotic sampling properties of the statistic are derived using both increasing domain and fixed-domain spatial asymptotics. These results are used to construct a statistic which is asymptotically pivotal.

REFERENCES

[1] BANDYOPADHYAY, S. and LAHIRI, S. N. (2009). Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhyā* **71** 221–259. MR2639292

[2] BANDYOPADHYAY, S., LAHIRI, S. N. and NORDMAN, D. J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *Ann. Statist.* **43** 519–545. MR3316189

[3] BANDYOPADHYAY, S. and SUBBA RAO, S. (2017). A test for stationarity for irregularly spaced spatial data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 95–123. MR3597966

[4] BEUTLER, F. J. (1970). Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inform. Theory* **16** 147–152. MR0273761

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

[6] BRILLINGER, D. (1969). The calculation of cumulants via conditioning. *Ann. Inst. Statist. Math.* **21** 215–218.

[7] BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*, 2nd ed. Holden-Day, Inc., Oakland, CA. MR0595684

[8] CAN, S. U., MIKOSCH, T. and SAMORODNITSKY, G. (2010). Weak convergence of the function-indexed integrated periodogram for infinite variance processes. *Bernoulli* **16** 995–1015. MR2759166

[9] CHEN, W. W., HURVICH, C. M. and LU, Y. (2006). On the correlation matrix of the discrete Fourier transform and the fast solution of large Toeplitz systems for long-memory time series. *J. Amer. Statist. Assoc.* **101** 812–822. MR2281249

[10] CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494

MSC2010 subject classifications. Primary 62M30; secondary 62M15.

Key words and phrases. Fixed and increasing domain asymptotics, irregular spaced locations, quadratic forms, spatial spectral density function, stationary spatial random fields.

[11] CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. Revised reprint of the 1991 edition. MR1239641

[12] DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766. MR1026311

[13] DAHLHAUS, R. and JANAS, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* **24** 1934–1963. MR1421155

[14] DEO, R. S. and CHEN, W. W. (2000). On the integral of the squared periodogram. *Stochastic Process. Appl.* **85** 159–176. MR1730613

[15] DUNSMUIR, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Ann. Statist.* **7** 490–506. MR0527485

[16] FOX, R. and TAQQU, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields* **74** 213–240. MR0871252

[17] FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.* **102** 321–331. MR2345545

[18] GINÉ, E. and NICKL, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** 47–61. MR2401653

[19] GIRAITIS, L. and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104. MR1061950

[20] HALL, P., FISHER, N. I. and HOFFMANN, B. (1994). On the nonparametric estimation of covariance functions. *Ann. Statist.* **22** 2115–2134. MR1329185

[21] HALL, P. and PATIL, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Related Fields* **99** 399–424. MR1283119

[22] HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probab.* **8** 767–780. MR0295498

[23] KAWATA, T. (1959). Some convergence theorems for stationary stochastic processes. *Ann. Math. Stat.* **30** 1192–1214. MR0109365

[24] LAHIRI, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā* **65** 356–388. MR2028905

[25] LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981

[26] MASRY, E. (1978). Poisson sampling and spectral estimation of continuous-time processes. *IEEE Trans. Inform. Theory* **24** 173–183. MR0483238

[27] MATSUDA, Y. and YAJIMA, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 191–217. MR2655530

[28] NEIBUHR, T. and KREISS, J.-P. (2014). Asymptotics for autocovariances and integrated periodograms for linear processes observed at lower frequencies. *Int. Stat. Rev.* **82** 123–140. MR3200539

[29] PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.* **28** 329–348. MR0088833

[30] PECCATI, G. and TAQQU, M. S. (2011). *Wiener Chaos: Moments, Cumulants and Diagrams: A Survey with Computer Implementation. Bocconi & Springer Series* **1**. Springer, Milan; Bocconi Univ. Press, Milan. MR2791919

[31] RICE, J. (1979). On the estimation of the parameters of a power spectrum. *J. Multivariate Anal.* **9** 378–392. MR0548788

[32] SHAPIRO, H. S. and SILVERMAN, R. A. (1960). Alias-free sampling of random noise. *J. Soc. Ind. Appl. Math.* **8** 225–248. MR0121948

[33] STEIN, M. L. (1995). Fixed-domain asymptotics for spatial periodograms. *J. Amer. Statist. Assoc.* **90** 1277–1288. MR1379470

[34] STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

[35] STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 275–296. MR2062376

[36] SUBBA RAO, S. (2018). Supplement to "Statistical inference for spatial statistics defined in the Fourier domain." DOI:10.1214/17-AOS1556SUPP.

[37] SUBBA RAO, S. (2017). Orthogonal samples for estimators in time series. Under revision.

[38] VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. R. Stat. Soc.*, *B* **50** 297–312. MR0964183

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

[40] ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303

[41] ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195

The Annals of Statistics
2018, Vol. 46, No. 2, 500–525
https://doi.org/10.1214/17-AOS1558

ON THE INFERENCE ABOUT THE SPECTRAL DISTRIBUTION OF HIGH-DIMENSIONAL COVARIANCE MATRIX BASED ON HIGH-FREQUENCY NOISY OBSERVATIONS

BY NINGNING XIA[1] AND XINGHUA ZHENG[2]

Shanghai University of Finance and Economics and Hong Kong University of Science and Technology

In practice, observations are often contaminated by noise, making the resulting sample covariance matrix a signal-plus-noise sample covariance matrix. Aiming to make inferences about the spectral distribution of the population covariance matrix under such a situation, we establish an asymptotic relationship that describes how the limiting spectral distribution of (signal) sample covariance matrices depends on that of signal-plus-noise-type sample covariance matrices. As an application, we consider inferences about the spectral distribution of integrated covolatility (ICV) matrices of high-dimensional diffusion processes based on high-frequency data with microstructure noise. The (slightly modified) pre-averaging estimator is a signal-plus-noise sample covariance matrix, and the aforementioned result, together with a (generalized) connection between the spectral distribution of signal sample covariance matrices and that of the population covariance matrix, enables us to propose a two-step procedure to consistently estimate the spectral distribution of ICV for a class of diffusion processes. An alternative approach is further proposed, which possesses several desirable properties: it is more robust, it eliminates the effects of microstructure noise, and the asymptotic relationship that enables consistent estimation of the spectral distribution of ICV is the standard Marčenko–Pastur equation. The performance of the two approaches is examined via simulation studies under both synchronous and asynchronous observation settings.

REFERENCES

AÏT-SAHALIA, Y., FAN, J. and LI, Y. (2010). The leverage effect puzzle: Disentangling sources of bias at high frequency. *J. Financ. Econ.* **109** 224–249.

AÏT-SAHALIA, Y., FAN, J. and XIU, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.* **105** 1504–1517. MR2796567

ANDERSEN, T. G. and BOLLERSLEV, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Internat. Econom. Rev.* **39** 885–905.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55. MR1952727

BAI, Z., CHEN, J. and YAO, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52** 423–437. MR2791528

MSC2010 subject classifications. Primary 62H12; secondary 62G99, 60F15.

Key words and phrases. High-dimension, high-frequency, integrated covariance matrices, Marčenko–Pastur equation, microstructure noise.

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

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 253–280. MR1904704

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536. MR2468558

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162** 149–169. MR2795610

CHRISTENSEN, K., KINNEBROCK, S. and PODOLSKIJ, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics* **159** 116–133. MR2720847

DOZIER, R. B. and SILVERSTEIN, J. W. (2007a). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.* **98** 678–694. MR2322123

EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012

EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248

EL KAROUI, N. (2010a). High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *Ann. Statist.* **38** 3487–3566. MR2766860

EL KAROUI, N. (2010b). On information plus noise kernel random matrices. *Ann. Statist.* **38** 3191–3216. MR2722468

EL KAROUI, N. (2013). On the realized risk of high-dimensional Markowitz portfolios. *SIAM J. Financial Math.* **4** 737–783. MR3118251

GLOTER, A. and JACOD, J. (2001). Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Stat.* **5** 243–260. MR1875673

HACHEM, W., LOUBATON, P., MESTRE, X., NAJIM, J. and VALLET, P. (2012). Large information plus noise random matrix models and consistent subspace estimation in large sensor networks. *Random Matrices Theory Appl.* **1** 1150006, 51. MR2934713

HANSEN, P. R. and LUNDE, A. (2006). Realized variance and market microstructure noise. *J. Bus. Econom. Statist.* **24** 127–218. MR2234447

JACOD, J., LI, Y. and ZHENG, X. (2017). Statistical properties of microstructure noise. *Econometrica*. To appear. Available at SSRN, http://ssrn.com/abstract=2212119.

JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. MR1617049

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. and VETTER, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119** 2249–2276. MR2531091

LEDOIT, O. and WOLF, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivariate Anal.* **139** 360–384. MR3349498

LIU, L. Y., PATTON, A. J. and SHEPPARD, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *J. Econometrics* **187** 293–311. MR3347308

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* **72** 507–536. MR0208649

MCNEIL, A. J., FREY, R. and EMBRECHTS, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton Univ. Press, Princeton, NJ. MR2175089

MESTRE, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inform. Theory* **54** 5113–5129. MR2589886

MYKLAND, P. A. and ZHANG, L. (2006). ANOVA for diffusions and Itô processes. *Ann. Statist.* **34** 1931–1963. MR2283722

PODOLSKIJ, M. and VETTER, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* **15** 634–658. MR2555193

UBUKATA, M. and OYA, K. (2009). Estimation and testing for dependence in market microstructure noise. *J. Financ. Econom.* **7** 106–151.

WANG, C. D. and MYKLAND, P. A. (2014). The estimation of leverage effect with high-frequency data. *J. Amer. Statist. Assoc.* **109** 197–215. MR3180557

XIA, N. and ZHENG, X. (2018). Supplement to "On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations." DOI:10.1214/17-AOS1558SUPP.

XIU, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *J. Econometrics* **159** 235–250. MR2720855

ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* **12** 1019–1043. MR2274854

ZHANG, L. (2011). Estimating covariation: Epps effect, microstructure noise. *J. Econometrics* **160** 33–47. MR2745865

ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450

ZHENG, X. and LI, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Ann. Statist.* **39** 3121–3151. MR3012403

The Annals of Statistics
2018, Vol. 46, No. 2, 526–554
https://doi.org/10.1214/17-AOS1559

ONLINE RULES FOR CONTROL OF FALSE DISCOVERY RATE AND FALSE DISCOVERY EXCEEDANCE

BY ADEL JAVANMARD[1] AND ANDREA MONTANARI[2]

University of Southern California and Stanford University

Multiple hypothesis testing is a core problem in statistical inference and arises in almost every scientific field. Given a set of null hypotheses $\mathcal{H}(n) = (H_1, \dots, H_n)$, Benjamini and Hochberg [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** (1995) 289–300] introduced the false discovery rate (FDR), which is the expected proportion of false positives among rejected null hypotheses, and proposed a testing procedure that controls FDR below a preassigned significance level. Nowadays FDR is the criterion of choice for large-scale multiple hypothesis testing.

In this paper we consider the problem of controlling FDR in an *online manner*. Concretely, we consider an ordered—possibly infinite—sequence of null hypotheses $\mathcal{H} = (H_1, H_2, H_3, \dots)$ where, at each step i, the statistician must decide whether to reject hypothesis H_i having access only to the previous decisions. This model was introduced by Foster and Stine [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** (2008) 429–444].

We study a class of *generalized alpha investing* procedures, first introduced by Aharoni and Rosset [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** (2014) 771–794]. We prove that any rule in this class controls online FDR, provided p-values corresponding to true nulls are independent from the other p-values. Earlier work only established mFDR control. Next, we obtain conditions under which generalized alpha investing controls FDR in the presence of general p-values dependencies. We also develop a modified set of procedures that allow to control the false discovery exceedance (the tail of the proportion of false discoveries). Finally, we evaluate the performance of online procedures on both synthetic and real data, comparing them with offline approaches, such as adaptive Benjamini–Hochberg.

REFERENCES

[1] AHARONI, E. and ROSSET, S. (2014). Generalized α-investing: Definitions, optimality results and application to public databases. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 771–794. MR3248676

[2] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876

[3] BARBER, R. F. and CANDES, E. J. (2016). A knockoff filter for high-dimensional selective inference. Available at arXiv:1602.03574.

[4] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300. MR1325392

MSC2010 subject classifications. Primary 62F03, 62F05; secondary 62L99.

Key words and phrases. Hypothesis testing, false discovery rate (FDR), false discovery exceedance (FDX), online decision making.

[5] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

[6] BLANCHARD, G. and ROQUAIN, É. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* **10** 2837–2871. MR2579914

[7] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—Adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717

[8] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields* **99** 277–303. MR1278886

[9] DWORK, C., FELDMAN, V., HARDT, M., PITASSI, T., REINGOLD, O. and ROTH, A. (2015). Preserving statistical validity in adaptive data analysis [extended abstract]. In *STOC'*15—*Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 117–126. ACM, New York. MR3388189

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

[11] FOSTER, D. P. and STINE, R. A. (2008). α-investing: A procedure for sequential control of expected false discoveries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 429–444. MR2424761

[12] GENOVESE, C. R., LAZAR, N. A. and NICHOLS, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15** 870–878.

[13] GENOVESE, C. R., ROEDER, K. and WASSERMAN, L. (2006). False discovery control with p-value weighting. *Biometrika* **93** 509–524. MR2261439

[14] GENOVESE, C. R. and WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.* **101** 1408–1417. MR2279468

[15] G'SELL, M. G., WAGER, S., CHOULDECHOVA, A. and TIBSHIRANI, R. (2016). Sequential selection procedures and false discovery rate control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 423–444. MR3454203

[16] IOANNIDIS, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Jornal of the American Medical Association* **294** 218–228.

[17] IOANNIDIS, J. P. A. (2005). Why most published research findings are false. *Chance* **18** 40–47. MR2216666

[18] JAVANMARD, A. and MONTANARI, A. (2013). Nearly optimal sample size in hypothesis testing for high-dimensional regression. In 51*st Annual Allerton Conference* 1427–1434, Monticello, IL.

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

[20] 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

[21] JAVANMARD, A. and MONTANARI, A. (2015). On online control of false discovery rate. Available at arXiv:1502.06197.

[22] JAVANMARD, A. and MONTANARI, A. (2018). Supplement to "Online rules for control of false discovery rate and false discovery exceedance." DOI:10.1214/17-AOS1559SUPP.

[23] JIN, J. (2008). Proportion of non-zero normal means: Universal oracle equivalences and uniformly consistent estimators. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 461–493. MR2420411

[24] JIN, J. and CAI, T. T. (2007). Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102** 495–506. MR2325113

[25] JOHNSTONE, I. M. (1994). On minimax estimation of a sparse normal mean vector. *Ann. Statist.* **22** 271–289. MR1272083

[26] LEHMANN, E. L. and ROMANO, J. P. (2012). *Generalizations of the Familywise Error Rate*. Springer, Berlin.

[27] LI, A. and BARBER, R. F. (2016). Accumulation tests for FDR control in ordered hypothesis testing. *J. Amer. Statist. Assoc.* **112** 1–38. MR3671774

[28] LIN, D., FOSTER, D. P. and UNGAR, L. H. (2011). VIF regression: A fast regression algorithm for large data. *J. Amer. Statist. Assoc.* **106** 232–247. MR2816717

[29] LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. MR3210970

[30] MEINSHAUSEN, N. and RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393. MR2275246

[31] OWEN, A. B. (2005). Variance of the number of false discoveries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 411–426. MR2155346

[32] PEKELIS, L., WALSH, D. and JOHARI, R. (2015). The new stats engine. Available at http://pages.optimizely.com/rs/optimizely/images/stats_engine_technical_paper.pdf.

[33] PRINZ, F., SCHLANGE, T. and ASADULLAH, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery* **10** 712–712.

[34] REINER, A., YEKUTIELI, D. and BENJAMINI, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19** 368–375.

[35] ROSSET, S., AHARONI, E. and NEUVIRTH, H. (2014). Novel statistical tools for management of public databases facilitate community-wide replicability and control of false discovery. *Genetic Epidemiology* **38** 477–481.

[36] STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498. MR1924302

[37] STOREY, J. D. and TIBSHIRANI, R. (2003). SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In *The Analysis of Gene Expression Data* 272–290. Springer, New York. MR2001400

[38] TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689

[39] 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

[40] VAN DER LAAN, M. J., DUDOIT, S. and POLLARD, K. S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat. Appl. Genet. Mol. Biol.* **3** Art. 15, 27. MR2101464

[41] 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

The Annals of Statistics
2018, Vol. 46, No. 2, 555–579
https://doi.org/10.1214/17-AOS1560

FREQUENCY DOMAIN MINIMUM DISTANCE INFERENCE FOR POSSIBLY NONINVERTIBLE AND NONCAUSAL ARMA MODELS

BY CARLOS VELASCO[1] AND IGNACIO N. LOBATO[2]

Universidad Carlos III de Madrid and Centro de Investigación Económica, ITAM

This article introduces frequency domain minimum distance procedures for performing inference in general, possibly non causal and/or noninvertible, autoregressive moving average (ARMA) models. We use information from higher order moments to achieve identification on the location of the roots of the AR and MA polynomials for non-Gaussian time series. We propose a minimum distance estimator that optimally combines the information contained in second, third, and fourth moments. Contrary to existing estimators, the proposed one is consistent under general assumptions, and may improve on the efficiency of estimators based on only second order moments. Our procedures are also applicable for processes for which either the third or the fourth order spectral density is the zero function.

REFERENCES

ALEKSEEV, V. G. (1993). Asymptotic properties of higher-order periodograms. *Theory Probab. Appl*. **40** 409–419.

ALESSI, L., BARIGOZZI, M. and CAPASSO, M. (2011). Non-fundamentalness in structural econometric models: A review. *Int. Stat. Rev.* **79** 16–47.

ANDREWS, D. W. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory* **3** 348–358.

ANDREWS, B., DAVIS, R. A. and BREIDT, F. J. (2007). Rank-based estimation for all-pass time series models. *Ann. Statist*. **35** 844–869. MR2336871

ANH, V. V., LEONENKO, N. N. and SAKHNO, L. M. (2007). Minimum contrast estimation of random processes based on information of second and third orders. *J. Statist. Plann. Inference* **137** 1302–1331. MR2301481

BREIDT, F. J., DAVIS, R. A. and TRINDADE, A. A. (2001). Least absolute deviation estimation for all-pass time series models. *Ann. Statist*. **29** 919–946. MR1869234

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco, CA. MR0595684

BRILLINGER, D. R. (1985). Fourier inference: Some methods for the analysis of array and nongaussian series data. *Water Resour. Bull*. **21** 744–756.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York. MR1093459

GOSPODINOV, N. and NG, S. (2015). Minimum distance estimation of possibly noninvertible moving average models. *J. Bus. Econom. Statist*. **33** 403–417. MR3372667

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York. MR0279952

MSC2010 subject classifications. 62M10, 62F12.

Key words and phrases. Higher-order moments, higher-order spectra, nonminimum phase, Whittle estimate.

HANSEN, L. P. and SARGENT, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *J. Econom. Dynam. Control* **2** 7–46. MR0601128

HANSEN, L. P. and SARGENT, T. J. (1991). Two difficulties in interpreting vector autoregressions. In *Rational Expectations Econometrics* 77–119. Westview Press, Boulder, CO.

HUANG, J. and PAWITAN, Y. (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scand. J. Stat.* **27** 689–702. MR1804170

KUMON, M. (1992). Identification of non-minimum phase transfer function using higher-order spectrum. *Ann. Inst. Statist. Math.* **44** 239–260. MR1177462

LANNE, M. and SAIKKONEN, P. (2011). Noncausal autoregressions for economic time series. *J. Time Ser. Econom.* **3** Art. 2, 32. MR2928654

LEEPER, E. M., WALKER, T. B. and YANG, S.-C. S. (2013). Fiscal foresight and information flows. *Econometrica* **81** 1115–1145. MR3064062

LII, K.-S. and ROSENBLATT, M. (1992). An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes. *J. Multivariate Anal.* **43** 272–299. MR1193615

MOUNTFORD, A. and UHLIG, H. (2009). What are the effects of fiscal policy shocks? *J. Appl. Econometrics* **24** 960–992. MR2750186

RAMSEY, J. B. and MONTENEGRO, A. (1992). Identification and estimation of non-invertible non-Gaussian MA(q) processes. *J. Econometrics* **54** 301–320.

RAO, C. R. and MITRA, S. K. (1972). Generalized inverse of a matrix and its applications. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 601–620. Univ. California Press, Berkeley, CA. MR0403093

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston, MA. MR0885090

SCHNEEWEISS, H. (2014). The linear GMM model with singular covariance matrix due to the elimination of a nuisance parameter. Technical Report 165, Dept. Statistics, Univ. Munich.

TERDIK, G. (1999). *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis: A Frequency Domain Approach. Lecture Notes in Statistics* **142**. Springer, New York. MR1702281

VELASCO, C. and LOBATO, I. N. (2018). Supplement to "Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models." DOI:10.1214/17-AOS1560SUPP.

WHITTLE, P. (1953). The analysis of multiple stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **15** 125–139. MR0056902

The Annals of Statistics
2018, Vol. 46, No. 2, 580–610
https://doi.org/10.1214/17-AOS1561

ON CONSISTENCY AND SPARSITY FOR SLICED INVERSE REGRESSION IN HIGH DIMENSIONS

BY QIAN LIN*,†,1 ZHIGEN ZHAO‡,2 AND JUN S. LIU*,†,3

*Tsinghua University *, Harvard University† and Temple University‡*

We provide here a framework to analyze the phase transition phenomenon of slice inverse regression (SIR), a supervised dimension reduction technique introduced by Li [*J. Amer. Statist. Assoc.* **86** (1991) 316–342]. Under mild conditions, the asymptotic ratio $\rho = \lim p/n$ is the phase transition parameter and the SIR estimator is consistent if and only if $\rho = 0$. When dimension p is greater than n, we propose a diagonal thresholding screening SIR (DT-SIR) algorithm. This method provides us with an estimate of the eigenspace of $\mathrm{var}(\mathbb{E}[\boldsymbol{x}|y])$, the covariance matrix of the conditional expectation. The desired dimension reduction space is then obtained by multiplying the inverse of the covariance matrix on the eigenspace. Under certain sparsity assumptions on both the covariance matrix of predictors and the loadings of the directions, we prove the consistency of DT-SIR in estimating the dimension reduction space in high-dimensional data analysis. Extensive numerical experiments demonstrate superior performances of the proposed method in comparison to its competitors.

REFERENCES

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885

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

COOK, R. D. (1996). Graphics for regressions with a binary response. *J. Amer. Statist. Assoc.* **91** 983–992. MR1424601

COOK, R. D., FORZANI, L. and ROTHMAN, A. J. (2012). Estimating sufficient reductions of the predictors in abundant high-dimensional regressions. *Ann. Statist.* **40** 353–384. MR3014310

CUI, H., LI, R. and ZHONG, W. (2015). Model-free feature screening for ultrahigh dimensional discriminant analysis. *J. Amer. Statist. Assoc.* **110** 630–641. MR3367253

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

HSING, T. and CARROLL, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20** 1040–1061. MR1165605

JIANG, B. and LIU, J. S. (2014). Variable selection for general index models via sliced inverse regression. *Ann. Statist.* **42** 1751–1786. MR3262467

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

MSC2010 subject classifications. Primary 62J02; secondary 62H25.
Key words and phrases. Dimension reduction, random matrix theory, sliced inverse regression.

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

LI, L. (2007). Sparse sufficient dimension reduction. *Biometrika* **94** 603–613. MR2410011

LI, L. and NACHTSHEIM, C. J. (2006). Sparse sliced inverse regression. *Technometrics* **48** 503–510. MR2328619

LIN, Q., ZHAO, Z. and LIU, J. S. (2018). Supplement to "On consistency and sparsity for sliced inverse regression in high dimensions." DOI:10.1214/17-AOS1561SUPP.

LUO, X., STEFANSKI, L. A. and BOOS, D. D. (2006). Tuning variable selection procedures by adding noise. *Technometrics* **48** 165–175. MR2277672

NEYKOV, M., LIN, Q. and LIU, J. S. (2015). Signed support recovery for single index models in high-dimensions. *Ann. Math. Sci. Appl.* **1** 379–426. DOI:10.4310/AMSA.2016.v1.n2.a5.

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

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

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

WU, Y., BOOS, D. D. and STEFANSKI, L. A. (2007). Controlling variable selection by the addition of pseudovariables. *J. Amer. Statist. Assoc.* **102** 235–243. MR2345541

YU, Z., DONG, Y. and ZHU, L.-X. (2016). Trace pursuit: A general framework for model-free variable selection. *J. Amer. Statist. Assoc.* **111** 813–821. MR3538707

YU, Z., ZHU, L., PENG, H. and ZHU, L. (2013). Dimension reduction and predictor selection in semiparametric models. *Biometrika* **100** 641–654. MR3094442

ZHONG, W., ZHANG, T., ZHU, Y. and LIU, J. S. (2012). Correlation pursuit: Forward stepwise variable selection for index models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 849–870. MR2988909

ZHU, L.-X. and FANG, K.-T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.* **24** 1053–1068. MR1401836

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

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

ZHU, L.-P., LI, L., LI, R. and ZHU, L.-X. (2011). Model-free feature screening for ultrahigh-dimensional data. *J. Amer. Statist. Assoc.* **106** 1464–1475. MR2896849

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
2018, Vol. 46, No. 2, 611–641
https://doi.org/10.1214/17-AOS1562

REGULARIZATION AND THE SMALL-BALL METHOD I: SPARSE RECOVERY

BY GUILLAUME LECUÉ[1,*] AND SHAHAR MENDELSON[2,†]

CREST, CNRS, Université Paris Saclay, and Technion—Israel Institute of Technology and Australian National University†*

We obtain bounds on estimation error rates for regularization procedures of the form

$$\hat{f} \in \underset{f \in F}{\operatorname{argmin}} \left(\frac{1}{N} \sum_{i=1}^{N} (Y_i - f(X_i))^2 + \lambda \Psi(f) \right)$$

when Ψ is a norm and F is convex.

Our approach gives a common framework that may be used in the analysis of learning problems and regularization problems alike. In particular, it sheds some light on the role various notions of sparsity have in regularization and on their connection with the size of subdifferentials of Ψ in a neighborhood of the true minimizer.

As "proof of concept" we extend the known estimates for the LASSO, SLOPE and trace norm regularization.

REFERENCES

[1] ARTSTEIN-AVIDAN, S., GIANNOPOULOS, A. and MILMAN, V. D. (2015). *Asymptotic Geometric Analysis*. *Part I*. *Mathematical Surveys and Monographs* **202**. Amer. Math. Soc., Providence, RI. MR3331351

[2] BACH, F. R. (2010). Structured sparsity-inducing norms through submodular functions. In *Advances in Neural Information Processing Systems* 23: 24*th Annual Conference on Neural Information Processing Systems* 2010. *Proceedings of a Meeting Held* 6-9 *December* 2010 118–126, Vancouver, British Columbia, Canada.

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

[4] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—Adaptive variable selection via convex optimization. *Ann*. *Appl*. *Stat*. **9** 1103–1140. MR3418717

[5] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*: *Methods*, *Theory and Applications*. Springer, Heidelberg. MR2807761

[6] CANDÈS, E. J. and PLAN, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans*. *Inform*. *Theory* **57** 2342–2359. MR2809094

[7] GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans*. *Inform*. *Theory* **57** 1548–1566. MR2815834

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

MSC2010 subject classifications. Primary 60K35, 62G08; secondary 62C20, 62G05, 62G20.
Key words and phrases. Empirical processes, high-dimensional statistics.

[9] KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869

[10] KOLTCHINSKII, V. and MENDELSON, S. (2015). Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN* **23** 12991–13008. MR3431642

[11] LECUÉ, G. and MENDELSON, S. (2013). Learning subgaussian classes: Upper and minimax bounds. Technical report, CNRS, Ecole polytechnique and Technion.

[12] LECUÉ, G. and MENDELSON, S. (2015). Regularization and the small-ball method II: Complexity-based bounds. Technical report, CNRS, ENSAE and Technion, I.I.T.

[13] LECUÉ, G. and MENDELSON, S. (2017). Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc.* (*JEMS*) **19** 881–904. MR3612870

[14] LECUÉ, G. and MENDELSON, S. (2018). Supplement to "Regularization and the small-ball method I: sparse recovery." DOI:10.1214/17-AOS1562SUPP.

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

[16] LOUNICI, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electron. J. Stat.* **2** 90–102. MR2386087

[17] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

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

[19] MENDELSON, S. (2013). Learning without concentration for general loss function. Technical report, Technion, I.I.T. Available at arXiv:1410.3192.

[20] MENDELSON, S. (2014). Learning without concentration. In *Proceedings of the* 27*th Annual Conference on Learning Theory COLT*14 25–39.

[21] MENDELSON, S. (2014). A remark on the diameter of random sections of convex bodies. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **2116** 395–404. Springer, Cham. MR3364699

[22] MENDELSON, S. (2015). Learning without concentration. *J. ACM* **62** Art. 21, 25. MR3367000

[23] MENDELSON, S. (2015). "Local vs. global parameters", breaking the Gaussian compexity barrier. Technical report, Technion, I.I.T.

[24] MENDELSON, S. (2015). On multiplier processes under weak moment assumptions Technical report, Technion, I.I.T.

[25] MENDELSON, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680. MR3565471

[26] NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

[27] 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

[28] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450

[29] RECHT, B., FAZEL, M. and PARRILO, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52** 471–501. MR2680543

[30] ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342

[31] RUDELSON, M. and VERSHYNIN, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Not. IMRN* **19** 9594–9617. MR3431603

[32] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. MR3485953

[33] TALAGRAND, M. (2014). Upper and lower bounds for stochastic processes. In *Modern Methods and Classical Problems. Ergebnisse der Mathematik und Ihrer Grenzgebiete.* 3. *Folge. A Series of Modern Surveys in Mathematics* **60**. Springer, Heidelberg.

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

[35] VAN DE GEER, S. (2014). Weakly decomposable regularization penalties and structured sparsity. *Scand. J. Stat.* **41** 72–86. MR3181133

[36] 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

[37] VAN DE GEER, S. A. (2007). The deterministic lasso. Technical report, ETH Zürich. Available at http://www.stat.math.ethz.ch/~geer/lasso.pdf.

[38] VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809

[39] WATSON, G. A. (1992). Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl.* **170** 33–45. MR1160950

[40] 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

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

The Annals of Statistics
2018, Vol. 46, No. 2, 642–678
https://doi.org/10.1214/17-AOS1563

GAUSSIAN AND BOOTSTRAP APPROXIMATIONS FOR HIGH-DIMENSIONAL U-STATISTICS AND THEIR APPLICATIONS[1]

BY XIAOHUI CHEN

University of Illinois at Urbana–Champaign

This paper studies the Gaussian and bootstrap approximations for the probabilities of a nondegenerate U-statistic belonging to the hyperrectangles in $\mathbb{R}^d$ when the dimension d is large. A two-step Gaussian approximation procedure that does not impose structural assumptions on the data distribution is proposed. Subject to mild moment conditions on the kernel, we establish the explicit rate of convergence uniformly in the class of all hyperrectangles in $\mathbb{R}^d$ that decays polynomially in sample size for a high-dimensional scaling limit, where the dimension can be much larger than the sample size. We also provide computable approximation methods for the quantiles of the maxima of centered U-statistics. Specifically, we provide a unified perspective for the empirical bootstrap, the randomly reweighted bootstrap and the Gaussian multiplier bootstrap with the jackknife estimator of covariance matrix as randomly reweighted quadratic forms and we establish their validity. We show that all three methods are inferentially first-order equivalent for high-dimensional U-statistics in the sense that they achieve the same uniform rate of convergence over all d-dimensional hyperrectangles. In particular, they are asymptotically valid when the dimension d can be as large as $O(e^{n^c})$ for some constant $c \in (0, 1/7)$.

The bootstrap methods are applied to statistical applications for high-dimensional non-Gaussian data including: (i) principled and data-dependent tuning parameter selection for regularized estimation of the covariance matrix and its related functionals; (ii) simultaneous inference for the covariance and rank correlation matrices. In particular, for the thresholded covariance matrix estimator with the bootstrap selected tuning parameter, we show that for a class of sub-Gaussian data, error bounds of the bootstrapped thresholded covariance matrix estimator can be much tighter than those of the minimax estimator with a universal threshold. In addition, we also show that the Gaussian-like convergence rates can be achieved for heavy-tailed data, which are less conservative than those obtained by the Bonferroni technique that ignores the dependency in the underlying data distribution.

REFERENCES

[1] ADAMCZAK, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985

[2] ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586

MSC2010 subject classifications. Primary 62E17; secondary 62F40.

Key words and phrases. U-statistics, high-dimensional inference, Gaussian approximation, bootstrap.

[3] ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542. MR1235426

[4] BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117

[5] BENTKUS, V., GÖTZE, F. and VAN ZWET, W. R. (1997). An Edgeworth expansion for symmetric statistics. *Ann. Statist.* **25** 851–896. MR1439326

[6] BENTKUS, V. Y. (1985). Lower bounds for the rate of convergence in the central limit theorem in Banach spaces. *Litovsk. Mat. Sb.* **25** 10–21. MR0823198

[7] BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

[8] BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1986). The Edgeworth expansion for U-statistics of degree two. *Ann. Statist.* **14** 1463–1484. MR0868312

[9] BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

[10] BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

[11] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. MR2807761

[12] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973

[13] CAI, T. T. and ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607

[14] CALLAERT, H. and VERAVERBEKE, N. (1981). The order of the normal approximation for a studentized U-statistic. *Ann. Statist.* **9** 194–200. MR0600547

[15] CHANG, J., ZHOU, W., ZHOU, W.-X. and WANG, L. (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73** 31–41. MR3632349

[16] CHEN, L. H. Y., FANG, X. and SHAO, Q.-M. (2013). From Stein identities to moderate deviations. *Ann. Probab.* **41** 262–293. MR3059199

[17] CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

[18] CHEN, X. (2016). Gaussian approximation for the sup-norm of high-dimensional matrix-variate U-statistics and its applications. Preprint. Available at arXiv:1602.00199.

[19] CHEN, X. (2018). Supplement to "Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications." DOI:10.1214/17-AOS1563SUPP.

[20] CHEN, X., XU, M. and WU, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* **41** 2994–3021. MR3161455

[21] CHEN, X., XU, M. and WU, W. B. (2016). Regularized estimation of linear functionals of precision matrices for high-dimensional time series. *IEEE Trans. Signal Process.* **64** 6459–6470. MR3566612

[22] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448

[23] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040

[24] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963

[25] DASGUPTA, A., LAHIRI, S. N. and STOYANOV, J. (2014). Sharp fixed n bounds and asymptotic expansions for the mean and the median of a Gaussian sample maximum, and applications to the Donoho–Jin model. *Stat. Methodol.* **20** 40–62. MR3205720

[26] DEHLING, H. and MIKOSCH, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal*. **51** 392–413. MR1321305

[27] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.

[28] DE LA PEÑA, V. C. H. and GINÉ, E. (1999). *Decoupling*: *From Dependence to Independence*, *Randomly Stopped Processes*. *U-Statistics and Processes*. *Martingales and Beyond*. Springer, New York. MR1666908

[29] EINMAHL, U. and LI, D. (2008). Characterization of LIL behavior in Banach space. *Trans*. *Amer. Math*. *Soc*. **360** 6677–6693. MR2434306

[30] EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann*. *Statist*. **36** 2717–2756. MR2485011

[31] FAN, J., LIAO, Y. and MINCHEVA, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Ann*. *Statist*. **39** 3320–3356. MR3012410

[32] GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability*, *II* (*Seattle*, *WA*, 1999). *Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. MR1857312

[33] GÖTZE, F. (1987). Approximations for multivariate U-statistics. *J. Multivariate Anal*. **22** 212–229. MR0899659

[34] GREGORY, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann*. *Statist*. **5** 110–123. MR0433669

[35] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann*. *Math*. *Stat*. **19** 293–325. MR0026294

[36] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist*. *Assoc*. **58** 13–30. MR0144363

[37] HOUDRÉ, C. and REYNAUD-BOURET, P. (2003). Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications*. *Progress in Probability* **56** 55–69. Birkhäuser, Basel. MR2073426

[38] HSING, T. and WU, W. B. (2004). On weighted U-statistics for stationary processes. *Ann*. *Probab*. **32** 1600–1631. MR2060311

[39] HUŠKOVÁ, M. and JANSSEN, P. (1993). Consistency of the generalized bootstrap for degenerate U-statistics. *Ann*. *Statist*. **21** 1811–1823. MR1245770

[40] HUŠKOVÁ, M. and JANSSEN, P. (1993). Generalized bootstrap for studentized U-statistics: A rank statistic approach. *Statist*. *Probab. Lett*. **16** 225–233. MR1208512

[41] JANSSEN, P. (1994). Weighted bootstrapping of U-statistics. *J. Statist*. *Plann*. *Inference* **38** 31–41. MR1256846

[42] KLEIN, T. and RIO, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann*. *Probab*. **33** 1060–1077. MR2135312

[43] LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann*. *Statist*. **37** 4254–4278. MR2572459

[44] 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

[45] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015

[46] LEHMANN, E. L. (1999). *Elements of Large-Sample Theory*. Springer, New York. MR1663158

[47] LO, A. Y. (1987). A large sample study of the Bayesian bootstrap. *Ann*. *Statist*. **15** 360–375. MR0885742

[48] MAI, Q., ZOU, H. and YUAN, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika* **99** 29–42. MR2899661

[49] MASON, D. M. and NEWTON, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann*. *Statist*. **20** 1611–1624. MR1186268

[50] MASSART, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863–884. MR1782276

[51] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

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

[53] NAGAEV, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR0542129

[54] PENG, J., WANG, P., ZHOU, N. and ZHU, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Amer. Statist. Assoc.* **104** 735–746. MR2541591

[55] PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York. MR0388499

[56] PORTNOY, S. (1986). On the central limit theorem in $\mathbf{R}^p$ when $p \to \infty$. *Probab. Theory Related Fields* **73** 571–583. MR0863546

[57] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

[58] ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. MR2417391

[59] RUBIN, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134. MR0600538

[60] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

[61] SHAO, Q.-M. and WANG, Q. (2013). Self-normalized limit theorems: A survey. *Probab. Surv.* **10** 69–93. MR3161676

[62] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006

[63] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

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

[65] WANG, Q. and JING, B.-Y. (2004). Weighted bootstrap for U-statistics. *J. Multivariate Anal.* **91** 177–198. MR2087842

[66] YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. MR2719856

[67] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824

[68] ZHANG, C.-H. (1999). Sub-Bernoulli functions, moment inequalities and strong laws for nonnegative and symmetrized U-statistics. *Ann. Probab.* **27** 432–453. MR1681165

[69] ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high-dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156

[70] ZHANG, X. and CHENG, G. (2014). Bootstrapping high dimensional time series. Available at arXiv:1406.1037.

The Annals of Statistics
2018, Vol. 46, No. 2, 679–710
https://doi.org/10.1214/17-AOS1564

SELECTIVE INFERENCE WITH A RANDOMIZED RESPONSE

BY XIAOYING TIAN AND JONATHAN TAYLOR[1]

Stanford University

Inspired by sample splitting and the reusable holdout introduced in the field of differential privacy, we consider selective inference with a randomized response. We discuss two major advantages of using a randomized response for model selection. First, the selectively valid tests are more powerful after randomized selection. Second, it allows consistent estimation and weak convergence of selective inference procedures. Under independent sampling, we prove a selective (or privatized) central limit theorem that transfers procedures valid under asymptotic normality without selection to their corresponding selective counterparts. This allows selective inference in nonparametric settings. Finally, we propose a framework of inference after combining multiple randomized selection procedures. We focus on the classical asymptotic setting, leaving the interesting high-dimensional asymptotic questions for future work.

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580. MR0189095

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

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BENJAMINI, Y. and STARK, P. B. (1996). Nonequivariant simultaneous confidence intervals less likely to contain zero. *J. Amer. Statist. Assoc.* **91** 329–337. MR1394088

BÜHLMANN, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19** 1212–1242. MR3102549

CHATTERJEE, S. (2005). A simple invariance theorem. Preprint. Available at arXiv:math/0508213.

CHUNG, E. and ROMANO, J. P. (2013). Exact and asymptotically robust permutation tests. *Ann. Statist.* **41** 484–507. MR3099111

COX, D. R. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika* **62** 441–444. MR0378189

DWORK, C., FELDMAN, V., HARDT, M., PITASSI, T., REINGOLD, O. and ROTH, A. (2015). Preserving statistical validity in adaptive data analysis [extended abstract]. In *STOC'15—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 117–126. ACM, New York. MR3388189

FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. Available at arXiv:1410.2597.

FITHIAN, W., TAYLOR, J., TIBSHIRANI, R. and TIBSHIRANI, R., (2015). Selective sequential model selection. Available at arXiv:1512.02565.

MSC2010 subject classifications. Primary 62M40; secondary 62J05.
Key words and phrases. Selective inference, nonparametric, differential privacy.

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

HARRIS, X. T. (2016). Prediction error after model search. Preprint. Available at arXiv:1610.06107.

HARRIS, X. T., PANIGRAHI, S., MARKOVIC, J., BI, N. and TAYLOR, J. (2016). Selective sampling after solving a convex problem. Preprint. Available at arXiv:1609.05609.

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

LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York. MR0852406

LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. MR3210970

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 417–473. MR2758523

MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584

PAKMAN, A. and PANINSKI, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *J. Comput. Graph. Statist.* **23** 518–542. MR3215823

ROSENTHAL, R. (1979). The file drawer problem and tolerance for null results. *Psychol. Bull.* **86** 638.

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

TIAN, X., BI, N. and TAYLOR, J. (2016). Magic: A general, powerful and tractable method for selective inference. Preprint. Available at arXiv:1607.02630.

TIAN, X., LOFTUS, J. R. and TAYLOR, J. E. (2015). Selective inference with unknown variance via the square-root lasso. Preprint. Available at arXiv:1504.08031.

TIAN, X. and TAYLOR, J. (2015). Asymptotics of selective inference. Available at arXiv:1501.03588.

TIAN, X. and TAYLOR, J. (2018). Supplement to "Selective inference with a randomized response." DOI:10.1214/17-AOS1564SUPP.

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

TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2015). Uniform asymptotic inference and the bootstrap after model selection. Preprint. Available at arXiv:1506.06266.

TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689

TUKEY, J. W. (1980). We need both exploratory and confirmatory. *Amer. Statist.* **34** 23–25.

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

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

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

The Annals of Statistics
2018, Vol. 46, No. 2, 711–744
https://doi.org/10.1214/17-AOS1565

MULTISCALE BLIND SOURCE SEPARATION

BY MERLE BEHR*,[1], CHRIS HOLMES†,[2] AND AXEL MUNK*,‡,[3]

*University of Goettingen**, *University of Oxford*†, *and*
Max Planck Institute for Biophysical Chemistry‡

We provide a new methodology for statistical recovery of single linear mixtures of piecewise constant signals (sources) with unknown mixing weights and change points in a multiscale fashion. We show exact recovery within an ε-neighborhood of the mixture when the sources take only values in a known finite alphabet. Based on this we provide the SLAM (Separates Linear Alphabet Mixtures) estimators for the mixing weights and sources. For Gaussian error, we obtain uniform confidence sets and optimal rates (up to log-factors) for all quantities. SLAM is efficiently computed as a nonconvex optimization problem by a dynamic program tailored to the finite alphabet assumption. Its performance is investigated in a simulation study. Finally, it is applied to assign copy-number aberrations from genetic sequencing data to different clones and to estimate their proportions.

REFERENCES

[1] AÏSSA-EL-BEY, A., PASTOR, D., SBAÏ, S. M. A. and FADLALLAH, Y. (2015). Sparsity-based recovery of finite alphabet solutions to underdetermined linear systems. *IEEE Trans. Inform. Theory* **61** 2008–2018. MR3332994

[2] ARORA, S., GE, R., KANNAN, R. and MOITRA, A. (2012). Computing a nonnegative matrix factorization—provably. In *STOC'12—Proceedings of the* 2012 *ACM Symposium on Theory of Computing* 145–161. ACM, New York. MR2961503

[3] ARORA, S., GE, R., MOITRA, A. and SACHDEVA, S. (2015). Provable ICA with unknown Gaussian noise, and implications for Gaussian mixtures and autoencoders. *Algorithmica* **72** 215–236. MR3332931

[4] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

[5] BEHR, M., HOLMES, C. and MUNK, A. (2018). Supplement to "Multiscale blind source separation." DOI:10.1214/17-AOS1565SUPP.

[6] BEHR, M. and MUNK, A. (2015). Identifiability for blind source separation of multiple finite alphabet linear mixtures. *IEEE Trans. Inform. Theory* **63** 5506–5517.

[7] BELKIN, M., RADEMACHER, L. and VOSS, J. (2013). Blind signal separation in the presence of Gaussian noise. *J. Mach. Learn. Res. Proc.* **30** 270–287.

[8] BEROUKHIM, R., MERMEL, C. H., PORTER, D., WEI, G., RAYCHAUDHURI, S., DONOVAN, J., BARRETINA, J., BOEHM, J. S., DOBSON, J., URASHIMA, M. et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* **463** 899–905.

[9] BIOGLIO, V., COLUCCIA, G. and MAGLI, E. (2014). Sparse image recovery using compressed sensing over finite alphabets. *IEEE Int. Conf. Image Process.* (*ICIP*) 1287–1291.

MSC2010 subject classifications. Primary 62G08, 62G15; secondary 92D10.
Key words and phrases. Multiscale inference, honest confidence sets, change point regression, finite alphabet linear mixture, exact recovery, genetic sequencing.

[10] BOFILL, P. and ZIBULEVSKY, M. (2001). Underdetermined blind source separation using sparse representations. *Signal Process*. **81** 2353–2362.

[11] BOYSEN, L., KEMPE, A., LIEBSCHER, V., MUNK, A. and WITTICH, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann*. *Statist*. **37** 157–183. MR2488348

[12] CANDES, E. J. and TAO, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans*. *Inform*. *Theory* **52** 5406–5425. MR2300700

[13] CARLSTEIN, E., MÜLLER, H.-G. and SIEGMUND, D., eds. (1994). *Change-Point Problems*. *Lecture Notes—Monograph Series* **23**. IMS, Hayward, CA. MR1477909

[14] CARTER, S. L., CIBULSKIS, K., HELMAN, E., MCKENNA, A., SHEN, H., ZACK, T., LAIRD, P. W., ONOFRIO, R. C., WINCKLER, W., WEIR, B. A. et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat*. *Biotechnol*. **30** 413–421.

[15] CHEN, H., XING, H. and ZHANG, N. R. (2011). Estimation of parent specific DNA copy number in tumors using high-density genotyping arrays. *PLoS Comput*. *Biol*. **7** e1001060. MR2776334

[16] CHENG, M.-Y. and HALL, P. (1999). Mode testing in difficult cases. *Ann*. *Statist*. **27** 1294–1315. MR1740110

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

[18] DAS, A. K. and VISHWANATH, S. (2013). On finite alphabet compressive sensing. *IEEE Int*. *Conf*. *Acoust*., *Speech Signal Process*. (*ICASSP*) 5890–5894.

[19] DAVIES, L., HÖHENRIEDER, C. and KRÄMER, W. (2012). Recursive computation of piecewise constant volatilities. *Comput*. *Statist*. *Data Anal*. **56** 3623–3631. MR2943916

[20] DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution. *Ann*. *Statist*. **29** 1–65. MR1833958

[21] DETTE, H., MUNK, A. and WAGNER, T. (1998). Estimating the variance in nonparametric regression—what is a reasonable choice? *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **60** 751–764. MR1649480

[22] DIAMANTARAS, K. I. (2006). A clustering approach for the blind separation of multiple finite alphabet sequences from a single linear mixture. *Signal Process*. **86** 877–891.

[23] DING, L., WENDL, M. C., MCMICHAEL, J. F. and RAPHAEL, B. J. (2014). Expanding the computational toolbox for mining cancer genomes. *Nat*. *Rev*. *Genet*. **15** 556–570.

[24] DONOHO, D. and STODDEN, V. (2003). When does non-negative matrix factorization give a correct decomposition into parts? *Adv*. *Neural Inf*. *Process*. *Syst*. **16**.

[25] DONOHO, D. L. (2006). Compressed sensing. *IEEE Trans*. *Inform*. *Theory* **52** 1289–1306. MR2241189

[26] DRAPER, S. C. and MALEKPOUR, S. (2009). Compressed sensing over finite fields. *Proceedings of the* 2009 *IEEE international conference on Symposium on Information Theory* **1** 669–673.

[27] DU, C., KAO, C.-L. M. and KOU, S. C. (2016). Stepwise signal extraction via marginal likelihood. *J*. *Amer*. *Statist*. *Assoc*. **111** 314–330. MR3494662

[28] DÜMBGEN, L., PITERBARG, V. I. and ZHOLUD, D. (2006). On the limit distribution of multiscale test statistics for nonparametric curve estimation. *Math*. *Methods Statist*. **15** 20–25. MR2225428

[29] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann*. *Statist*. **29** 124–152. MR1833961

[30] DÜMBGEN, L. and WALTHER, G. (2008). Multiscale inference about a density. *Ann*. *Statist*. **36** 1758–1785. MR2435455

[31] FEARNHEAD, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Stat*. *Comput*. **16** 203–213. MR2227396

[32] FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728
[33] FRIEDRICH, F., KEMPE, A., LIEBSCHER, V. and WINKLER, G. (2008). Complexity penalized M-estimation: Fast computation. *J. Comput. Graph. Statist.* **17** 201–224. MR2424802
[34] FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979
[35] FUTSCHIK, A., HOTZ, T., MUNK, A. and SIELING, H. (2014). Multiscale DNA partitioning: Statistical evidence for segments. *Bioinformatics* **30** 2255–2262.
[36] GREAVES, M. and MALEY, C. C. (2012). Clonal evolution in cancer. *Nature* **481** 306–313.
[37] GU, F., ZHANG, H., LI, N. and LU, W. (2010). Blind separation of multiple sequences from a single linear mixture using finite alphabet. *IEEE Int. Conf. Wirel. Commun. Signal Process. (WCSP)* 1–5.
[38] HA, G., ROTH, A., KHATTRA, J., HO, J., YAP, D., PRENTICE, L. M., MELNYK, N., MCPHERSON, A., BASHASHATI, A., LAKS, E. et al. (2014). TITAN: Inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* **24** 1881–1893.
[39] HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842
[40] HARCHAOUI, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. MR2796565
[41] JENG, X. J., CAI, T. T. and LI, H. (2010). Optimal sparse segment identification with application in copy number variation analysis. *J. Amer. Statist. Assoc.* **105** 1156–1166. MR2752611
[42] KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418
[43] KOFIDIS, N., MARGARIS, A., DIAMANTARAS, K. and ROUMELIOTIS, M. (2008). Blind system identification: Instantaneous mixtures of n sources. *Int. J. Comput. Math.* **85** 1333–1340. MR2451492
[44] LEE, D. and SEUNG, S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* **401** 788–791.
[45] LEE, T. W., LEWICKI, M. S., GIROLAMI, M. and SEJNOWSKI, T. J. (1999). Blind source separation of more sources than mixtures using overcomplete representations. *Signal Process. Lett.* **6** 87–90.
[46] LI, J., RAY, S. and LINDSAY, B. G. (2007). A nonparametric statistical approach to clustering via mode identification. *J. Mach. Learn. Res.* **8** 1687–1723. MR2332445
[47] LI, Y., AMARI, S. I., CICHOCKI, A., HO, D. W. and XIE, S. (2006). Underdetermined blind source separation based on sparse representation. *IEEE Trans. Signal Process.* **54** 423–437.
[48] LIU, B., MORRISON, C. D., JOHNSON, C. S., TRUMP, D. L., QIN, M., CONROY, J. C., WANG, J. and LIU, S. (2013). Computational methods for detecting copy number variations in cancer genome using next generation sequencing: Principles and challenges. *Oncotarget* **4** 1868.
[49] MATTESON, D. S. and JAMES, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Amer. Statist. Assoc.* **109** 334–345. MR3180567
[50] MÜLLER, H.-G. and STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625. MR0888429
[51] NIU, Y. S. and ZHANG, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat.* **6** 1306–1326. MR3012531

[52] OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostat*. **5** 557–572.

[53] OOI, H. (2002). Density visualization and mode hunting using trees. *J. Comput. Graph. Statist.* **11** 328–347. MR1938139

[54] PAJUNEN, P. (1997). Blind separation of binary sources with less sensors than sources. *IEEE Int. Conf. Neural Netw.* **3** 1994–1997.

[55] POLONIK, W. (1998). The silhouette, concentration functions and ML-density estimation under order restrictions. *Ann. Statist.* **26** 1857–1877. MR1673281

[56] PROAKIS, J. G. (1995). *Digital Communications*. McGraw-Hill, New York.

[57] RECHT, B., RE, C., TROPP, J. and BITTORF, V. (2012). Factoring nonnegative matrices with linear programs. *Adv. Neural Inf. Process. Syst.* **25** 1214–1222.

[58] ROSENBERG, A. and HIRSCHBERG, J. (2007). V-measure: A conditional entropy-based external cluster evaluation measure. *EMNLP-CoNLL* **7** 410–420.

[59] ROSTAMI, M., BABAIE-ZADEH, M., SAMADI, S. and JUTTEN, C. (2011). Blind source separation of discrete finite alphabet sources using a single mixture. *IEEE Stat. Signal Process. Workshop* (*SSP*) 709–712.

[60] ROTH, A., KHATTRA, J., YAP, D., WAN, A., LAKS, E., BIELE, J., HA, G., APARICIO, S., BOUCHARD-CÔTÉ, A. and SHAH, S. P. (2014). PyClone: Statistical inference of clonal population structure in cancer. *Nat. Methods* **11** 396–398.

[61] SHAH, S. P., ROTH, A., GOYA, R., OLOUMI, A., HA, G., ZHAO, Y., TURASHVILI, G., DING, J., TSE, K., HAFFARI, G. et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486** 395–399.

[62] SIEGMUND, D. (2013). Change-points: From sequential detection to biology and back. *Sequential Anal.* **32** 2–14. MR3023983

[63] SIEGMUND, D. and YAKIR, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6** 191–213. MR1748719

[64] SPIELMAN, D. A., WANG, H. and WRIGHT, J. (2012). Exact recovery of sparsely-used dictionaries. *J. Mach. Learn. Res. Proc.* **23** 37.1–37.18.

[65] SPOKOINY, V. (2009). Multiscale local change point detection with applications to value-at-risk. *Ann. Statist.* **37** 1405–1436. MR2509078

[66] TALWAR, S., VIBERG, M. and PAULRAJ, A. (1996). Blind separation of synchronous co-channel digital signals using an antenna array—Part I. algorithms. *IEEE Trans. Signal Process.* **44** 1184–1197.

[67] TIBSHIRANI, R., WALTHER, G. and HASTIE, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503

[68] TIBSHIRANI, R. and WANG, P. (2008). Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostat*. **9** 18–29.

[69] VERDÚ, S. (1998). *Multiuser Detection*. Cambridge University Press, Cambridge.

[70] WALTHER, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033. MR2604703

[71] YAU, C., PAPASPILIOPOULOS, O., ROBERTS, G. O. and HOLMES, C. (2011). Bayesian non-parametric hidden Markov models with applications in genomics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 37–57. MR2797735

[72] YUANQING, L., CICHOCKI, A. and ZHANG, L. (2003). Blind separation and extraction of binary sources. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **86** 580–589.

[73] ZHANG, N. R. and SIEGMUND, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63** 22–32. MR2345571

[74] ZHANG, N. R. and SIEGMUND, D. O. (2012). Model selection for high-dimensional, multi-sequence change-point problems. *Statist. Sinica* **22** 1507–1538. MR3027097

The Annals of Statistics
2018, Vol. 46, No. 2, 745–780
https://doi.org/10.1214/17-AOS1566

SHARP ORACLE INEQUALITIES FOR LEAST SQUARES ESTIMATORS IN SHAPE RESTRICTED REGRESSION[1]

BY PIERRE C. BELLEC

ENSAE, UMR CNRS 9194 and Rutgers University

The performance of Least Squares (LS) estimators is studied in shape-constrained regression models under Gaussian and sub-Gaussian noise. General bounds on the performance of LS estimators over closed convex sets are provided. These results have the form of sharp oracle inequalities that account for the model misspecification error. In the presence of misspecification, these bounds imply that the LS estimator estimates the projection of the true parameter at the same rate as in the well-specified case.

In isotonic and unimodal regression, the LS estimator achieves the nonparametric rate $n^{-2/3}$ as well as a parametric rate of order k/n up to logarithmic factors, where k is the number of constant pieces of the true parameter. In univariate convex regression, the LS estimator satisfies an adaptive risk bound of order q/n up to logarithmic factors, where q is the number of affine pieces of the true regression function. This adaptive risk bound holds for any collection of design points. While Guntuboyina and Sen [*Probab. Theory Related Fields* **163** (2015) 379–411] established that the nonparametric rate of convex regression is of order $n^{-4/5}$ for equispaced design points, we show that the nonparametric rate of convex regression can be as slow as $n^{-2/3}$ for some worst-case design points. This phenomenon can be explained as follows: Although convexity brings more structure than unimodality, for some worst-case design points this extra structure is uninformative and the nonparametric rates of unimodal regression and convex regression are both $n^{-2/3}$. Higher order cones, such as the cone of β-monotone sequences, are also studied.

REFERENCES

[1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453

[2] BARTLETT, P. L. and MENDELSON, S. (2006). Empirical minimization. *Probab. Theory Related Fields* **135** 311–334. MR2240689

[3] BELLEC, P. C. (2018). Supplement to "Sharp oracle inequalities for Least Squares estimators in shape restricted regression." DOI:10.1214/17-AOS1566SUPP.

[4] BELLEC, P. C., LECUÉ, G. and TSYBAKOV, A. B. (2016). Slope meets lasso: Improved oracle bounds and optimality. ArXiv preprint. Available at arXiv:1605.08651.

[5] BELLEC, P. C. and TSYBAKOV, A. B. (2015). Sharp oracle bounds for monotone and convex regression through aggregation. *J. Mach. Learn. Res.* **16** 1879–1892. MR3417801

MSC2010 subject classifications. 62G08, 62C20.
Key words and phrases. Shape restricted regression, convexity, minimax rates, Gaussian width, concentration.

[6] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*. *A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193
[7] CHANDRASEKARAN, V. and JORDAN, M. I. (2013). Computational and statistical tradeoffs via convex relaxation. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **110** E1181–E1190. MR3047651
[8] 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
[9] CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann*. *Statist*. **42** 2340–2381. MR3269982
[10] CHATTERJEE, S. (2016). An improved global risk bound in concave regression. *Electron*. *J*. *Stat*. **10** 1608–1629. MR3522655
[11] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann*. *Statist*. **43** 1774–1800. MR3357878
[12] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On matrix estimation under monotonicity constraints. ArXiv preprint. Available at arXiv:1506.03430.
[13] CHATTERJEE, S. and LAFFERTY, J. (2017). Adaptive risk bounds in unimodal regression. *Bernoulli*. To appear. Available at arXiv:1512.02956.
[14] FLAMMARION, N., MAO, C. and RIGOLLET, P. (2016). Optimal rates of statistical seriation. ArXiv preprint. Available at arXiv:1607.02435.
[15] GAO, F. and WELLNER, J. A. (2007). Entropy estimate for high-dimensional monotonic functions. *J*. *Multivariate Anal*. **98** 1751–1764. MR2392431
[16] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab*. *Theory Related Fields* **163** 379–411. MR3405621
[17] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015
[18] MEYER, M. and WOODROOFE, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann*. *Statist*. **28** 1083–1104. MR1810920
[19] OYMAK, S. and HASSIBI, B. (2016). Sharp MSE bounds for proximal denoising. *Found*. *Comput*. *Math*. **16** 965–1029. MR3529131
[20] OYMAK, S., RECHT, B. and SOLTANOLKOTABI, M. (2015). Sharp time–data tradeoffs for linear inverse problems. ArXiv preprint. Available at arXiv:1507.04793.
[21] PLAN, Y., VERSHYNIN, R. and YUDOVINA, E. (2017). High-dimensional estimation with geometric constraints. *Inf*. *Inference* **6** 1–40. MR3636866
[22] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original, translated by Vladimir Zaiats. MR2724359
[23] VAN DE GEER, S. and WAINWRIGHT, M. (2015). On concentration for (regularized) empirical risk minimization. ArXiv preprint. Available at arXiv:1512.00677.
[24] VERSHYNIN, R. (2015). Estimation in high dimensions: A geometric perspective. In *Sampling Theory*, *A Renaissance*. *Appl*. *Numer*. *Harmon*. *Anal*. 3–66. Birkhäuser, Basel. MR3467418
[25] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann*. *Statist*. **30** 528–555. MR1902898

The Annals of Statistics
2018, Vol. 46, No. 2, 781–813
https://doi.org/10.1214/17-AOS1567

ORACLE INEQUALITIES FOR SPARSE ADDITIVE QUANTILE REGRESSION IN REPRODUCING KERNEL HILBERT SPACE

BY SHAOGAO LV[*,1], HUAZHEN LIN[†,1], HENG LIAN[‡] AND JIAN HUANG[§]

Nanjing Audit University[*], *Southwestern University of Finance and Economics*[†], *City University of Hong Kong*[‡], *and University of Iowa*[§]

This paper considers the estimation of the sparse additive quantile regression (SAQR) in high-dimensional settings. Given the nonsmooth nature of the quantile loss function and the nonparametric complexities of the component function estimation, it is challenging to analyze the theoretical properties of ultrahigh-dimensional SAQR. We propose a regularized learning approach with a two-fold Lasso-type regularization in a reproducing kernel Hilbert space (RKHS) for SAQR. We establish nonasymptotic oracle inequalities for the excess risk of the proposed estimator without any coherent conditions. If additional assumptions including an extension of the restricted eigenvalue condition are satisfied, the proposed method enjoys sharp oracle rates without the light tail requirement. In particular, the proposed estimator achieves the minimax lower bounds established for sparse additive mean regression. As a by-product, we also establish the concentration inequality for estimating the population mean when the general Lipschitz loss is involved. The practical effectiveness of the new method is demonstrated by competitive numerical results.

REFERENCES

[1] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. MR0051437

[2] BACH, F., JENATTON, R., MAIRAL, J. and OBOZINSKI, G. (2012). Convex optimization with sparsity-inducing norms. In *Optimization for Machine Learning*. MIT Press, Cambridge, MA.

[3] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. MR2166554

[4] BARTLETT, P. L. and MENDELSON, S. (2002). Rademacher and Gaussian complexities: Risk bounds and structural results. *J. Mach. Learn. Res.* **3** 463–482. MR1984026

[5] BECK, A. and TEBOULLE, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** 183–202. MR2486527

[6] BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. MR2797841

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

[8] BOUSQUET, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** 495–500. MR1890640

MSC2010 subject classifications. Primary 62G20; secondary 62G05.

Key words and phrases. Quantile regression, additive models, sparsity, regularization methods, reproducing kernel Hilbert space.

[9] BREHENY, P. and HUANG, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Stat. Comput.* **25** 173–187. MR3306699

[10] BUCHINSKY, M. (1994). Changes in the U.S. wage structure 1963–1987: Application of quantile regression. *Econometrica* **62** 405–458.

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

[12] CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. MR2847974

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

[14] CHRISTMANN, A. and ZHOU, D.-X. (2016). Learning rates for the risk of kernel-based quantile regression estimators in additive models. *Anal. Appl.* (*Singap.*) **14** 449–477. MR3486095

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

[16] HE, X. (2009). Modeling and inference by quantile regression. Technical report, Dept. Statistics, Univ. Illinois at Urbana–Champaign.

[17] HE, X., WANG, L. and HONG, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41** 342–369. MR3059421

[18] HOROWITZ, J. L. and LEE, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100** 1238–1249. MR2236438

[19] HUANG, J., HOROWITZ, J. L. and WEI, F. (2010). Variable selection in nonparametric additive models. *Ann. Statist.* **38** 2282–2313. MR2676890

[20] HUNTER, D. R. and LANGE, K. (2000). Quantile regression via an MM algorithm. *J. Comput. Graph. Statist.* **9** 60–77. MR1819866

[21] HUNTER, D. R. and LANGE, K. (2004). A tutorial on MM algorithms. *Amer. Statist.* **58** 30–37. MR2055509

[22] KATO, K. (2016). Group Lasso for high dimensional sparse quantile regression models. arXiv:1103.1458.

[23] KOENKER, R. (2005). *Quantile Regression*. *Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657

[24] KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[25] KOENKER, R., ROGER, W. and D'OREY, V. (1987). Algorithm AS 229: Computing regression quantiles. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **36** 383–384.

[26] KOLTCHINSKII, V. and YUAN, M. (2010). Sparsity in multiple kernel learning. *Ann. Statist.* **38** 3660–3695. MR2766864

[27] LI, Y., LIU, Y. and ZHU, J. (2007). Quantile regression in reproducing kernel Hilbert spaces. *J. Amer. Statist. Assoc.* **102** 255–268. MR2293307

[28] LIAN, H. (2012). Semiparametric estimation of additive quantile regression models by two-fold penalty. *J. Bus. Econom. Statist.* **30** 337–350. MR2969217

[29] LIN, Y. and ZHANG, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* **34** 2272–2297. MR2291500

[30] LV, J. and FAN, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.* **37** 3498–3528. MR2549567

[31] LV, S., HE, X. and WANG, J. (2017). A unified penalized method for sparse additive quantile models: An RKHS approach. *Ann. Inst. Statist. Math.* **69** 897–923. MR3671643

[32] LV, S., LIN, H., LIAN, H. and HUANG, J. (2018). Supplement to "Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space." DOI:10.1214/17-AOS1567SUPP.
[33] MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Ann. Statist.* **37** 3779–3821. MR2572443
[34] MENDELSON, S. (2002). Geometric parameters of kernel machines. In *Computational Learning Theory* (*Sydney*, 2002). *Lecture Notes in Computer Science* **2375** 29–43. Springer, Berlin. MR2040403
[35] PEARCE, N. D. and WAND, M. P. (2006). Penalized splines and reproducing kernel methods. *Amer. Statist.* **60** 233–240. MR2246756
[36] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704
[37] RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 1009–1030. MR2750255
[38] RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771. MR2816337
[39] SCHOLKÖPF, B. and SMOLA, A. (2002). *Learning with Kernels*: *Support Vector Machine*, *Regularization*, *Optimization*, *and Beyond*. MIT Press, Cambridge, MA.
[40] STEINWART, I. and CHRISTMANN, A. (2008). *Support Vector Machines*. Springer, New York. MR2450103
[41] STEINWART, I. and CHRISTMANN, A. (2011). Estimating conditional quantiles with the help of the pinball loss. *Bernoulli* **17** 211–225. MR2797989
[42] SUZUKI, T. and SUGIYAMA, M. (2013). Fast learning rate of multiple kernel learning: Trade-off between sparsity and smoothness. *Ann. Statist.* **41** 1381–1405. MR3113815
[43] TARIGAN, B. and VAN DE GEER, S. A. (2006). Classifiers of support vector machine type with l_1 complexity regularization. *Bernoulli* **12** 1045–1076. MR2274857
[44] THE CANCER GENOME ATLAS NETWORK (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490** 61–70.
[45] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. MR1379242
[46] TSENG, P. and YUN, S. (2009). A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program.* **117** 387–423. MR2421312
[47] VAN DE GEER, S. (2002). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.
[48] VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809
[49] WANG, L., WU, Y. and LI, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107** 214–222. MR2949353
[50] WEI, F., HUANG, J. and LI, H. (2011). Variable selection and estimation in high-dimensional varying-coefficient models. *Statist. Sinica* **21** 1515–1540. MR2895107
[51] WU, Y. and LIU, Y. (2009). Variable selection in quantile regression. *Statist. Sinica* **19** 801–817. MR2514189
[52] 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
[53] ZHANG, X., WU, Y., WANG, L. and LI, R. (2016). Variable selection for support vector machines in moderately high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 53–76. MR3453646
[54] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

The Annals of Statistics
2018, Vol. 46, No. 2, 814–841
https://doi.org/10.1214/17-AOS1568

I-LAMM FOR SPARSE LEARNING: SIMULTANEOUS CONTROL OF ALGORITHMIC COMPLEXITY AND STATISTICAL ERROR[1]

BY JIANQING FAN[*,†], HAN LIU[†], QIANG SUN[†,‡] AND TONG ZHANG[*,§]

Fudan University[*], *Princeton University*[†], *University of Toronto*[‡] *and Tecent AI Lab*[§]

We propose a computational framework named iterative local adaptive majorize-minimization (I-LAMM) to simultaneously control algorithmic complexity and statistical error when fitting high-dimensional models. I-LAMM is a two-stage algorithmic implementation of the local linear approximation to a family of folded concave penalized quasi-likelihood. The first stage solves a convex program with a crude precision tolerance to obtain a coarse initial estimator, which is further refined in the second stage by iteratively solving a sequence of convex programs with smaller precision tolerances. Theoretically, we establish a phase transition: the first stage has a sublinear iteration complexity, while the second stage achieves an improved linear rate of convergence. Though this framework is completely algorithmic, it provides solutions with optimal statistical performances and controlled algorithmic complexity for a large family of nonconvex optimization problems. The iteration effects on statistical errors are clearly demonstrated via a contraction property. Our theory relies on a localized version of the sparse/restricted eigenvalue condition, which allows us to analyze a large family of loss and penalty functions and provide optimality guarantees under very weak assumptions (e.g., I-LAMM requires much weaker minimal signal strength than other procedures). Thorough numerical results are provided to support the obtained theory.

REFERENCES

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Fast global convergence of gradient methods for high-dimensional statistical recovery. *Ann. Statist.* **40** 2452–2482. MR3097609

BECK, A. and TEBOULLE, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** 183–202. MR2486527

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

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

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575

BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396

MSC2010 subject classifications. Primary 62J07; secondary 62C20, 62H35.

Key words and phrases. Algorithmic statistics, iteration complexity, local adaptive MM, nonconvex statistical optimization, optimal rate of convergence.

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*. Springer, Heidelberg. MR2807761

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

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

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

FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans*. *Inform*. *Theory* **57** 5467–5484. MR2849368

FAN, J., XUE, L. and ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *Ann*. *Statist*. **42** 819–849. MR3210988

FAN, J., LIU, H., SUN, Q. and ZHANG, T. (2018). Supplement to "I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error." DOI:10.1214/17-AOS1568SUPP.

FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann*. *Appl*. *Stat*. **1** 302–332. MR2415737

HUNTER, D. R. and LANGE, K. (2004). A tutorial on MM algorithms. *Amer*. *Statist*. **58** 30–37. MR2055509

KIM, Y., CHOI, H. and OH, H.-S. (2008). Smoothly clipped absolute deviation on high dimensions. *J*. *Amer*. *Statist*. *Assoc*. **103** 1665–1673. MR2510294

KIM, Y. and KWON, S. (2012). Global optimality of nonconvex penalized estimators. *Biometrika* **99** 315–325. MR2931256

LANGE, K., HUNTER, D. R. and YANG, I. (2000). Optimization transfer using surrogate objective functions. *J*. *Comput*. *Graph*. *Statist*. **9** 1–59. MR1819865

LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann*. *Statist*. **45** 866–896. MR3650403

LOH, P.-L. and WAINWRIGHT, M. J. (2014). Support recovery without incoherence: A case for nonconvex regularization. To appear. Available at arXiv:1412.5632.

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

LOZANO, A. C. and MEINSHAUSEN, N. (2013). Minimum distance estimation for robust high-dimensional regression. Available at arXiv:1307.3227.

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

NESTEROV, Y. (2013). Gradient methods for minimizing composite functions. *Math*. *Program*. **140** 125–161. MR3071865

RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J*. *Mach*. *Learn*. *Res*. **11** 2241–2259. MR2719855

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

VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron*. *J*. *Stat*. **3** 1360–1392. MR2576316

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

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

ZHANG, T. (2009). Some sharp performance bounds for least squares regression with L_1 regularization. *Ann*. *Statist*. **37** 2109–2144. MR2543687

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

ZHANG, T. (2010b). Analysis of multi-stage convex relaxation for sparse regularization. *J. Mach. Learn. Res.* **11** 1081–1107. MR2629825

ZHANG, C.-H. and ZHANG, T. (2012). A general theory of concave regularization for high-dimensional sparse estimation problems. *Statist. Sci.* **27** 576–593. MR3025135

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443

The Annals of Statistics
2018, Vol. 46, No. 2, 842–865
https://doi.org/10.1214/17-AOS1569

ON BAYESIAN INDEX POLICIES FOR SEQUENTIAL RESOURCE ALLOCATION

BY EMILIE KAUFMANN[1]

CNRS, Univ. Lille, Inria SequeL, UMR 9189—CRIStAL—Centre de Recherche en Informatique Signal et Automatique de Lille

This paper is about index policies for minimizing (frequentist) regret in a stochastic multi-armed bandit model, inspired by a Bayesian view on the problem. Our main contribution is to prove that the Bayes-UCB algorithm, which relies on quantiles of posterior distributions, is asymptotically optimal when the reward distributions belong to a one-dimensional exponential family, for a large class of prior distributions. We also show that the Bayesian literature gives new insight on what kind of exploration rates could be used in frequentist, UCB-type algorithms. Indeed, approximations of the Bayesian optimal solution or the Finite-Horizon Gittins indices provide a justification for the kl-UCB$^+$ and kl-UCB-H$^+$ algorithms, whose asymptotic optimality is also established.

REFERENCES

[1] AGRAWAL, R. (1995). Sample mean based index policies with $O(\log n)$ regret for the multi-armed bandit problem. *Adv. in Appl. Probab.* **27** 1054–1078. MR1358906

[2] AGRAWAL, S. and GOYAL, N. (2012). Analysis of Thompson sampling for the multi-armed bandit problem. In *Proceedings of the* 25*th Conference on Learning Theory*.

[3] AGRAWAL, S. and GOYAL, N. (2013). Further optimal regret bounds for Thompson sampling. In *Proceedings of the* 16*th Conference on Artificial Intelligence and Statistics*.

[4] AUDIBERT, J.-Y. and BUBECK, S. (2010). Regret bounds and minimax policies under partial monitoring. *J. Mach. Learn. Res.* **11** 2785–2836. MR2738783

[5] AUDIBERT, J.-Y., MUNOS, R. and SZEPESVÁRI, C. (2009). Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. *Theoret. Comput. Sci.* **410** 1876–1902. MR2514714

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

[7] BELLMAN, R. (1956). A problem in the sequential design of experiments. *Sankhyā* **16** 221–229. MR0079386

[8] BERRY, D. A. and FRISTEDT, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*. Chapman & Hall, London. MR0813698

[9] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193

[10] BRADT, R. N., JOHNSON, S. M. and KARLIN, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Stat.* **27** 1060–1074. MR0087288

MSC2010 subject classifications. 62L05.
Key words and phrases. Multi-armed bandit problems, Bayesian methods, upper-confidence bounds, Gittins indices.

[11] BROCHU, E., CORA, V. M. and DE FREITAS, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. Technical report, Univ. British Columbia.

[12] BUBECK, S. and LIU, C.-Y. (2013). Prior-free and prior-dependent regret bounds for Thompson sampling. In *Advances in Neural Information Processing Systems*.

[13] BURNETAS, A. N. and KATEHAKIS, M. N. (2003). Asymptotic Bayes analysis for the finite-horizon one-armed-bandit problem. *Probab. Engrg. Inform. Sci.* **17** 53–82. MR1959385

[14] CAPPÉ, O., GARIVIER, A., MAILLARD, O.-A., MUNOS, R. and STOLTZ, G. (2013). Kullback–Leibler upper confidence bounds for optimal sequential allocation. *Ann. Statist.* **41** 1516–1541. MR3113820

[15] CHANG, F. and LAI, T. L. (1987). Optimal stopping and dynamic allocation. *Adv. in Appl. Probab.* **19** 829–853. MR0914595

[16] CHAPELLE, O. and LI, L. (2011). An empirical evaluation of Thompson sampling. In *Advances in Neural Information Processing Systems*.

[17] GARIVIER, A. and CAPPÉ, O. (2011). The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the* 24*th Conference on Learning Theory*.

[18] GINEBRA, J. and CLAYTON, M. K. (1999). Small-sample performance of Bernoulli two-armed bandit Bayesian strategies. *J. Statist. Plann. Inference* **79** 107–122. MR1704187

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

[20] GITTINS, J., GLAZEBROOK, K. and WEBER, R. (2011). *Multi-Armed Bandit Allocation Indices*, 2nd ed. Wiley, Chichester.

[21] HONDA, J. and TAKEMURA, A. (2010). An asymptotically optimal bandit algorithm for bounded support models. In *Proceedings of the* 23*rd Conference on Learning Theory*.

[22] HONDA, J. and TAKEMURA, A. (2014). Optimality of Thompson sampling for Gaussian bandits depends on priors. In *Proceedings of the* 17*th Conference on Artificial Intelligence and Statistics*.

[23] KAUFMANN, E. (2018). Supplement to "On Bayesian index policies for sequential resource allocation." DOI:10.1214/17-AOS1569SUPP.

[24] KAUFMANN, E., CAPPÉ, O. and GARIVIER, A. (2012). On Bayesian upper-confidence bounds for bandit problems. In *Proceedings of the* 15*th Conference on Artificial Intelligence and Statistics*.

[25] KAUFMANN, E., KORDA, N. and MUNOS, R. (2012). Thompson sampling: An asymptotically optimal finite-time analysis. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **7568** 199–213. Springer, Heidelberg. MR3042891

[26] KORDA, N., KAUFMANN, E. and MUNOS, R. (2013). Thompson sampling for 1-dimensional exponential family bandits. In *Advances in Neural Information Processing Systems*.

[27] LAI, T. L. (1987). Adaptive treatment allocation and the multi-armed bandit problem. *Ann. Statist.* **15** 1091–1114. MR0902248

[28] LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22. MR0776826

[29] LATTIMORE, T. (2016). Regret analysis of the finite-horizon Gittins index strategy for multi-armed bandits. In *Proceedings of the* 29*th Conference on Learning Theory*, *COLT* 2016 1214–1245. Available at http://jmlr.org/proceedings/papers/v49/lattimore16.html.

[30] LIU, C.-Y. and LI, L. (2016). On the prior sensitivity of Thompson sampling. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **9925** 321–336. Springer, Cham. MR3591000

[31] NIÑO-MORA, J. (2011). Computing a classic index for finite-horizon bandits. *INFORMS J. Comput.* **23** 254–267. MR2816898

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

[33] REVERDY, P., SRIVASTAVA, V. and LEONARD, N. E. (2014). Modeling human decision making in generalized Gaussian multiarmed bandits. *Proc. IEEE* **102** 544–571.

[34] RUSSO, D. and VAN ROY, B. (2014). Learning to optimize via posterior sampling. *Math. Oper. Res.* **39** 1221–1243.

[35] RUSSO, D. and VAN ROY, B. (2014). Learning to optimize via information direct sampling. In *Advances in Neural Information Processing Systems*.

[36] SCOTT, S. L. (2010). A modern Bayesian look at the multi-armed bandit. *Appl. Stoch. Models Bus. Ind.* **26** 639–658. MR2752378

[37] SRINIVAS, N., KRAUSE, A., KAKADE, S. M. and SEEGER, M. W. (2010). Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the International Conference on Machine Learning*.

[38] THOMPSON, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25** 285–294.

The Annals of Statistics
2018, Vol. 46, No. 2, 866–894
https://doi.org/10.1214/17-AOS1571

TESTING INDEPENDENCE WITH HIGH-DIMENSIONAL CORRELATED SAMPLES

BY XI CHEN AND WEIDONG LIU[1]

New York University and Shanghai Jiao Tong University

Testing independence among a number of (ultra) high-dimensional random samples is a fundamental and challenging problem. By arranging n identically distributed p-dimensional random vectors into a $p \times n$ data matrix, we investigate the problem of testing independence among columns under the matrix-variate normal modeling of data. We propose a computationally simple and tuning-free test statistic, characterize its limiting null distribution, analyze the statistical power and prove its minimax optimality. As an important by-product of the test statistic, a ratio-consistent estimator for the quadratic functional of a covariance matrix from correlated samples is developed. We further study the effect of correlation among samples to an important high-dimensional inference problem—large-scale multiple testing of Pearson's correlation coefficients. Indeed, blindly using classical inference results based on the assumed independence of samples will lead to many false discoveries, which suggests the need for conducting independence testing before applying existing methods. To address the challenge arising from correlation among samples, we propose a "sandwich estimator" of Pearson's correlation coefficient by de-correlating the samples. Based on this approach, the resulting multiple testing procedure asymptotically controls the overall false discovery rate at the nominal level while maintaining good statistical power. Both simulated and real data experiments are carried out to demonstrate the advantages of the proposed methods.

REFERENCES

ALLEN, G. I. and TIBSHIRANI, R. (2012). Inference with transposable data: Modelling the effects of row and column correlations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 721–743. MR2965957

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley-Interscience, Hoboken, NJ. MR1990662

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

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

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300. MR1325392

BIEN, J. and TIBSHIRANI, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika* **98** 807–820. MR2860325

MSC2010 subject classifications. Primary 62F05; secondary 62H10.

Key words and phrases. Independence test, multiple testing of correlations, false discovery rate, matrix-variate normal, quadratic functional estimation, high-dimensional sample correlation matrix.

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210

CAI, T. T. and JIANG, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *J. Multivariate Anal.* **107** 24–39. MR2890430

CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949

CAI, T. T. and LIU, W. (2016). Large-scale multiple testing of correlations. *J. Amer. Statist. Assoc.* **111** 229–240. MR3494655

CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973

CAI, T., LIU, W. and ZHOU, H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **44** 455–488. MR3476606

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

CARTER, S. L., BRECHBÜHLER, C. M., GRIFFIN, M. and BOND, A. T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20** 2242–2250.

CHEN, X. and LIU, W. (2018). Supplement to "Testing independence with high-dimensional correlated samples." DOI:10.1214/17-AOS1571SUPP.

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

DAWID, A. P. (1977). Spherical matrix distributions and a multivariate model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **39** 254–261. MR0518916

DAWID, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68** 265–274. MR0614963

EFRON, B. (2009). Are a set of microarrays independent of each other? *Ann. Appl. Stat.* **3** 922–942. MR2750220

FAN, J., RIGOLLET, P. and WANG, W. (2015). Estimation of functionals of sparse covariance matrices. *Ann. Statist.* **43** 2706–2737. MR3405609

FANG, K. T. and ZHANG, Y. T. (1990). *Generalized Multivariate Analysis*. Springer, Berlin. MR1079542

HAN, F., CHEN, S. Z. and LIU, H. (2017). Distribution-free tests of independence with applications to testing more structures. *Biometrika* **104** 813–828.

HIRAI, M. Y., SUGIYAMA, K., SAWADA, Y., TOHGE, T., OBAYASHI, T., SUZUKI, A., ARAKI, R., SAKURAI, N., SUZUKI, H., AOKI, K., GODA, H., NISHIZAWA, O. I., SHIBATA, D. and SAITO, K. (2007). Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* **104** 6478–6483.

HONG, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 429–453. MR1616065

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906

JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857

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

KIM, K., JIANG, K., TENG, S. L., FELDMAN, L. J. and HUANG, H. (2012). Using biologically interrelated experiments to identify pathway genes in Arabidopsis. *Bioinformatics* **28** 815–822.

LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459

LAZZERONI, L. and OWEN, A. (2002). Plaid models for gene expression data. *Statist. Sinica* **12** 61–86. MR1894189

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

LEE, H. K., HSU, A. K., SAJDAK, J., QIN, J. and PAVLIDIS, P. (2004). Coexpression analysis of human genes across many microarray data sets. *Genome Res.* **14** 1085–1094.

LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453

LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539

LIU, W. and SHAO, Q.-M. (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *Ann. Statist.* **42** 2003–2025. MR3262475

MURALIDHARAN, O. (2010). Detecting column dependence when rows are correlated and estimating the strength of the row correlation. *Electron. J. Stat.* **4** 1527–1546. MR2747132

NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

PAN, G., GAO, J. and YANG, Y. (2014). Testing independence among a large number of high-dimensional random vectors. *J. Amer. Statist. Assoc.* **109** 600–612. MR3223736

ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* **104** 177–186. MR2504372

SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. MR2234197

SHAW, P., GREENSTEIN, D., LERCH, J., CLASEN, L., LENROOT, R., GOGTAY, N., EVANS, A., RAPOPORT, J. and GIEDD, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature* **440** 676–679.

TENG, S. L. and HUANG, H. (2009). A statistical framework to inter functional gene relationships from biologically interrelated microarray experiments. *J. Amer. Statist. Assoc.* **104** 465–473. MR2751431

YIN, J. and LI, H. (2012). Model selection and estimation in the matrix normal graphical model. *J. Multivariate Anal.* **107** 119–140. MR2890437

ZHOU, S. (2014). Gemini: Graph estimation with matrix variate normal instances. *Ann. Statist.* **42** 532–562. MR3210978

ZHU, D., HERO, A. O., QIN, Z. S. and SWAROOP, A. (2005). High throughput screening of co-expressed gene pairs with controlled false discovery rate (FDR) and minimum acceptable strength (MAS). *J. Comput. Biol.* **12** 1029–1045.

The Annals of Statistics
2018, Vol. 46, No. 2, 895–923
https://doi.org/10.1214/17-AOS1574

DETECTING RARE AND FAINT SIGNALS VIA THRESHOLDING MAXIMUM LIKELIHOOD ESTIMATORS[1]

BY YUMOU QIU*, SONG XI CHEN† AND DAN NETTLETON‡

University of Nebraska–Lincoln, Peking University† and Iowa State University‡*

Motivated by the analysis of RNA sequencing (RNA-seq) data for genes differentially expressed across multiple conditions, we consider detecting rare and faint signals in high-dimensional response variables. We address the signal detection problem under a general framework, which includes generalized linear models for count-valued responses as special cases. We propose a test statistic that carries out a multi-level thresholding on maximum likelihood estimators (MLEs) of the signals, based on a new Cramér-type moderate deviation result for multidimensional MLEs. Based on the multi-level thresholding test, a multiple testing procedure is proposed for signal identification. Numerical simulations and a case study on maize RNA-seq data are conducted to demonstrate the effectiveness of the proposed approaches on signal detection and identification.

REFERENCES

ANDERS, S. and HUBER, W. (2010). Differential expression analysis for sequence count data. *Genome Biol*. **11** R106.

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist*. **39** 2533–2556. MR2906877

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **57** 289–300. MR1325392

BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist*. **38** 808–835. MR2604697

DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **73** 283–301. MR2815777

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist*. **32** 962–994. MR2065195

FAN, Y., JIN, J. and YAO, Z. (2013). Optimal classification in sparse Gaussian graphic model. *Ann. Statist*. **41** 2537–2571. MR3161437

FAN, J. and SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist*. **38** 3567–3604. MR2766861

GENOVESE, C. R. and WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc*. **101** 1408–1417. MR2279468

MSC2010 subject classifications. Primary 62H15; secondary 62G20, 62G32.

Key words and phrases. Detection boundary, false discovery proportion, generalized linear model, moderate deviation, multiple testing procedure, RNA-seq data.

GOEMAN, J. J., VAN HOUWELINGEN, H. C. and FINOS, L. (2011). Testing against a high-dimensional alternative in the generalized linear model: Asymptotic type I error control. *Biometrika* **98** 381–390. MR2806435

GUO, B. and CHEN, S. X. (2016). Tests for high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1079–1102. MR3557190

HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

INGLOT, T. and KALLENBERG, W. C. M. (2003). Moderate deviations of minimum contrast estimators under contamination. *Ann. Statist.* **31** 852–879. MR1994733

INGSTER, YU. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

JENSEN, J. L. and WOOD, A. T. A. (1998). Large deviation and other results for minimum contrast estimators. *Ann. Inst. Statist. Math.* **50** 673–695. MR1671986

JI, P. and JIN, J. (2012). UPS delivers optimal phase diagram in high-dimensional variable selection. *Ann. Statist.* **40** 73–103. MR3013180

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York. MR0107933

LUND, S. P., NETTLETON, D., MCCARTHY, D. J. and SMYTH, G. K. (2012). Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat. Appl. Genet. Mol. Biol.* **11** Art. 8. MR2997311

MCCULLOCH, C. E., SEARLE, S. R. and NEUHAUS, J. M. (2008). *Generalized, Linear, and Mixed Models*, 2nd ed. Wiley, Hoboken, NJ. MR2431553

PASCHOLD, A., LARSON, N. B., MARCON, C., SCHNABLE, J. C., YEH, C. T., LANZ, C., NETTLETON, D., PIEPHO, H.-P., SCHNABLE, P. S. and HOCHHOLDINGER, F. (2017). Non-syntenic genes frive highly dynamic complementation of gene expression in maize hybrids. *Plant Cell*. To appear.

PETROV, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Clarendon Press, London.

QIU, Y., CHEN, S. X and NETTLETON, D. (2018). Supplement to "Detecting rare and faint signals via thresholding maximum likelihood estimators." DOI:10.1214/17-AOS1574SUPP.

ROBINSON, M. D. and SMYTH, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23** 2881–2887.

ROBINSON, M. D. and SMYTH, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9** 321–332.

SAULIS, L. and STATULEVIČIUS, V. A. (1991). *Limit Theorems for Large Deviations. Mathematics and Its Applications (Soviet Series)* **73**. Kluwer Academic, Dordrecht. MR1171883

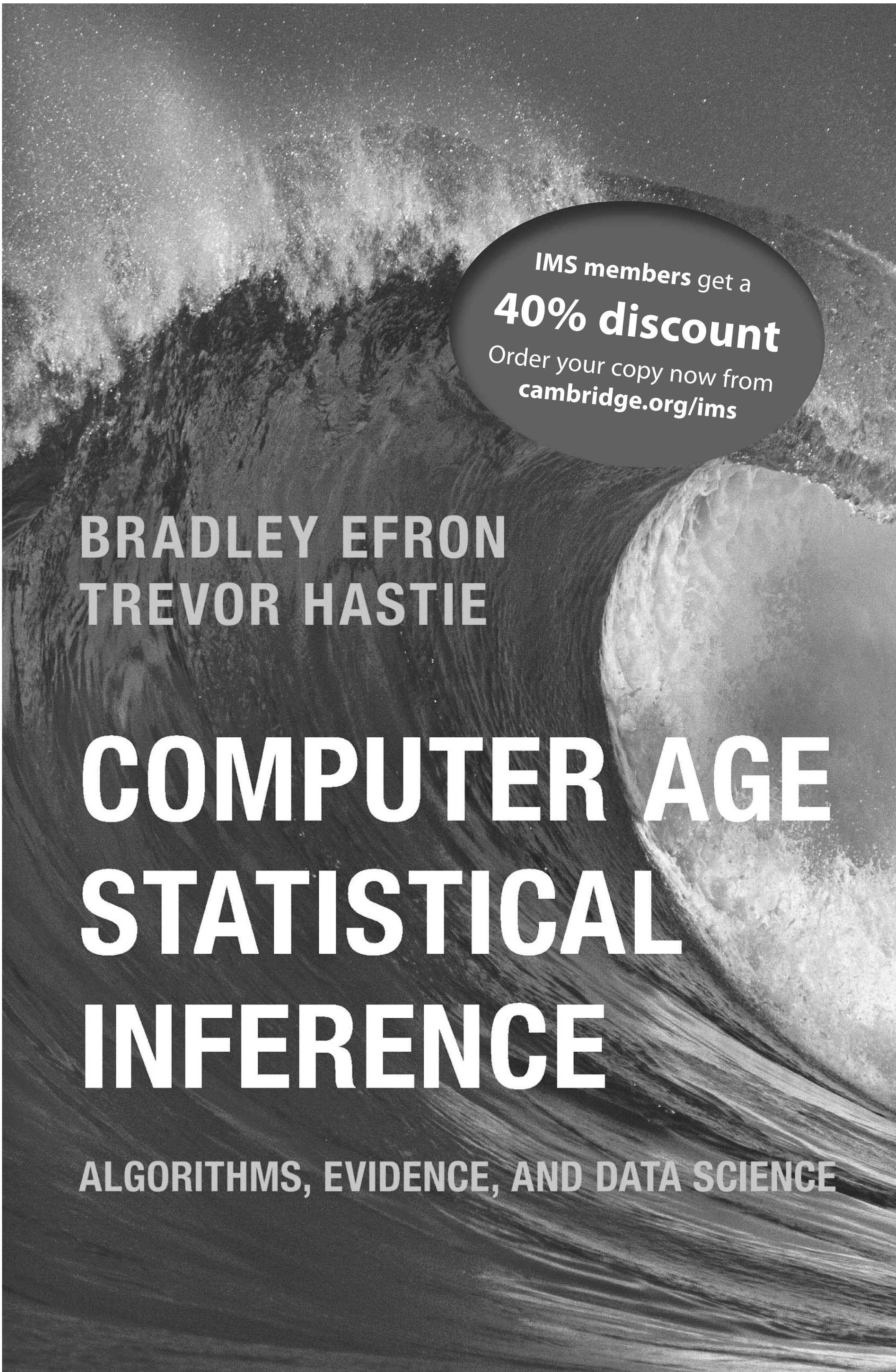
VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285

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

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

ZHONG, P.-S. and CHEN, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260–274. MR2816719

ZHONG, P.-S., CHEN, S. X. and XU, M. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann. Statist.* **41** 2820–2851. MR3161449


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE