

---

## References

- Adamczak, R. 2008. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability*, **34**, 1000–1034.
- Adamczak, R., Litvak, A. E., Pajor, A., and Tomczak-Jaegermann, N. 2010. Quantitative estimations of the convergence of the empirical covariance matrix in log-concave ensembles. *Journal of the American Mathematical Society*, **23**, 535–561.
- Agarwal, A., Negahban, S., and Wainwright, M. J. 2012. Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Annals of Statistics*, **40**(2), 1171–1197.
- Ahlsvede, R., and Winter, A. 2002. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, **48**(3), 569–579.
- Aizerman, M. A., Braverman, E. M., and Rozonoer, L. I. 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, **25**, 821–837.
- Akçakaya, M., and Tarokh, V. 2010. Shannon theoretic limits on noisy compressive sampling. *IEEE Transactions on Information Theory*, **56**(1), 492–504.
- Alexander, K. S. 1987. Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probability Theory and Related Fields*, **75**, 379–423.
- Alliney, S., and Ruzinsky, S. A. 1994. An algorithm for the minimization of mixed  $\ell_1$  and  $\ell_2$  norms with application to Bayesian estimation. *IEEE Transactions on Signal Processing*, **42**(3), 618–627.
- Amini, A. A., and Wainwright, M. J. 2009. High-dimensional analysis of semidefinite relaxations for sparse principal component analysis. *Annals of Statistics*, **5B**, 2877–2921.
- Anandkumar, A., Tan, V. Y. F., Huang, F., and Willsky, A. S. 2012. High-dimensional structure learning of Ising models: Local separation criterion. *Annals of Statistics*, **40**(3), 1346–1375.
- Anderson, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Mathematical Statistics. New York, NY: Wiley.
- Ando, R. K., and Zhang, T. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, **6**(December), 1817–1853.
- Aronszajn, N. 1950. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68**, 337–404.
- Assouad, P. 1983. Deux remarques sur l'estimation. *Comptes Rendus de l'Académie des Sciences, Paris*, **296**, 1021–1024.
- Azuma, K. 1967. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, **19**, 357–367.
- Bach, F., Jenatton, R., Mairal, J., and Obozinski, G. 2012. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning*, **4**(1), 1–106.
- Bahadur, R. R., and Rao, R. R. 1960. On deviations of the sample mean. *Annals of Mathematical Statistics*, **31**, 1015–1027.
- Bai, Z., and Silverstein, J. W. 2010. *Spectral Analysis of Large Dimensional Random Matrices*. New York, NY: Springer. Second edition.
- Baik, J., and Silverstein, J. W. 2006. Eigenvalues of large sample covariance matrices of spiked populations models. *Journal of Multivariate Analysis*, **97**(6), 1382–1408.
- Balabdaoui, F., Rufibach, K., and Wellner, J. A. 2009. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of Statistics*, **62**(3), 1299–1331.

- Ball, K. 1997. An elementary introduction to modern convex geometry. Pages 1–55 of: *Flavors of Geometry*. MSRI Publications, vol. 31. Cambridge, UK: Cambridge University Press.
- Banerjee, O., El Ghaoui, L., and d'Aspremont, A. 2008. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research*, **9**(March), 485–516.
- Baraniuk, R. G., Cevher, V., Duarte, M. F., and Hegde, C. 2010. Model-based compressive sensing. *IEEE Transactions on Information Theory*, **56**(4), 1982–2001.
- Barndorff-Nielsen, O. E. 1978. *Information and Exponential Families*. Chichester, UK: Wiley.
- Bartlett, P. L., and Mendelson, S. 2002. Gaussian and Rademacher complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, **3**, 463–482.
- Bartlett, P. L., Bousquet, O., and Mendelson, S. 2005. Local Rademacher complexities. *Annals of Statistics*, **33**(4), 1497–1537.
- Baxter, R. J. 1982. *Exactly Solved Models in Statistical Mechanics*. New York, NY: Academic Press.
- Bean, D., Bickel, P. J., El Karoui, N., and Yu, B. 2013. Optimal M-estimation in high-dimensional regression. *Proceedings of the National Academy of Sciences of the USA*, **110**(36), 14563–14568.
- Belloni, A., Chernozhukov, V., and Wang, L. 2011. Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, **98**(4), 791–806.
- Bennett, G. 1962. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, **57**(297), 33–45.
- Bento, J., and Montanari, A. 2009 (December). Which graphical models are difficult to learn? In: *Proceedings of the NIPS Conference*.
- Berlinet, A., and Thomas-Agnan, C. 2004. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Norwell, MA: Kluwer Academic.
- Bernstein, S. N. 1937. On certain modifications of Chebyshev's inequality. *Doklady Akademii Nauk SSSR*, **16**(6), 275–277.
- Berthet, Q., and Rigollet, P. 2013 (June). Computational lower bounds for sparse PCA. In: *Conference on Computational Learning Theory*.
- Bertsekas, D. P. 2003. *Convex Analysis and Optimization*. Boston, MA: Athena Scientific.
- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, **36**, 192–236.
- Besag, J. 1975. Statistical analysis of non-lattice data. *The Statistician*, **24**(3), 179–195.
- Besag, J. 1977. Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika*, **64**(3), 616–618.
- Bethe, H. A. 1935. Statistics theory of superlattices. *Proceedings of the Royal Society of London, Series A*, **150**(871), 552–575.
- Bhatia, R. 1997. *Matrix Analysis*. Graduate Texts in Mathematics. New York, NY: Springer.
- Bickel, P. J., and Doksum, K. A. 2015. *Mathematical Statistics: Basic Ideas and Selected Topics*. Boca Raton, FL: CRC Press.
- Bickel, P. J., and Levina, E. 2008a. Covariance regularization by thresholding. *Annals of Statistics*, **36**(6), 2577–2604.
- Bickel, P. J., and Levina, E. 2008b. Regularized estimation of large covariance matrices. *Annals of Statistics*, **36**(1), 199–227.
- Bickel, P. J., Ritov, Y., and Tsybakov, A. B. 2009. Simultaneous analysis of lasso and Dantzig selector. *Annals of Statistics*, **37**(4), 1705–1732.
- Birgé, L. 1983. Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. verw. Gebiete*, **65**, 181–327.
- Birgé, L. 1987. Estimating a density under order restrictions: Non-asymptotic minimax risk. *Annals of Statistics*, **15**(3), 995–1012.
- Birgé, L. 2005. A new lower bound for multiple hypothesis testing. *IEEE Transactions on Information Theory*, **51**(4), 1611–1614.
- Birgé, L., and Massart, P. 1995. Estimation of integral functionals of a density. *Annals of Statistics*, **23**(1), 11–29.

- Birnbaum, A., Johnstone, I. M., Nadler, B., and Paul, D. 2012. Minimax bounds for sparse PCA with noisy high-dimensional data. *Annals of Statistics*, **41**(3), 1055–1084.
- Bobkov, S. G. 1999. Isoperimetric and analytic inequalities for log-concave probability measures. *Annals of Probability*, **27**(4), 1903–1921.
- Bobkov, S. G., and Götze, F. 1999. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis*, **163**, 1–28.
- Bobkov, S. G., and Ledoux, M. 2000. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geometric and Functional Analysis*, **10**, 1028–1052.
- Borgwardt, K., Gretton, A., Rasch, M., Kriegel, H. P., Schölkopf, B., and Smola, A. J. 2006. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, **22**(14), 49–57.
- Borwein, J., and Lewis, A. 1999. *Convex Analysis*. New York, NY: Springer.
- Boser, B. E., Guyon, I. M., and Vapnik, V. N. 1992. A training algorithm for optimal margin classifiers. Pages 144–152 of: *Proceedings of the Conference on Learning Theory (COLT)*. New York, NY: ACM.
- Boucheron, S., Lugosi, G., and Massart, P. 2003. Concentration inequalities using the entropy method. *Annals of Probability*, **31**(3), 1583–1614.
- Boucheron, S., Lugosi, G., and Massart, P. 2013. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford, UK: Oxford University Press.
- Bourgain, J., Dirksen, S., and Nelson, J. 2015. Toward a unified theory of sparse dimensionality reduction in Euclidean space. *Geometric and Functional Analysis*, **25**(4).
- Bousquet, O. 2002. A Bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus de l'Académie des Sciences, Paris, Série I*, **334**, 495–500.
- Bousquet, O. 2003. Concentration inequalities for sub-additive functions using the entropy method. *Stochastic Inequalities and Applications*, **56**, 213–247.
- Boyd, S., and Vandenberghe, L. 2004. *Convex optimization*. Cambridge, UK: Cambridge University Press.
- Brascamp, H. J., and Lieb, E. H. 1976. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, **22**, 366–389.
- Breiman, L. 1992. *Probability*. Classics in Applied Mathematics. Philadelphia, PA: S IAM.
- Bresler, G. 2014. *Efficiently learning Ising models on arbitrary graphs*. Tech. rept. MIT.
- Bresler, G., Mossel, E., and Sly, A. 2013. Reconstruction of Markov Random Fields from samples: Some observations and algorithms. *SIAM Journal on Computing*, **42**(2), 563–578.
- Bronstein, E. M. 1976.  $\epsilon$ -entropy of convex sets and functions. *Siberian Mathematical Journal*, **17**, 393–398.
- Brown, L. D. 1986. *Fundamentals of statistical exponential families*. Hayward, CA: Institute of Mathematical Statistics.
- Brunk, H. D. 1955. Maximum likelihood estimates of monotone parameters. *Annals of Math. Statistics*, **26**, 607–616.
- Brunk, H. D. 1970. Estimation of isotonic regression. Pages 177–197 of: *Nonparametric techniques in statistical inference*. New York, NY: Cambridge University Press.
- Bühlmann, P., and van de Geer, S. 2011. *Statistics for high-dimensional data*. Springer Series in Statistics. Springer.
- Buja, A., Hastie, T. J., and Tibshirani, R. 1989. Linear smoothers and additive models. *Annals of Statistics*, **17**(2), 453–510.
- Buldygin, V. V., and Kozachenko, Y. V. 2000. *Metric characterization of random variables and random processes*. Providence, RI: American Mathematical Society.
- Bunea, F., Tsybakov, A. B., and Wegkamp, M. 2007. Sparsity oracle inequalities for the Lasso. *Electronic Journal of Statistics*, 169–194.
- Bunea, F., She, Y., and Wegkamp, M. 2011. Optimal selection of reduced rank estimators of high-dimensional matrices. *Annals of Statistics*, **39**(2), 1282–1309.
- Cai, T. T., Zhang, C. H., and Zhou, H. H. 2010. Optimal rates of convergence for covariance matrix estimation. *Annals of Statistics*, **38**(4), 2118–2144.
- Cai, T. T., Liu, W., and Luo, X. 2011. A constrained  $\ell_1$ -minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **106**, 594–607.

- Cai, T. T., Liang, T., and Rakhlin, A. 2015. *Computational and statistical boundaries for submatrix localization in a large noisy matrix*. Tech. rept. Univ. Penn.
- Candès, E. J., and Plan, Y. 2010. Matrix completion with noise. *Proceedings of the IEEE*, **98**(6), 925–936.
- Candès, E. J., and Recht, B. 2009. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, **9**(6), 717–772.
- Candès, E. J., and Tao, T. 2005. Decoding by linear programming. *IEEE Transactions on Information Theory*, **51**(12), 4203–4215.
- Candès, E. J., and Tao, T. 2007. The Dantzig selector: statistical estimation when  $p$  is much larger than  $n$ . *Annals of Statistics*, **35**(6), 2313–2351.
- Candès, E. J., Li, X., Ma, Y., and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM*, **58**(3), 11 (37pp).
- Candès, E. J., Strohmer, T., and Voroninski, V. 2013. PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, **66**(8), 1241–1274.
- Cantelli, F. P. 1933. Sulla determinazione empirica della legge di probabilita. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 421–424.
- Carl, B., and Pajor, A. 1988. Gelfand numbers of operators with values in a Hilbert space. *Inventiones Mathematicae*, **94**, 479–504.
- Carl, B., and Stephani, I. 1990. *Entropy, Compactness and the Approximation of Operators*. Cambridge Tracts in Mathematics. Cambridge, UK: Cambridge University Press.
- Carlen, E. 2009. Trace inequalities and quantum entropy: an introductory course. In: *Entropy and the Quantum*. Providence, RI: American Mathematical Society.
- Carroll, R. J., Ruppert, D., and Stefanski, L. A. 1995. *Measurement Error in Nonlinear Models*. Boca Raton, FL: Chapman & Hall/CRC.
- Chai, A., Moscoso, M., and Papanicolaou, G. 2011. Array imaging using intensity-only measurements. *Inverse Problems*, **27**(1), 1—15.
- Chandrasekaran, V., Sanghavi, S., Parrilo, P. A., and Willsky, A. S. 2011. Rank-Sparsity Incoherence for Matrix Decomposition. *SIAM Journal on Optimization*, **21**, 572–596.
- Chandrasekaran, V., Recht, B., Parrilo, P. A., and Willsky, A. S. 2012a. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, **12**(6), 805–849.
- Chandrasekaran, V., Parrilo, P. A., and Willsky, A. S. 2012b. Latent variable graphical model selection via convex optimization. *Annals of Statistics*, **40**(4), 1935–1967.
- Chatterjee, S. 2005 (October). *An error bound in the Sudakov-Fernique inequality*. Tech. rept. UC Berkeley. arXiv:math.PR/0510424.
- Chatterjee, S. 2007. Stein's method for concentration inequalities. *Probability Theory and Related Fields*, **138**(1–2), 305–321.
- Chatterjee, S., Guntuboyina, A., and Sen, B. 2015. On risk bounds in isotonic and other shape restricted regression problems. *Annals of Statistics*, **43**(4), 1774–1800.
- Chen, S., Donoho, D. L., and Saunders, M. A. 1998. Atomic decomposition by basis pursuit. *SIAM J. Sci. Computing*, **20**(1), 33–61.
- Chernoff, H. 1952. A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Annals of Mathematical Statistics*, **23**, 493–507.
- Chernozhukov, V., Chetverikov, D., and Kato, K. 2013. *Comparison and anti-concentration bounds for maxima of Gaussian random vectors*. Tech. rept. MIT.
- Chung, F.R.K. 1991. *Spectral Graph Theory*. Providence, RI: American Mathematical Society.
- Clifford, P. 1990. Markov random fields in statistics. In: Grimmett, G.R., and Welsh, D. J. A. (eds), *Disorder in physical systems*. Oxford Science Publications.
- Cohen, A., Dahmen, W., and DeVore, R. A. 2008. Compressed sensing and best  $k$ -term approximation. *J. of American Mathematical Society*, **22**(1), 211–231.
- Cormode, G. 2012. Synopses for massive data: Samples, histograms, wavelets and sketches. *Foundations and Trends in Databases*, **4**(2), 1–294.
- Cover, T.M., and Thomas, J.A. 1991. *Elements of Information Theory*. New York, NY: Wiley.

- Cule, M., Samworth, R. J., and Stewart, M. 2010. Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. B*, **62**, 545–607.
- Dalalyan, A. S., Hebiri, M., and Lederer, J. 2014. *On the prediction performance of the Lasso*. Tech. rept. ENSAE. arxiv:1402.1700, to appear in Bernoulli.
- d'Aspremont, A., El Ghaoui, L., Jordan, M. I., and Lanckriet, G. R. 2007. A direct formulation for sparse PCA using semidefinite programming. *SIAM Review*, **49**(3), 434–448.
- d'Aspremont, A., Banerjee, O., and El Ghaoui, L. 2008. First order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Its Applications*, **30**(1), 55–66.
- Davidson, K. R., and Szarek, S. J. 2001. Local operator theory, random matrices, and Banach spaces. Pages 317–336 of: *Handbook of Banach Spaces*, vol. 1. Amsterdam, NL: Elsevier.
- Dawid, A. P. 2007. The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics*, **59**, 77–93.
- de La Pena, V., and Giné, E. 1999. *Decoupling: From dependence to independence*. New York, NY: Springer.
- Dembo, A. 1997. Information inequalities and concentration of measure. *Annals of Probability*, **25**(2), 927–939.
- Dembo, A., and Zeitouni, O. 1996. Transportation approach to some concentration inequalities in product spaces. *Electronic Communications in Probability*, **1**, 83–90.
- DeVore, R. A., and Lorentz, G. G. 1993. *Constructive Approximation*. New York, NY: Springer.
- Devroye, L., and Györfi, L. 1986. *Nonparametric density estimation: the  $L_1$  view*. New York, NY: Wiley.
- Donoho, D. L. 2006a. For most large underdetermined systems of linear equations, the minimal  $\ell_1$ -norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*, **59**(7), 907–934.
- Donoho, D. L. 2006b. For most large underdetermined systems of linear equations, the minimal  $\ell_1$ -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, **59**(6), 797–829.
- Donoho, D. L., and Huo, X. 2001. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, **47**(7), 2845–2862.
- Donoho, D. L., and Johnstone, I. M. 1994. Minimax risk over  $\ell_p$ -balls for  $\ell_q$ -error. *Probability Theory and Related Fields*, **99**, 277–303.
- Donoho, D. L., and Montanari, A. 2013. *High dimensional robust M-estimation: asymptotic variance via approximate message passing*. Tech. rept. Stanford University. Posted as arxiv:1310.7320.
- Donoho, D. L., and Stark, P. B. 1989. Uncertainty principles and signal recovery. *SIAM Journal of Applied Mathematics*, **49**, 906–931.
- Donoho, D. L., and Tanner, J. M. 2008. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, July.
- Duchi, J. C., Wainwright, M. J., and Jordan, M. I. 2013. *Local privacy and minimax bounds: Sharp rates for probability estimation*. Tech. rept. UC Berkeley.
- Duchi, J. C., Wainwright, M. J., and Jordan, M. I. 2014. Privacy-aware learning. *Journal of the ACM*, **61**(6), Article 37.
- Dudley, R. M. 1967. The sizes of compact subsets of Hilbert spaces and continuity of Gaussian processes. *Journal of Functional Analysis*, **1**, 290–330.
- Dudley, R. M. 1978. Central limit theorems for empirical measures. *Annals of Probability*, **6**, 899–929.
- Dudley, R. M. 1999. *Uniform central limit theorems*. Cambridge, UK: Cambridge University Press.
- Dümbgen, L., Samworth, R. J., and Schuhmacher, D. 2011. Approximation by log-concave distributions with applications to regression. *Annals of Statistics*, **39**(2), 702–730.
- Durrett, R. 2010. *Probability: Theory and examples*. Cambridge, UK: Cambridge University Press.
- Dvoretzky, A., Kiefer, J., and Wolfowitz, J. 1956. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, **27**, 642–669.
- Eggermont, P. P. B., and LaRiccia, V. N. 2001. *Maximum penalized likelihood estimation: V. I Density estimation*. Springer Series in Statistics, vol. 1. New York, NY: Springer.
- Eggermont, P. P. B., and LaRiccia, V. N. 2007. *Maximum penalized likelihood estimation: V. II Regression*. Springer Series in Statistics, vol. 2. New York, NY: Springer.

- El Karoui, N. 2008. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Annals of Statistics*, **36**(6), 2717–2756.
- El Karoui, N. 2013. *Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators : rigorous results*. Tech. rept. UC Berkeley. Posted as arxiv:1311.2445.
- El Karoui, N., Bean, D., Bickel, P. J., and Yu, B. 2013. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences of the USA*, **110**(36), 14557–14562.
- Elad, M., and Bruckstein, A. M. 2002. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, **48**(9), 2558–2567.
- Fan, J., and Li, R. 2001. Variable selection via non-concave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**(456), 1348–1360.
- Fan, J., and Lv, J. 2011. Nonconcave penalized likelihood with NP-dimensionality. *IEEE Transactions on Information Theory*, **57**(8), 5467–5484.
- Fan, J., Liao, Y., and Mincheva, M. 2013. Large covariance estimation by thresholding principal orthogonal components. *Journal of the Royal Statistical Society B*, **75**, 603–680.
- Fan, J., Xue, L., and Zou, H. 2014. Strong oracle optimality of folded concave penalized estimation. *Annals of Statistics*, **42**(3), 819–849.
- Fazel, M. 2002. *Matrix Rank Minimization with Applications*. Ph.D. thesis, Stanford. Available online: <http://faculty.washington.edu/mfazel/thesis-final.pdf>.
- Fernique, X. M. 1974. Des resultats nouveaux sur les processus Gaussiens. *Comptes Rendus de l'Académie des Sciences, Paris*, **278**, A363–A365.
- Feuer, A., and Nemirovski, A. 2003. On sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, **49**(6), 1579–1581.
- Fienup, J. R. 1982. Phase retrieval algorithms: a comparison. *Applied Optics*, **21**(15), 2758–2769.
- Fienup, J. R., and Wackerman, C. C. 1986. Phase-retrieval stagnation problems and solutions. *Journal of the Optical Society of America A*, **3**, 1897–1907.
- Fletcher, A. K., Rangan, S., and Goyal, V. K. 2009. Necessary and Sufficient Conditions for Sparsity Pattern Recovery. *IEEE Transactions on Information Theory*, **55**(12), 5758–5772.
- Foygel, R., and Srebro, N. 2011. *Fast rate and optimistic rate for  $\ell_1$ -regularized regression*. Tech. rept. Toyoto Technological Institute. arXiv:1108.037v1.
- Friedman, J. H., and Stuetzle, W. 1981. Projection pursuit regression. *Journal of the American Statistical Association*, **76**(376), 817–823.
- Friedman, J. H., and Tukey, J. W. 1994. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, **C-23**, 881–889.
- Friedman, J. H., Hastie, T. J., and Tibshirani, R. 2007. Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics*.
- Fuchs, J. J. 2004. Recovery of exact sparse representations in the presence of noise. Pages 533–536 of: *ICASSP*, vol. 2.
- Gallager, R. G. 1968. *Information theory and reliable communication*. New York, NY: Wiley.
- Gao, C., Ma, Z., and Zhou, H. H. 2015. *Sparse CCA: Adaptive estimation and computational barriers*. Tech. rept. Yale University.
- Gardner, R. J. 2002. The Brunn-Minkowski inequality. *Bulletin of the American Mathematical Society*, **39**, 355–405.
- Geman, S. 1980. A limit theorem for the norm of random matrices. *Annals of Probability*, **8**(2), 252–261.
- Geman, S., and Geman, D. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Geman, S., and Hwang, C. R. 1982. Nonparametric maximum likelihood estimation by the method of sieves. *Annals of Statistics*, **10**(2), 401–414.
- Glivenko, V. 1933. Sulla determinazione empirica della legge di probabilita. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 92–99.
- Gneiting, T., and Raftery, A. E. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**(477), 359–378.
- Goldberg, K., Roeder, T., Gupta, D., and Perkins, C. 2001. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, **4**(2), 133–151.

- Good, I. J., and Gaskins, R. A. 1971. Nonparametric roughness penalties for probability densities. *Biometrika*, **58**, 255–277.
- Gordon, Y. 1985. Some inequalities for Gaussian processes and applications. *Israel Journal of Mathematics*, **50**, 265–289.
- Gordon, Y. 1986. On Milman's inequality and random subspaces which escape through a mesh in  $\mathbb{R}^n$ . Pages 84–106 of: *Geometric aspects of functional analysis*. Lecture Notes in Mathematics, vol. 1317. Springer-Verlag.
- Gordon, Y. 1987. Elliptically contoured distributions. *Probability Theory and Related Fields*, **76**, 429–438.
- Götze, F., and Tikhomirov, A. 2004. Rate of convergence in probability to the Marčenko-Pastur law. *Bernoulli*, **10**(3), 503–548.
- Grechberg, R. W., and Saxton, W. O. 1972. A practical algorithm for the determination of phase from image and diffraction plane intensities. *Optik*, **35**, 237–246.
- Greenshtein, E., and Ritov, Y. 2004. Persistency in high dimensional linear predictor-selection and the virtue of over-parametrization. *Bernoulli*, **10**, 971–988.
- Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., and Smola, A. 2012. A kernel two-sample test. *Journal of Machine Learning Research*, **13**, 723–773.
- Griffin, D., and Lim, J. 1984. Signal estimation from modified short-time Fourier transforms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **32**(2), 236–243.
- Grimmett, G. R. 1973. A theorem about random fields. *Bulletin of the London Mathematical Society*, **5**, 81–84.
- Gross, D. 2011. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, **57**(3), 1548–1566.
- Gross, L. 1975. Logarithmic Sobolev inequalities. *American Journal Math.*, **97**, 1061–1083.
- Gu, C. 2002. *Smoothing spline ANOVA models*. Springer Series in Statistics. New York, NY: Springer.
- Guédon, O., and Litvak, A. E. 2000. Euclidean projections of a  $p$ -convex body. Pages 95–108 of: *Geometric aspects of functional analysis*. Springer.
- Guntuboyina, A. 2011. Lower bounds for the minimax risk using  $f$ -divergences and applications. *IEEE Transactions on Information Theory*, **57**(4), 2386–2399.
- Guntuboyina, A., and Sen, B. 2013. Covering numbers for convex functions. *IEEE Transactions on Information Theory*, **59**, 1957–1965.
- Gyorfi, L., Kohler, M., Krzyzak, A., and Walk, H. 2002. *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer.
- Hammersley, J. M., and Clifford, P. 1971. *Markov fields on finite graphs and lattices*. Unpublished.
- Hanson, D. L., and Pledger, G. 1976. Consistency in concave regression. *Annals of Statistics*, **4**, 1038–1050.
- Hanson, D. L., and Wright, F. T. 1971. A bound on tail probabilities for quadratic forms in independent random variables. *Annals of Mathematical Statistics*, **42**(3), 1079–1083.
- Härdle, W. K., and Stoker, T. M. 1989. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, **84**, 986–995.
- Härdle, W. K., Hall, P., and Ichimura, H. 1993. Optimal smoothing in single-index models. *Annals of Statistics*, **21**, 157–178.
- Härdle, W. K., Müller, M., Sperlich, S., and Werwatz, A. 2004. *Nonparametric and semiparametric models*. Springer Series in Statistics. New York, NY: Springer.
- Harper, L. H. 1966. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, **1**, 385–393.
- Harrison, R. W. 1993. Phase problem in crystallography. *Journal of the Optical Society of America A*, **10**(5), 1046–1055.
- Hasminskii, R. Z. 1978. A lower bound on the risks of nonparametric estimates of densities in the uniform metric. *Theory of Probability and Its Applications*, **23**, 794–798.
- Hasminskii, R. Z., and Ibragimov, I. 1981. *Statistical estimation: Asymptotic theory*. New York, NY: Springer.
- Hasminskii, R. Z., and Ibragimov, I. 1990. On density estimation in the view of Kolmogorov's ideas in approximation theory. *Annals of Statistics*, **18**(3), 999–1010.
- Hastie, T. J., and Tibshirani, R. 1986. Generalized additive models. *Statistical Science*, **1**(3), 297–310.

- Hastie, T. J., and Tibshirani, R. 1990. *Generalized Additive Models*. Boca Raton, FL: Chapman & Hall/CRC.
- Hildreth, C. 1954. Point estimates of ordinates of concave functions. *Journal of the American Statistical Association*, **49**, 598–619.
- Hiriart-Urruty, J., and Lemaréchal, C. 1993. *Convex Analysis and Minimization Algorithms*. Vol. 1. New York, NY: Springer.
- Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**, 13–30.
- Hoerl, A. E., and Kennard, R. W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, **12**, 55–67.
- Höfling, H., and Tibshirani, R. 2009. Estimation of sparse binary pairwise Markov networks using pseudo-likelihoods. *Journal of Machine Learning Research*, **19**, 883–906.
- Holley, R., and Stroock, D. 1987. Log Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics*, **46**(5), 1159–1194.
- Horn, R. A., and Johnson, C. R. 1985. *Matrix Analysis*. Cambridge, UK: Cambridge University Press.
- Horn, R. A., and Johnson, C. R. 1991. *Topics in Matrix Analysis*. Cambridge, UK: Cambridge University Press.
- Hristache, M., Juditsky, A., and Spokoiny, V. 2001. Direct estimation of the index coefficient in a single index model. *Annals of Statistics*, **29**, 595–623.
- Hsu, D., Kakade, S. M., and Zhang, T. 2012a. Tail inequalities for sums of random matrices that depend on the intrinsic dimension. *Electronic Communications in Probability*, **17**(14), 1–13.
- Hsu, D., Kakade, S. M., and Zhang, T. 2012b. A tail inequality for quadratic forms of sub-Gaussian random vectors. *Electronic Journal of Probability*, **52**, 1–6.
- Huang, J., and Zhang, T. 2010. The benefit of group sparsity. *Annals of Statistics*, **38**(4), 1978–2004.
- Huang, J., Ma, S., and Zhang, C. H. 2008. Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, **18**, 1603–1618.
- Huber, P. J. 1973. Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics*, **1**(5), 799–821.
- Huber, P. J. 1985. Projection pursuit. *Annals of Statistics*, **13**(2), 435–475.
- Ichimura, H. 1993. Semiparametric least squares (SLS) and weighted (SLS) estimation of single index models. *Journal of Econometrics*, **58**, 71–120.
- Ising, E. 1925. Beitrag zur Theorie der Ferromagnetismus. *Zeitschrift für Physik*, **31**(1), 253–258.
- Iturria, S. J., Carroll, R. J., and Firth, D. 1999. Polynomial Regression and Estimating Functions in the Presence of Multiplicative Measurement Error. *Journal of the Royal Statistical Society B*, **61**, 547–561.
- Izenman, A. J. 1975. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, **5**, 248–264.
- Izenman, A. J. 2008. *Modern multivariate statistical techniques: Regression, classification and manifold learning*. New York, NY: Springer.
- Jacob, L., Obozinski, G., and Vert, J. P. 2009. Group Lasso with overlap and graph Lasso. Pages 433–440 of: *International Conference on Machine Learning (ICML)*.
- Jalali, A., Ravikumar, P., Sanghavi, S., and Ruan, C. 2010. A Dirty Model for Multi-task Learning. Pages 964–972 of: *Advances in Neural Information Processing Systems 23*.
- Johnson, W. B., and Lindenstrauss, J. 1984. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, **26**, 189–206.
- Johnstone, I. M. 2001. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, **29**(2), 295–327.
- Johnstone, I. M. 2015. *Gaussian estimation: Sequence and wavelet models*. New York, NY: Springer.
- Johnstone, I. M., and Lu, A. Y. 2009. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, **104**, 682–693.
- Jolliffe, I. T. 2004. *Principal Component Analysis*. New York, NY: Springer.
- Jolliffe, I. T., Trendafilov, N. T., and Uddin, M. 2003. A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, **12**, 531–547.



- Juditsky, A., and Nemirovski, A. 2000. Functional aggregation for nonparametric regression. *Annals of Statistics*, **28**, 681–712.
- Kahane, J. P. 1986. Une inégalité du type de Slepian et Gordon sur les processus Gaussiens. *Israel Journal of Mathematics*, **55**, 109–110.
- Kalisch, M., and Bühlmann, P. 2007. Estimating high-dimensional directed acyclic graphs with the PC algorithm. *Journal of Machine Learning Research*, **8**, 613–636.
- Kane, D. M., and Nelson, J. 2014. Sparser Johnson-Lindenstrauss transforms. *Journal of the ACM*, **61**(1).
- Kantorovich, L. V., and Rubinstein, G. S. 1958. On the space of completely additive functions. *Vestnik Leningrad Univ. Ser. Math. Mekh. i. Astron.*, **13**(7), 52–59. In Russian.
- Keener, R. W. 2010. *Theoretical Statistics: Topics for a Core Class*. New York, NY: Springer.
- Keshavan, R. H., Montanari, A., and Oh, S. 2010a. Matrix Completion from Few Entries. *IEEE Transactions on Information Theory*, **56**(6), 2980–2998.
- Keshavan, R. H., Montanari, A., and Oh, S. 2010b. Matrix Completion from Noisy Entries. *Journal of Machine Learning Research*, **11**(July), 2057–2078.
- Kim, Y., Kim, J., and Kim, Y. 2006. Blockwise sparse regression. *Statistica Sinica*, **16**(2).
- Kimeldorf, G., and Wahba, G. 1971. Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**, 82–95.
- Klein, T., and Rio, E. 2005. Concentration around the mean for maxima of empirical processes. *Annals of Probability*, **33**(3), 1060–1077.
- Koller, D., and Friedman, N. 2010. *Graphical Models*. New York, NY: MIT Press.
- Kolmogorov, A. N. 1956. Asymptotic characterization of some completely bounded metric spaces. *Doklady Akademii Nauk SSSR*, **108**, 585–589.
- Kolmogorov, A. N. 1958. Linear dimension of topological vector spaces. *Doklady Akademii Nauk SSSR*, **120**, 239–241–589.
- Kolmogorov, A. N., and Tikhomirov, B. 1959.  $\epsilon$ -entropy and  $\epsilon$ -capacity of sets in functional spaces. *Uspekhi Mat. Nauk.*, **86**, 3–86. Appeared in English as 1961. *American Mathematical Society Translations*, **17**, 277–364.
- Koltchinskii, V. 2001. Rademacher penalties and structural risk minimization. *IEEE Transactions on Information Theory*, **47**(5), 1902–1914.
- Koltchinskii, V. 2006. Local Rademacher complexities and oracle inequalities in risk minimization. *Annals of Statistics*, **34**(6), 2593–2656.
- Koltchinskii, V., and Panchenko, D. 2000. Rademacher processes and bounding the risk of function learning. Pages 443–459 of: *High-dimensional probability II*. Springer.
- Koltchinskii, V., and Yuan, M. 2010. Sparsity in multiple kernel learning. *Annals of Statistics*, **38**, 3660–3695.
- Koltchinskii, V., Lounici, K., and Tsybakov, A. B. 2011. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Annals of Statistics*, **39**, 2302–2329.
- Kontorovich, L. A., and Ramanan, K. 2008. Concentration inequalities for dependent random variables via the martingale method. *Annals of Probability*, **36**(6), 2126–2158.
- Kruskal, J. B. 1969. Towards a practical method which helps uncover the structure of a set of multivariate observation by finding the linear transformation which optimizes a new ‘index of condensation’. In: *Statistical computation*. New York, NY: Academic Press.
- Kühn, T. 2001. A lower estimate for entropy numbers. *Journal of Approximation Theory*, **110**, 120–124.
- Kullback, S., and Leibler, R. A. 1951. On information and sufficiency. *Annals of Mathematical Statistics*, **22**(1), 79–86.
- Lam, C., and Fan, J. 2009. Sparsistency and Rates of Convergence in Large Covariance Matrix Estimation. *Annals of Statistics*, **37**, 4254–4278.
- Laurent, M. 2001. Matrix Completion Problems. Pages 221–229 of: *The Encyclopedia of Optimization*. Kluwer Academic.
- Laurent, M. 2003. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, **28**, 470–496.
- Lauritzen, S. L. 1996. *Graphical Models*. Oxford: Oxford University Press.
- Le Cam, L. 1973. Convergence of estimates under dimensionality restrictions. *Annals of Statistics*, January.

- Ledoux, M. 1996. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, **1**(July), 63–87.
- Ledoux, M. 2001. *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society.
- Ledoux, M., and Talagrand, M. 1991. *Probability in Banach Spaces: Isoperimetry and Processes*. New York, NY: Springer.
- Lee, J. D., Sun, Y., and Taylor, J. 2013. *On model selection consistency of M-estimators with geometrically decomposable penalties*. Tech. rept. Stanford University. arxiv1305.7477v4.
- Leindler, L. 1972. On a certain converse of Hölder's inequality. *Acta Scientiarum Mathematicarum (Szeged)*, **33**, 217–223.
- Levy, S., and Fullagar, P. K. 1981. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, **46**(9), 1235–1243.
- Lieb, E. H. 1973. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, **11**, 267–288.
- Lindley, D. V. 1956. On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, **27**(4), 986–1005.
- Liu, H., Lafferty, J. D., and Wasserman, L. A. 2009. The nonparanormal: Semiparametric estimation of high-dimensional undirected graphs. *Journal of Machine Learning Research*, **10**, 1–37.
- Liu, H., Han, F., Yuan, M., Lafferty, J. D., and Wasserman, L. A. 2012. High-dimensional semiparametric Gaussian copula graphical models. *Annals of Statistics*, **40**(4), 2293–2326.
- Loh, P., and Wainwright, M. J. 2012. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Annals of Statistics*, **40**(3), 1637–1664.
- Loh, P., and Wainwright, M. J. 2013. Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Annals of Statistics*, **41**(6), 3022–3049.
- Loh, P., and Wainwright, M. J. 2015. Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Journal of Machine Learning Research*, **16**(April), 559–616.
- Loh, P., and Wainwright, M. J. 2017. Support recovery without incoherence: A case for nonconvex regularization. *Annals of Statistics*, **45**(6), 2455–2482. Appeared as arXiv:1412.5632.
- Lorentz, G. G. 1966. Metric entropy and approximation. *Bulletin of the AMS*, **72**(6), 903–937.
- Lounici, K., Pontil, M., Tsybakov, A. B., and van de Geer, S. 2011. Oracle inequalities and optimal inference under group sparsity. *Annals of Statistics*, **39**(4), 2164–2204.
- Lovász, L., and Schrijver, A. 1991. Cones of matrices and set-functions and 0 – 1 optimization. *SIAM Journal of Optimization*, **1**, 166–190.
- Ma, Z. 2010. *Contributions to high-dimensional principal component analysis*. Ph.D. thesis, Department of Statistics, Stanford University.
- Ma, Z. 2013. Sparse principal component analysis and iterative thresholding. *Annals of Statistics*, **41**(2), 772–801.
- Ma, Z., and Wu, Y. 2013. Computational barriers in minimax submatrix detection. *arXiv preprint arXiv:1309.5914*.
- Mackey, L. W., Jordan, M. I., Chen, R. Y., Farrell, B., and Tropp, J. A. 2014. Matrix concentration inequalities via the method of exchangeable pairs. *Annals of Probability*, **42**(3), 906–945.
- Mahoney, M. W. 2011. Randomized algorithms for matrices and data. *Foundations and Trends in Machine Learning*, **3**(2), 123–224.
- Marton, K. 1996a. Bounding  $d$ -distance by information divergence: a method to prove measure concentration. *Annals of Probability*, **24**, 857–866.
- Marton, K. 1996b. A measure concentration inequality for contracting Markov chains. *Geometric and Functional Analysis*, **6**(3), 556–571.
- Marton, K. 2004. Measure concentration for Euclidean distance in the case of dependent random variables. *Annals of Probability*, **32**(3), 2526–2544.
- Marčenko, V. A., and Pastur, L. A. 1967. Distribution of eigenvalues for some sets of random matrices. *Annals of Probability*, **4**(1), 457–483.
- Massart, P. 1990. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Annals of Probability*, **18**, 1269–1283.

- Massart, P. 2000. Some applications of concentration inequalities to statistics. *Annales de la Faculté des Sciences de Toulouse*, **IX**, 245–303.
- Maurey, B. 1991. Some deviation inequalities. *Geometric and Functional Analysis*, **1**, 188–197.
- McDiarmid, C. 1989. On the method of bounded differences. Pages 148–188 of: *Surveys in Combinatorics*. London Mathematical Society Lecture Notes, no. 141. Cambridge, UK: Cambridge University Press.
- Mehta, M. L. 1991. *Random Matrices*. New York, NY: Academic Press.
- Meier, L., van de Geer, S., and Bühlmann, P. 2009. High-dimensional additive modeling. *Annals of Statistics*, **37**, 3779–3821.
- Meinshausen, N. 2008. A note on the lasso for graphical Gaussian model selection. *Statistics and Probability Letters*, **78**(7), 880–884.
- Meinshausen, N., and Bühlmann, P. 2006. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, **34**, 1436–1462.
- Mendelson, S. 2002. Geometric parameters of kernel machines. Pages 29–43 of: *Proceedings of COLT*.
- Mendelson, S. 2010. Empirical processes with a bounded  $\psi_1$ -diameter. *Geometric and Functional Analysis*, **20**(4), 988–1027.
- Mendelson, S. 2015. Learning without concentration. *Journal of the ACM*, **62**(3), 1–25.
- Mendelson, S., Pajor, A., and Tomczak-Jaegermann, N. 2007. Reconstruction of subgaussian operators. *Geometric and Functional Analysis*, **17**(4), 1248–1282.
- Mézard, M., and Montanari, A. 2008. *Information, Physics and Computation*. New York, NY: Oxford University Press.
- Milman, V., and Schechtman, G. 1986. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Lecture Notes in Mathematics, vol. 1200. New York, NY: Springer.
- Minsker, S. 2011. *On some extensions of Bernstein's inequality for self-adjoint operators*. Tech. rept. Duke University.
- Mitjagin, B. S. 1961. The approximation dimension and bases in nuclear spaces. *Uspekhi. Mat. Naut.*, **61**(16), 63–132.
- Muirhead, R. J. 2008. *Aspects of multivariate statistical theory*. Wiley Series in Probability and Mathematical Statistics. New York, NY: Wiley.
- Müller, A. 1997. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, **29**(2), 429–443.
- Negahban, S., and Wainwright, M. J. 2011a. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Annals of Statistics*, **39**(2), 1069–1097.
- Negahban, S., and Wainwright, M. J. 2011b. Simultaneous support recovery in high-dimensional regression: Benefits and perils of  $\ell_{1,\infty}$ -regularization. *IEEE Transactions on Information Theory*, **57**(6), 3481–3863.
- Negahban, S., and Wainwright, M. J. 2012. Restricted strong convexity and (weighted) matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, **13**(May), 1665–1697.
- Negahban, S., Ravikumar, P., Wainwright, M. J., and Yu, B. 2010 (October). *A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers*. Tech. rept. UC Berkeley. Arxiv pre-print 1010.2731v1, Version 1.
- Negahban, S., Ravikumar, P., Wainwright, M. J., and Yu, B. 2012. A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statistical Science*, **27**(4), 538–557.
- Nemirovski, A. 2000. Topics in non-parametric statistics. In: Bernard, P. (ed), *Ecole d'Été de Probabilités de Saint-Flour XXVIII*. Lecture Notes in Mathematics. Berlin, Germany: Springer.
- Nesterov, Y. 1998. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software*, **9**(1), 141–160.
- Netrapalli, P., Banerjee, S., Sanghavi, S., and Shakkottai, S. 2010. Greedy learning of Markov network structure. Pages 1295–1302 of: *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE.
- Obozinski, G., Wainwright, M. J., and Jordan, M. I. 2011. Union support recovery in high-dimensional multivariate regression. *Annals of Statistics*, **39**(1), 1–47.
- Oldenburg, D. W., Scheuer, T., and Levy, S. 1983. Recovery of the acoustic impedance from reflection seismograms. *Geophysics*, **48**(10), 1318–1337.

- Oliveira, R. I. 2010. Sums of random Hermitian matrices and an inequality by Rudelson. *Electronic Communications in Probability*, **15**, 203–212.
- Oliveira, R. I. 2013. *The lower tail of random quadratic forms, with applicaitons to ordinary least squares and restricted eigenvalue properties*. Tech. rept. IMPA, Rio de Janeiro, Brazil.
- Ortega, J. M., and Rheinboldt, W. C. 2000. *Iterative Solution of Nonlinear Equations in Several Variables*. Classics in Applied Mathematics. New York, NY: SIAM.
- Pastur, L. A. 1972. On the spectrum of random matrices. *Theoretical and Mathematical Physics*, **10**, 67–74.
- Paul, D. 2007. Asymptotics of sample eigenstructure for a large-dimensional spiked covariance model. *Statistica Sinica*, **17**, 1617–1642.
- Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann.
- Petrov, V. V. 1995. *Limit theorems of probability theory: Sequence of independent random variables*. Oxford, UK: Oxford University Press.
- Pilanci, M., and Wainwright, M. J. 2015. Randomized sketches of convex programs with sharp guarantees. *IEEE Transactions on Information Theory*, **9**(61), 5096–5115.
- Pinkus, A. 1985. *N-Widths in Approximation Theory*. New York: Springer.
- Pisier, G. 1989. *The Volume of Convex Bodies and Banach Space Geometry*. Cambridge Tracts in Mathematics, vol. 94. Cambridge, UK: Cambridge University Press.
- Pollard, D. 1984. *Convergence of Stochastic Processes*. New York, NY: Springer.
- Portnoy, S. 1984. Asymptotic behavior of M-estimators of  $p$  regression parameters when  $p^2/n$  is large: I. Consistency. *Annals of Statistics*, **12**(4), 1296–1309.
- Portnoy, S. 1985. Asymptotic behavior of M-estimators of  $p$  regression parameters when  $p^2/n$  is large: II. Normal approximation. *Annals of Statistics*, **13**(4), 1403–1417.
- Portnoy, S. 1988. Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Annals of Statistics*, **16**(1), 356–366.
- Prékopa, A. 1971. Logarithmic concave measures with application to stochastic programming. *Acta Scientiarum Mathematicarum (Szeged)*, **32**, 301–315.
- Prékopa, A. 1973. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum (Szeged)*, **33**, 335–343.
- Rachev, S. T., and Ruschendorf, L. 1998. *Mass Transportation Problems, Volume II, Applications*. New York, NY: Springer.
- Rachev, S. T., Klebanov, L., Stoyanov, S. V., and Fabozzi, F. 2013. *The Method of Distances in the Theory of Probability and Statistics*. New York, NY: Springer.
- Rao, C. R. 1949. On some problems arising out of discrimination with multiple characters. *Sankhya (Indian Journal of Statistics)*, **9**(4), 343–366.
- Raskutti, G., Wainwright, M. J., and Yu, B. 2010. Restricted eigenvalue conditions for correlated Gaussian designs. *Journal of Machine Learning Research*, **11**(August), 2241–2259.
- Raskutti, G., Wainwright, M. J., and Yu, B. 2011. Minimax rates of estimation for high-dimensional linear regression over  $\ell_q$ -balls. *IEEE Transactions on Information Theory*, **57**(10), 6976–6994.
- Raskutti, G., Wainwright, M. J., and Yu, B. 2012. Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *Journal of Machine Learning Research*, **12**(March), 389–427.
- Raudys, V., and Young, D. M. 2004. Results in Statistical Discriminant Analysis: A Review of the Former Soviet Union Literature. *Journal of Multivariate Analysis*, **89**(1), 1–35.
- Ravikumar, P., Liu, H., Lafferty, J. D., and Wasserman, L. A. 2009. SpAM: sparse additive models. *Journal of the Royal Statistical Society, Series B*, **71**(5), 1009–1030.
- Ravikumar, P., Wainwright, M. J., and Lafferty, J. D. 2010. High-dimensional Ising model selection using  $\ell_1$ -regularized logistic regression. *Annals of Statistics*, **38**(3), 1287–1319.
- Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. 2011. High-dimensional covariance estimation by minimizing  $\ell_1$ -penalized log-determinant divergence. *Electronic Journal of Statistics*, **5**, 935–980.
- Recht, B. 2011. A Simpler Approach to Matrix Completion. *Journal of Machine Learning Research*, **12**, 3413–3430.
- Recht, B., Xu, W., and Hassibi, B. 2009. *Null space conditions and thresholds for rank minimization*. Tech. rept. U. Madison. Available at <http://pages.cs.wisc.edu/brecht/papers/10.RecXuHas.Thresholds.pdf>.

- Recht, B., Fazel, M., and Parrilo, P. A. 2010. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, **52**(3), 471–501.
- Reeves, G., and Gastpar, M. 2008 (July). Sampling Bounds for Sparse Support Recovery in the Presence of Noise. In: *International Symposium on Information Theory*.
- Reinsel, G. C., and Velu, R. P. 1998. *Multivariate Reduced-Rank Regression*. Lecture Notes in Statistics, vol. 136. New York, NY: Springer.
- Ren, Z., and Zhou, H. H. 2012. Discussion: Latent variable graphical model selection via convex optimization. *Annals of Statistics*, **40**(4), 1989–1996.
- Richardson, T., and Urbanke, R. 2008. *Modern Coding Theory*. Cambridge University Press.
- Rockafellar, R. T. 1970. *Convex Analysis*. Princeton: Princeton University Press.
- Rohde, A., and Tsybakov, A. B. 2011. Estimation of high-dimensional low-rank matrices. *Annals of Statistics*, **39**(2), 887–930.
- Rosenbaum, M., and Tsybakov, A. B. 2010. Sparse recovery under matrix uncertainty. *Annals of Statistics*, **38**, 2620–2651.
- Rosenthal, H. P. 1970. On the subspaces of  $\ell^p$  ( $p > 2$ ) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, **8**, 1546–1570.
- Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. 2008. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, **2**, 494–515.
- Rudelson, M. 1999. Random vectors in the isotropic position. *Journal of Functional Analysis*, **164**, 60–72.
- Rudelson, M., and Vershynin, R. 2013. Hanson–Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, **18**(82), 1–9.
- Rudelson, M., and Zhou, S. 2013. Reconstruction from anisotropic random measurements. *IEEE Transactions on Information Theory*, **59**(6), 3434–3447.
- Rudin, W. 1964. *Principles of Mathematical Analysis*. New York, NY: McGraw-Hill.
- Rudin, W. 1990. *Fourier Analysis on Groups*. New York, NY: Wiley-Interscience.
- Samson, P. M. 2000. Concentration of measure inequalities for Markov chains and  $\Phi$ -mixing processes. *Annals of Probability*, **28**(1), 416–461.
- Santhanam, N. P., and Wainwright, M. J. 2012. Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Transactions on Information Theory*, **58**(7), 4117–4134.
- Santosa, F., and Symes, W. W. 1986. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, **7**(4), 1307–1330.
- Saulis, L., and Statulevicius, V. 1991. *Limit Theorems for Large Deviations*. London: Kluwer Academic.
- Schölkopf, B., and Smola, A. 2002. *Learning with Kernels*. Cambridge, MA: MIT Press.
- Schütt, C. 1984. Entropy numbers of diagonal operators between symmetric Banach spaces. *Journal of Approximation Theory*, **40**, 121–128.
- Scott, D. W. 1992. *Multivariate Density Estimation: Theory, Practice and Visualization*. New York, NY: Wiley.
- Seijo, E., and Sen, B. 2011. Nonparametric least squares estimation of a multivariate convex regression function. *Annals of Statistics*, **39**(3), 1633–1657.
- Serdobolskii, V. 2000. *Multivariate Statistical Analysis*. Dordrecht, The Netherlands: Kluwer Academic.
- Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423.
- Shannon, C. E. 1949. Communication in the presence of noise. *Proceedings of the IRE*, **37**(1), 10–21.
- Shannon, C. E., and Weaver, W. 1949. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.
- Shao, J. 2007. *Mathematical Statistics*. New York, NY: Springer.
- Shor, N. Z. 1987. Quadratic optimization problems. *Soviet Journal of Computer and System Sciences*, **25**, 1–11.
- Silverman, B. W. 1982. On the estimation of a probability density function by the maximum penalized likelihood method. *Annals of Statistics*, **10**(3), 795–810.
- Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. Boca Raton, FL: CRC Press.
- Silverstein, J. 1995. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, **55**, 331–339.

- Slepian, D. 1962. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, **42**(2), 463–501.
- Smale, S., and Zhou, D. X. 2003. Estimating the approximation error in learning theory. *Analysis and Its Applications*, **1**(1), 1–25.
- Spirtes, P., Glymour, C., and Scheines, R. 2000. *Causation, Prediction and Search*. Cambridge, MA: MIT Press.
- Srebro, N. 2004. *Learning with Matrix Factorizations*. Ph.D. thesis, MIT. Available online: <http://ttic.uchicago.edu/~nati/Publications/thesis.pdf>.
- Srebro, N., Rennie, J., and Jaakkola, T. S. 2005a (December 2004). Maximum-margin matrix factorization. In: *Advances in Neural Information Processing Systems 17 (NIPS 2004)*.
- Srebro, N., Alon, N., and Jaakkola, T. S. 2005b (December). Generalization error bounds for collaborative prediction with low-rank matrices. In: *Advances in Neural Information Processing Systems 17 (NIPS 2004)*.
- Srivastava, N., and Vershynin, R. 2013. Covariance estimation for distributions with  $2 + \epsilon$  moments. *Annals of Probability*, **41**, 3081–3111.
- Steele, J. M. 1978. Empirical discrepancies and sub-additive processes. *Annals of Probability*, **6**, 118–127.
- Steinwart, I., and Christmann, A. 2008. *Support vector machines*. New York, NY: Springer.
- Stewart, G. W. 1971. Error bounds for approximate invariant subspaces of closed linear operators. *SIAM Journal on Numerical Analysis*, **8**(4), 796–808.
- Stewart, G. W., and Sun, J. 1980. *Matrix Perturbation Theory*. New York, NY: Academic Press.
- Stone, C. J. 1982. Optimal global rates of convergence for non-parametric regression. *Annals of Statistics*, **10**(4), 1040–1053.
- Stone, C. J. 1985. Additive regression and other non-parametric models. *Annals of Statistics*, **13**(2), 689–705.
- Szarek, S. J. 1991. Condition numbers of random matrices. *J. Complexity*, **7**(2), 131–149.
- Talagrand, M. 1991. A new isoperimetric inequality and the concentration of measure phenomenon. Pages 94–124 of: Lindenstrauss, J., and Milman, V. D. (eds), *Geometric Aspects of Functional Analysis*. Lecture Notes in Mathematics, vol. 1469. Berlin, Germany: Springer.
- Talagrand, M. 1995. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. I.H.E.S.*, **81**, 73–205.
- Talagrand, M. 1996a. New concentration inequalities in product spaces. *Inventiones Mathematicae*, **126**, 503–563.
- Talagrand, M. 1996b. A new look at independence. *Annals of Probability*, **24**(1), 1–34.
- Talagrand, M. 2000. *The Generic Chaining*. New York, NY: Springer.
- Talagrand, M. 2003. *Spin Glasses: A Challenge for Mathematicians*. New York, NY: Springer.
- Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58**(1), 267–288.
- Tibshirani, R., Saunders, M. A., Rosset, S., Zhu, J., and Knight, K. 2005. Sparsity and smoothness via the smoothed Lasso. *Journal of the Royal Statistical Society B*, **67**(1), 91–108.
- Tropp, J. A. 2006. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, **52**(3), 1030–1051.
- Tropp, J. A. 2010 (April). *User-friendly tail bounds for matrix martingales*. Tech. rept. Caltech.
- Tsybakov, A. B. 2009. *Introduction to non-parametric estimation*. New York, NY: Springer.
- Turlach, B., Venables, W.N., and Wright, S.J. 2005. Simultaneous variable selection. *Technometrics*, **27**, 349–363.
- van de Geer, S. 2000. *Empirical Processes in M-Estimation*. Cambridge University Press.
- van de Geer, S. 2014. Weakly decomposable regularization penalties and structured sparsity. *Scandinavian Journal of Statistics*, **41**, 72–86.
- van de Geer, S., and Bühlmann, P. 2009. On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, **3**, 1360–1392.
- van der Vaart, A. W., and Wellner, J. A. 1996. *Weak Convergence and Empirical Processes*. New York, NY: Springer.

- Vempala, S. 2004. *The Random Projection Method*. Discrete Mathematics and Theoretical Computer Science. Providence, RI: American Mathematical Society.
- Vershynin, R. 2011. *Introduction to the non-asymptotic analysis of random matrices*. Tech. rept. Univ. Michigan.
- Villani, C. 2008. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften, vol. 338. New York, NY: Springer.
- Vu, V. Q., and Lei, J. 2012. Minimax rates of estimation for sparse PCA in high dimensions. In: *15th Annual Conference on Artificial Intelligence and Statistics*.
- Wachter, K. 1978. The strong limits of random matrix spectra for samples matrices of independent elements. *Annals of Probability*, **6**, 1–18.
- Wahba, G. 1990. *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PN: SIAM.
- Wainwright, M. J. 2009a. Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting. *IEEE Transactions on Information Theory*, **55**(December), 5728–5741.
- Wainwright, M. J. 2009b. Sharp thresholds for high-dimensional and noisy sparsity recovery using  $\ell_1$ -constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory*, **55**(May), 2183–2202.
- Wainwright, M. J. 2014. Constrained forms of statistical minimax: Computation, communication and privacy. In: *Proceedings of the International Congress of Mathematicians*.
- Wainwright, M. J., and Jordan, M. I. 2008. Graphical models, exponential families and variational inference. *Foundations and Trends in Machine Learning*, **1**(1–2), 1–305.
- Waldspurger, I., d'Aspremont, A., and Mallat, S. 2015. Phase recovery, MaxCut and complex semidefinite programming. *Mathematical Programming A*, **149**(1–2), 47–81.
- Wang, T., Berthet, Q., and Samworth, R. J. 2014 (August). *Statistical and computational trade-offs in estimation of sparse principal components*. Tech. rept. arxiv:1408.5369. University of Cambridge.
- Wang, W., Wainwright, M. J., and Ramchandran, K. 2010. Information-theoretic limits on sparse signal recovery: dense versus sparse measurement matrices. *IEEE Transactions on Information Theory*, **56**(6), 2967–2979.
- Wang, W., Ling, Y., and Xing, E. P. 2015. Collective Support Recovery for Multi-Design Multi-Response Linear Regression. *IEEE Transactions on Information Theory*, **61**(1), 513–534.
- Wasserman, L. A. 2006. *All of Non-Parametric Statistics*. Springer Series in Statistics. New York, NY: Springer.
- Widom, H. 1963. Asymptotic behaviour of Eigenvalues of Certain Integral Operators. *Transactions of the American Mathematical Society*, **109**, 278–295.
- Widom, H. 1964. Asymptotic behaviour of Eigenvalues of Certain Integral Operators II. *Archive for Rational Mechanics and Analysis*, **17**(3), 215–229.
- Wigner, E. 1955. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, **62**, 548–564.
- Wigner, E. 1958. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, **67**, 325–327.
- Williams, D. 1991. *Probability with Martingales*. Cambridge, UK: Cambridge University Press.
- Witten, D., Tibshirani, R., and Hastie, T. J. 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biometrika*, **10**, 515–534.
- Woodruff, D. 2014. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, **10**(10), 1–157.
- Wright, F. T. 1973. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Annals of Probability*, **1**(6), 1068–1070.
- Xu, M., Chen, M., and Lafferty, J. D. 2014. *Faithful variable selection for high dimensional convex regression*. Tech. rept. Univ. Chicago. arxiv:1411.1805.
- Xu, Q., and You, J. 2007. Covariate selection for linear errors-in-variables regression models. *Communications in Statistics – Theory and Methods*, **36**(2), 375–386.
- Xue, L., and Zou, H. 2012. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Annals of Statistics*, **40**(5), 2541–2571.

- Yang, Y., and Barron, A. 1999. Information-theoretic determination of minimax rates of convergence. *Annals of Statistics*, **27**(5), 1564–1599.
- Ye, F., and Zhang, C. H. 2010. Rate minimaxity of the Lasso and Dantzig selector for the  $\ell_q$ -loss in  $\ell_r$ -balls. *Journal of Machine Learning Research*, **11**, 3519–3540.
- Yu, B. 1996. Assouad, Fano and Le Cam. *Research Papers in Probability and Statistics: Festschrift in Honor of Lucien Le Cam*, 423–435.
- Yuan, M. 2010. High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, **11**, 2261–2286.
- Yuan, M., and Lin, Y. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society B*, **1**(68), 49.
- Yuan, M., and Lin, Y. 2007. Model selection and estimation in the Gaussian graphical model. *Biometrika*, **94**(1), 19–35.
- Yuan, X. T., and Zhang, T. 2013. Truncated power method for sparse eigenvalue problems. *Journal of Machine Learning Research*, **14**, 899–925.
- Yurinsky, V. 1995. *Sums and Gaussian Vectors*. Lecture Notes in Mathematics. New York, NY: Springer.
- Zhang, C. H. 2012. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, **38**(2), 894–942.
- Zhang, C. H., and Zhang, T. 2012. A general theory of concave regularization for high-dimensional sparse estimation problems. *Statistical Science*, **27**(4), 576–593.
- Zhang, Y., Wainwright, M. J., and Jordan, M. I. 2014 (June). Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In: *Proceedings of the Conference on Learning Theory (COLT)*. Full length version at <http://arxiv.org/abs/1402.1918>.
- Zhang, Y., Wainwright, M. J., and Jordan, M. I. 2017. Optimal prediction for sparse linear models? Lower bounds for coordinate-separable M-estimators. *Electronic Journal of Statistics*, **11**, 752–799.
- Zhao, P., and Yu, B. 2006. On model selection consistency of Lasso. *Journal of Machine Learning Research*, **7**, 2541–2567.
- Zhao, P., Rocha, G., and Yu, B. 2009. Grouped and hierarchical model selection through composite absolute penalties. *Annals of Statistics*, **37**(6A), 3468–3497.
- Zhou, D. X. 2013. Density problem and approximation error in learning theory. *Abstract and Applied Analysis*, **2013**(715683).
- Zou, H. 2006. The Adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, **101**(476), 1418–1429.
- Zou, H., and Hastie, T. J. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, **67**(2), 301–320.
- Zou, H., and Li, R. 2008. One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics*, **36**(4), 1509–1533.