References

Aït-Sahalia, Yacine, and Duarte, Jefferson. 2003. Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, **116**, 9–47.

Aitken, Alexander C. 1936. IV.—On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, **55**, 42–48.

Aizerman, A. 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, **25**, 821–837.

Ali, Syed Mumtaz, and Silvey, Samuel D. 1966. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B*, **28**, 131–142.

Amelunxen, Dennis, Lotz, Martin, McCoy, Michael B, and Tropp, Joel A. 2014. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference: A Journal of the IMA*, **3**, 224–294.

Anderson, Theodore W. 2003. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics.

Arlot, Sylvain, and Celisse, Alain. 2010. A survey of cross-validation procedures for model selection. *Statistical Surveys*, **4**, 40–79.

Aronszajn, Nachman. 1950. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68**, 337–404.

Arora, Sanjeev, and Barak, Boaz. 2009. *Computational Complexity: A Modern Approach*. Cambridge University Press.

Assouad, Patrice. 1983. Deux remarques sur l'estimation. *Comptes Rendus des Séances de l'Académie des Sciences. Série 1, Mathématique*, **296**, 1021–1024.

Axelrod, Brian, Diakonikolas, Ilias, Stewart, Alistair, Sidiropoulos, Anastasios, and Valiant, Gregory. 2019. A polynomial time algorithm for log-concave maximum likelihood via locally exponential families. *Advances in Neural Information Processing Systems*, **32**.

Ayer, Miriam, Brunk, H Daniel, Ewing, George M, Reid, William T, and Silverman, Edward. 1955. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, **26**, 641–647.

Azuma, Kazuoki. 1967. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, **19**, 357–367.

Balabdaoui, Fadoua, and Doss, Charles R. 2018. Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli*, **24**, 1053–1071.

Balabdaoui, Fadoua, Rufibach, Kaspar, and Wellner, Jon A. 2009. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of Statistics*, **37**, 1299–1331.

Balabdaoui, Fadoua, Jankowski, Hanna, Pavlides, Marios, Seregin, Arseni, and Wellner, Jon. 2011. On the Grenander estimator at zero. *Statistica Sinica*, **21**, 873.

Balabdaoui, Fadoua, Durot, Cécile, and Jankowski, Hanna. 2019. Least squares estimation in the monotone single index model. *Bernoulli*, **25**, 3276–3310.

Published online by Cambridge University Press

Balázs, Gábor, György, András, and Szepesvári, Csaba. 2015. Near-optimal max-affine estimators for convex regression. Pages 56–64 of: *Artificial Intelligence and Statistics*. PMLR.

Banerjee, Moulinath, and Wellner, Jon A. 2001. Likelihood ratio tests for monotone functions. *Annals of Statistics*, **29**, 1699–1731.

Banerjee, Onureena, El Ghaoui, Laurent, and d'Aspremont, Alexandre. 2008. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research*, **9**, 485–516.

Baranchik, Alvin J. 1964. *Multiple regression and estimation of the mean of a multivariate normal distribution*. Tech. rept. 51. Department of Statistics, Stanford University.

Baranchik, Alvin J. 1970. A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, **41**, 642–645.

Baraud, Yannick, and Birgé, Lucien. 2016. Rho-estimators for shape restricted density estimation. *Stochastic Processes and their Applications*, **126**, 3888–3912.

Baraud, Yannick, Halconruy, Hèlène, and Maillard, Guillaume. 2025. Robust density estimation with the $\mathbb{L}_1$-loss. Applications to the estimation of a density on the line satisfying a shape constraint. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **61**, 2098–2148.

Barber, Rina Foygel, and Samworth, Richard J. 2021. Local continuity of log-concave projection, with applications to estimation under model misspecification. *Bernoulli*, **27**, 2437–2472.

Bartle, Robert G. 2014. *The Elements of Integration and Lebesgue Measure*. John Wiley & Sons.

Bartlett, Peter L, Bousquet, Olivier, and Mendelson, Shahar. 2005. Local Rademacher complexities. *Annals of Statistics*, **33**, 1497–1537.

Bauer, Heinz. 2011. *Measure and Integration Theory*. de Gruyter.

Bellec, Pierre C. 2018. Sharp oracle inequalities for least squares estimators in shape restricted regression. *Annals of Statistics*, **46**, 745–780.

Bellec, Pierre C, Lecué, Guillaume, and Tsybakov, Alexandre B. 2018. Slope meets Lasso: Improved oracle bounds and optimality. *Annals of Statistics*, **46**, 3603–3642.

Belloni, Alexandre, Chernozhukov, Victor, and Wang, Lie. 2011. Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, **98**, 791–806.

Beltrami, Eugenio. 1873. Sulle Funzioni Bilineari. *Giornale di Mathematiche*, **11**, 98–106.

Ben-Hamou, Anna, Boucheron, Stéphane, and Ohannessian, Mesrob I. 2017. Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli*, **23**, 249–287.

Bengs, Viktor, and Holzmann, Hajo. 2019. Uniform approximation in classical weak convergence theory. *arXiv:1903.09864*.

Benjamini, Yoav, and Hochberg, Yosef. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, **57**, 289–300.

Benjamini, Yoav, and Yekutieli, Daniel. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.

Bennett, George. 1962. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, **57**, 33–45.

Berend, Daniel, and Kontorovich, Aryeh. 2013. On the concentration of the missing mass. *Electronic Communications in Probability*, **18**, 1–7.

Berezin, Felix Aleksandrovich, and Gel'fand, Izrail Moiseevich. 1956. Some remarks on the theory of spherical functions on symmetric Riemannian manifolds. *Trudy Moskovskogo Matematicheskogo Obshchestva*, **5**, 311–351.

Berger, James O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer Science+Business Media.

Published online by Cambridge University Press

Bernstein, Sergei N. 1937. On certain modifications of Chebyshev's inequality. *Doklady Akademii Nauk SSSR*, **17**, 275–277.

Berrett, Thomas B, Wang, Yi, Barber, Rina Foygel, and Samworth, Richard J. 2020. The conditional permutation test for independence while controlling for confounders. *Journal of the Royal Statistical Society, Series B*, **82**, 175–197.

Berry, Andrew C. 1941. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, **49**, 122–136.

Berthet, Quentin, and Rigollet, Philippe. 2013. Optimal detection of sparse principal components in high dimension. *Annals of Statistics*, **41**, 1780–1815.

Bhatia, Rajendra. 2007. *Perturbation Bounds for Matrix Eigenvalues*. SIAM.

Bickel, Peter J. 1975. One-step Huber estimates in the linear model. *Journal of the American Statistical Association*, **70**, 428–434.

Bickel, Peter J, Klaassen, Chris AJ, Ritov, Ya'acov, and Wellner, Jon A. 1993. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

Bickel, Peter J, Ritov, Ya'acov, and Tsybakov, Alexandre B. 2009. Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics*, **37**, 1705–1732.

Billingsley, Patrick. 1995. *Probability and Measure*. John Wiley & Sons.

Birgé, Lucien. 1983. Approximation dans les espaces métriques et théorie de l'estimation. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **65**, 181–237.

Birgé, Lucien. 1987. Estimating a density under order restrictions: Nonasymptotic minimax risk. *Annals of Statistics*, **15**, 995–1012.

Birgé, Lucien. 1989. The Grenander estimator: A nonasymptotic approach. *Annals of Statistics*, **17**, 1532–1549.

Bjerhammer, Arne. 1951. *Application of Calculus of Matrices to Method of Least Squares with Special Reference to Geodetic Calculations*. Lindståhl, Stockholm.

Bobkov, Sergey, and Madiman, Mokshay. 2011. Concentration of the information in data with log-concave distributions. *Annals of Probability*, **39**, 1528–1543.

Bobkov, Sergey G. 1999. Isoperimetric and analytic inequalities for log-concave probability measures. *Annals of Probability*, **27**, 1903–1921.

Bochner, Salomon. 1932. *Vorlesungen uber Fouriersche Integrale*. Akademische Verlagsgesellschaft.

Bochner, Salomon. 1933. Monotone funktionen, stieltjessche integrale und harmonische analyse. *Mathematische Annalen*, **108**, 378–410.

Bollobás, Béla. 1990. *Linear Analysis*. Cambridge University Press.

Bonferroni, Carlo. 1936. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, **8**, 3–62.

Borel, Émile M. 1909. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, **27**, 247–271.

Boser, Bernhard E, Guyon, Isabelle M, and Vapnik, Vladimir N. 1992. A training algorithm for optimal margin classifiers. Pages 144–152 of: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*.

Boucheron, Stéphane, Lugosi, Gábor, and Massart, Pascal. 2013. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Boyd, Stephen, and Vandenberghe, Lieven. 2004. *Convex Optimization*. Cambridge University Press.

Brandwein, Ann Cohen, and Strawderman, William E. 1990. Stein estimation: The spherically symmetric case. *Statistical Science*, **5**, 356–369.

Breiman, Leo. 1996. Stacked regressions. *Machine Learning*, **24**, 49–64.

Bretagnolle, Jean, and Huber, Catherine. 1978. Estimation des densités: risque minimax. *Séminaire de Probabilités de Strasbourg*, **12**, 342–363.

Bretagnolle, Jean, and Huber, Catherine. 1979. Estimation des densités: risque minimax. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **47**, 119–137.

Published online by Cambridge University Press

Brown, Lawrence D, and Low, Mark G. 1996. A constrained risk inequality with applications to nonparametric functional estimation. *Annals of Statistics*, **24**, 2524–2535.

Brunk, Hugh D. 1955. Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics*, **26**, 607–616.

Brunk, Hugh D, Barlow, Richard E, Bartholomew, Daniel J, and Bremner, James M. 1972. *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*. John Wiley & Sons.

Bühlmann, Peter, and van de Geer, Sara. 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Science+Business Media.

Bühlmann, Peter, Rütimann, Philipp, van de Geer, Sara, and Zhang, Cun-Hui. 2013. Correlated variables in regression: Clustering and sparse estimation. *Journal of Statistical Planning and Inference*, **143**, 1835–1858.

Buldygin, Valerii Vladimirovich, and Kozachenko, Yu V. 2000. *Metric Characterization of Random Variables and Random Processes*. American Mathematical Society.

Bunea, Florentina, Tsybakov, Alexandre, and Wegkamp, Marten. 2007. Sparsity oracle inequalities for the Lasso. *Electronic Journal of Statistics*, **1**, 169–194.

Burkill, John Charles. 1978. *A First Course in Mathematical Analysis*. Cambridge University Press.

Cai, Tony, Liu, Weidong, and Luo, Xi. 2011. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **106**, 594–607.

Candès, Emmanuel, and Tao, Terence. 2007. The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, **35**, 2313–2351.

Candès, Emmanuel, Fan, Yingying, Janson, Lucas, and Lv, Jinchi. 2018. Panning for gold: 'Model-X' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society, Series B*, **80**, 551–577.

Candès, Emmanuel J, and Plan, Yaniv. 2009. Near-ideal model selection by ℓ_1 minimization. *Annals of Statistics*, **37**, 2145–2177.

Canonne, Clément L. 2022. A short note on an inequality between KL and TV. *arXiv:2202.07198*.

Cantelli, Francesco P. 1933. Sulla determinazione empirica delle leggi di probabilita. *Giorn. Ist. Ital. Attuari*, **4**, 421–424.

Cantelli, Francesco, P. 1917. Sulla probabilità come limite della frequenza. *Rendiconti della R. Accademia dei Lincei*, **26**, 39–45.

Cao, Liying, Shi, Pei-Jian, Li, Lin, and Chen, Guifen. 2019. A new flexible sigmoidal growth model. *Symmetry*, **11**, 204.

Carpenter, Timothy, Diakonikolas, Ilias, Sidiropoulos, Anastasios, and Stewart, Alistair. 2018. Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. Pages 1234–1262 of: *Conference On Learning Theory*. PMLR.

Casella, George, and Hwang, Jiunn Tzon. 1983. Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *Journal of the American Statistical Association*, **78**, 688–698.

Castro, M, Menegatto, V, and Peron, A. 2012. Integral operators generated by Mercer-like kernels on topological spaces. *Colloquium Mathematicum*, **126**, 125–138.

Catoni, Olivier. 2012. Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **48**, 1148–1185.

Čech, Eduard. 1937. On bicompact spaces. *Annals of Mathematics*, **38**, 823–844.

Celentano, Michael, Montanari, Andrea, and Wei, Yuting. 2023. The Lasso with general Gaussian designs with applications to hypothesis testing. *Annals of Statistics*, **51**, 2194–2220.

Čencov, Nikolai N. 1962. Estimation of an unknown distribution density from observations. *Soviet Mathematics*, **3**, 1559–1566.

Published online by Cambridge University Press

Chang, George T, and Walther, Guenther. 2007. Clustering with mixtures of log-concave distributions. *Computational Statistics & Data Analysis*, **51**, 6242–6251.

Chang, Joseph T, and Pollard, David. 1997. Conditioning as disintegration. *Statistica Neerlandica*, **51**, 287–317.

Charig, C. R., Webb, D. R., Payne, S. R., and Wickham, J. E. A. 1986. Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *British Medical Journal (Clin. Res. Ed.)*, **292**, 879–882.

Charles, Denis, Chickering, Max, and Simard, Patrice. 2013. Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research*, **14**, 3207–3260.

Chatfield, Christopher, and Collins, Alexander J. 2018. *Introduction to Multivariate Analysis*. Routledge.

Chatterjee, Sabyasachi. 2016. An improved global risk bound in concave regression. *Electronic Journal of Statistics*, **10**, 1608–1629.

Chatterjee, Sabyasachi, Guntuboyina, Adityanand, and Sen, Bodhisattva. 2015. On risk bounds in isotonic and other shape restricted regression problems. *Annals of Statistics*, **43**, 1774–1800.

Chatterjee, Sourav. 2014. A new perspective on least squares under convex constraint. *Annals of Statistics*, **42**, 2340–2381.

Chen, Louis HY, Goldstein, Larry, and Shao, Qi-Man. 2011. *Normal Approximation by Stein's Method*. Springer Science+Business Media.

Chen, Wenyu, Mazumder, Rahul, and Samworth, Richard J. 2024. A new computational framework for log-concave density estimation. *Mathematical Programming Computation*, **16**, 185–228.

Chen, Yining, and Samworth, Richard J. 2013. Smoothed log-concave maximum likelihood estimation with applications. *Statistica Sinica*, **23**, 1373–1398.

Chen, Yining, and Samworth, Richard J. 2014. scar: Shape-Constrained Additive Regression: a Maximum Likelihood Approach. *R package version 0.2-2*.

Chen, Yining, and Samworth, Richard J. 2016. Generalized additive and index models with shape constraints. *Journal of the Royal Statistical Society, Series B*, **78**, 729–754.

Chen, Yuansi. 2021. An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geometric and Functional Analysis*, **31**, 34–61.

Chernoff, Herman. 1952. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, **23**, 493–507.

Chernozhukov, Victor, Hansen, Christian, and Liao, Yuan. 2017. A lava attack on the recovery of sums of dense and sparse signals. *Annals of Statistics*, **45**, 39–76.

Chernozhukov, Victor, Chetverikov, Denis, Demirer, Mert, Duflo, Esther, Hansen, Christian, Newey, Whitney, and Robins, James. 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, **21**, C1–C68.

Chetverikov, Denis, Liao, Zhipeng, and Chernozhukov, Victor. 2021. On cross-validated Lasso in high dimensions. *Annals of Statistics*, **49**, 1300–1317.

Chhor, Julien, Klopp, Olga, and Tsybakov, Alexandre. 2024. Generalized multi-view model: Adaptive density estimation under low-rank constraints. *arXiv:2404.17209*.

Claerbout, Jon F, and Muir, Francis. 1973. Robust modeling with erratic data. *Geophysics*, **38**, 826–844.

Cochran, William G. 1934. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Mathematical Proceedings of the Cambridge Philosophical Society*, **30**, 178–191.

Cohn, Donald L. 2013. *Measure Theory*. Birkhäuser.

Comon, Pierre. 1994. Independent component analysis, a new concept? *Signal Processing*, **36**, 287–314.

Conway, John B. 2019. *A Course in Functional Analysis*. Springer.

Published online by Cambridge University Press

Cortes, Corinna, and Vapnik, Vladimir. 1995. Support-vector networks. *Machine Learning*, **20**, 273–297.

Courant, Richard. 1920. Über die Eigenwerte bei den Differentialgleichungen der mathematischen Physik. *Mathematische Zeitschrift*, **7**, 1–57.

Cover, Thomas M, and Thomas, Joy A. 1999. *Elements of Information Theory*. John Wiley & Sons.

Cristianini, Nello, and Shawe-Taylor, John. 2000. *An Introduction to Support Vector Machines*. Cambridge University Press.

Csiszár, Imre. 1967. Information-type measures of difference of probability distributions and indirect observation. *Studia Scientiarum Mathematicarum Hungarica*, **2**, 229–318.

Cule, Madeleine, and Samworth, Richard. 2010. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic Journal of Statistics*, **4**, 254–270.

Cule, Madeleine, Gramacy, Robert B, Samworth, Richard, and Chen, Yining. 2007. LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density. *R package version 1.6-12*.

Cule, Madeleine, Gramacy, Robert B, and Samworth, Richard. 2009. LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density. *Journal of Statistical Software*, **29**, 1–20.

Cule, Madeleine, Samworth, Richard, and Stewart, Michael. 2010. Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society, Series B (with discussion)*, **72**, 545–607.

Dalalyan, Arnak S. 2017. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society, Series B*, **79**, 651–676.

Dalalyan, Arnak S, Hebiri, Mohamed, and Lederer, Johannes. 2017. On the prediction performance of the Lasso. *Bernoulli*, **23**, 552–581.

Daubechies, Ingrid, Defrise, Michel, and De Mol, Christine. 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, **57**, 1413–1457.

Davis, Chandler, and Kahan, William Morton. 1970. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, **7**, 1–46.

Dawid, A Philip. 1979. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B*, **41**, 1–15.

De Boor, Carl. 1976. *Splines as Linear Combinations of B-Splines: A Survey*. University of Wisconsin–Madison. Mathematics Research Center.

de Moivre, Abraham. 1733. *Approximatio ad Summam Terminorum Binomii* $(a+b)^n$ *in Seriem expansi*. Printed for private circulation.

Debnath, Lokenath, and Mikusiński, Piotr. 1999. *Introduction to Hilbert Spaces with Applications*. Academic Press.

Deng, Hang, and Zhang, Cun-Hui. 2020. Isotonic regression in multi-dimensional spaces and graphs. *Annals of Statistics*, **48**, 3672–3698.

Deng, Hang, Han, Qiyang, and Zhang, Cun-Hui. 2021. Confidence intervals for multiple isotonic regression and other monotone models. *Annals of Statistics*, **49**, 2021–2052.

Deng, Hang, Han, Qiyang, and Sen, Bodhisattva. 2023. Inference for local parameters in convexity constrained models. *Journal of the American Statistical Association*, **118**, 2721–2735.

Depersin, Jules, and Lecué, Guillaume. 2022. Optimal robust mean and location estimation via convex programs with respect to any pseudo-norms. *Probability Theory and Related Fields*, **183**, 997–1025.

Devroye, Luc. 1987. *A Course in Density Estimation*. Birkhauser Boston Inc.

Devroye, Luc, and Györfi, Laszlo. 1985. *Nonparametric Density Estimation: The* L_1 *View*. John Wiley & Sons.

Published online by Cambridge University Press

Devroye, Luc, Lerasle, Matthieu, Lugosi, Gabor, and Oliveira, Roberto I. 2016. Sub-Gaussian mean estimators. *Annals of Statistics*, **44**, 2695–2725.

Devroye, Luc P, and Wise, Gary L. 1979. On the recovery of discrete probability densities from imperfect measurements. *Journal of the Franklin Institute*, **307**, 1–20.

Dharmadhikari, Sudhakar, and Joag-Dev, Kumar. 1988. *Unimodality, Convexity, and Applications*. Elsevier.

Donoho, David L. 1994. Statistical estimation and optimal recovery. *Annals of Statistics*, **22**, 238–270.

Donoho, David L. 2006. For most large underdetermined systems of linear equations, the minimal ℓ_1 solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, **59**, 797–829.

Donoho, David L, Maleki, Arian, and Montanari, Andrea. 2009. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, **106**, 18914–18919.

Doob, Joseph L. 2012. *Measure Theory*. Springer Science+Business Media.

Doss, Charles R, and Wellner, Jon A. 2016. Global rates of convergence of the MLEs of log-concave and s-concave densities. *Annals of Statistics*, **44**, 954–981.

Doss, Charles R, and Wellner, Jon A. 2019. Inference for the mode of a log-concave density. *Annals of Statistics*, **47**, 2950–2976.

Doss, Charles R, and Weng, Guangwei. 2018. Bandwidth selection for kernel density estimators of multivariate level sets and highest density regions. *Electronic Journal of Statistics*, **12**, 4313–4376.

Dossal, Charles. 2012. A necessary and sufficient condition for exact sparse recovery by ℓ_1 minimization. *Comptes Rendus Mathematique*, **350**, 117–120.

Duchi, John, and Haque, Saminul. 2024. An information-theoretic lower bound in time-uniform estimation. Pages 1486–1500 of: Agrawal, Shipra, and Roth, Aaron (eds), *Proceedings of Thirty Seventh Conference on Learning Theory*, vol. 247.

Duchi, John C, Jordan, Michael I, and Wainwright, Martin J. 2013. Local privacy and statistical minimax rates. Pages 429–438 of: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE.

Dudley, Richard M. 2002. *Real Analysis and Probability*. Cambridge University Press.

Dümbgen, Lutz. 2024. Shape-constrained statistical inference. *Annual Review of Statistics and its Application*, **11**, 373–391.

Dümbgen, Lutz, and Rufibach, Kaspar. 2009. Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, **15**, 40–68.

Dümbgen, Lutz, and Rufibach, Kaspar. 2011. logcondens: Computations related to univariate log-concave density estimation. *Journal of Statistical Software*, **39**, 1–28.

Dümbgen, Lutz, Hüsler, André, and Rufibach, Kaspar. 2007a. Active set and EM algorithms for log-concave densities based on complete and censored data. *arXiv:0707.4643*.

Dümbgen, Lutz, Rufibach, Kaspar, and Wellner, Jon A. 2007b. Marshall's lemma for convex density estimation. *IMS Lecture Notes Monograph Series, Vol. 55*, 101–107.

Dümbgen, Lutz, Samworth, Richard, and Schuhmacher, Dominic. 2011. Approximation by log-concave distributions, with applications to regression. *Annals of Statistics*, **39**, 702–730.

Dümbgen, Lutz, Samworth, Richard J, and Schuhmacher, Dominic. 2013. Stochastic search for semiparametric linear regression models. Pages 78–90 of: Banerjee, M, Bunea, F, Huang, J, Koltchinskii, V, and Maathuis, MH (eds), *From Probability to Statistics and Back: High-Dimensional Models and Processes – A Festschrift in Honor of Jon A. Wellner*. Institute of Mathematical Statistics Collections, Beechwood, Ohio.

Dümbgen, Lutz, Rufibach, Kaspar, and Schuhmacher, Dominic. 2014. Maximum-likelihood estimation of a log-concave density based on censored data. *Electronic Journal of Statistics*, **8**, 1405–1437.

Published online by Cambridge University Press

Durot, Cécile. 2007. On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Annals of Statistics*, **35**, 1080–1104.

Durot, Cécile, Kulikov, Vladimir N, and Lopuhaä, Hendrik P. 2012. The limit distribution of the L_∞-error of Grenander-type estimators. *Annals of Statistics*, **40**, 1578–1608.

Durrett, Rick. 2019. *Probability: Theory and Examples (5th Edition)*. Cambridge University Press.

Dvoretzky, Aryeh, Kiefer, Jack, and Wolfowitz, Jacob. 1956. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, **27**, 642–669.

Džeroski, Saso, and Ženko, Bernard. 2004. Is combining classifiers with stacking better than selecting the best one? *Machine Learning*, **54**, 255–273.

Eckart, Carl, and Young, Gale. 1936. The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.

Edner, H, Ragnarson, P, Svanberg, Sune, Wallinder, E, De Liso, A, Ferrara, R, and Maserti, BE. 1992. Differential absorption lidar mapping of atmospheric atomic mercury in Italian geothermal fields. *Journal of Geophysical Research: Atmospheres*, **97**, 3779–3786.

Edner, Hans, Faris, Gregory W, Sunesson, Anders, and Svanberg, Sune. 1989. Atmospheric atomic mercury monitoring using differential absorption lidar techniques. *Applied Optics*, **28**, 921–930.

Efron, Bradley. 1979. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

Efron, Bradley, Hastie, Trevor, Johnstone, Iain, and Tibshirani, Robert. 2004. Least angle regression. *Annals of Statistics (with discussion)*, **32**, 407–499.

Eilers, Paul HC, and Borgdorff, MW. 2007. Non-parametric log-concave mixtures. *Computational Statistics & Data Analysis*, **51**, 5444–5451.

El Ghaoui, Laurent, Viallon, Vivian, and Rabbani, Tarek. 2012. Safe feature elimination in sparse supervised learning. *Pacific Journal of Optimization*, **8**, 667–698.

Epanechnikov, Vassiliy A. 1969. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, **14**, 153–158.

Eriksson, Jan, and Koivunen, Visa. 2004. Identifiability, separability, and uniqueness of linear ICA models. *IEEE Signal Processing Letters*, **11**, 601–604.

Esseen, Carl-Gustav. 1942. On the Liapounoff limit error in the theory of probability. *Arkiv för Matematik, Astronomi och Fysik*, **28**, 1–19.

Eubank, Randall L. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker Inc.

Evans, Lawrence C, and Gariepy, Ronald F. 2018. *Measure Theory and Fine Properties of Functions*. Routledge.

Ezekiel, Mordecai. 1930. *Methods of Correlation Analysis*. John Wiley & Sons.

Fan, Jianqing, and Gijbels, Irene. 1996. *Local Polynomial Modelling and its Applications*. Chapman & Hall.

Fan, Jianqing, and Li, Runze. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

Fan, Jianqing, Samworth, Richard, and Wu, Yichao. 2009. Ultrahigh dimensional feature selection: Beyond the linear model. *Journal of Machine Learning Research*, **10**, 2013–2038.

Fang, Zheng. 2019. Refinements of the Kiefer–Wolfowitz theorem and a test of concavity. *Electronic Journal of Statistics*, **13**, 4596–4645.

Fano, Robert M. 1961. *Transmission of Information: A Statistical Theory of Communication*. MIT Press.

Farrell, RH. 1972. On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Annals of Mathematical Statistics*, **43**, 170–180.

Published online by Cambridge University Press

Fathi, Max, Goldstein, Larry, Reinert, Gesine, and Saumard, Adrien. 2022. Relaxing the Gaussian assumption in shrinkage and SURE in high dimension. *Annals of Statistics*, **50**, 2737–2766.

Feige, Uriel, and Krauthgamer, Robert. 2003. The probable value of the Lovász–Schrijver relaxations for maximum independent set. *SIAM Journal on Computing*, **32**, 345–370.

Feldman, Vitaly, Grigorescu, Elena, Reyzin, Lev, Vempala, Santosh S, and Xiao, Ying. 2017. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, **64**, 1–37.

Feller, Willy. 1936. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **40**, 521–559.

Feng, Oliver Y, Guntuboyina, Adityanand, Kim, Arlene KH, and Samworth, Richard J. 2021. Adaptation in multivariate log-concave density estimation. *Annals of Statistics*, **49**, 129–153.

Feng, Oliver Y, Chen, Yining, Han, Qiyang, Carroll, Raymond J, and Samworth, Richard J. 2022a. Nonparametric, tuning-free estimation of S-shaped functions. *Journal of the Royal Statistical Society, Series B*, **84**, 1324–1352.

Feng, Oliver Y, Chen, Yining, Han, Qiyang, Carroll, Raymond J, and Samworth, Richard J. 2022b. Sshaped: Estimation of an S-shaped function. *R package version 0.99*.

Feng, Oliver Y, Venkataramanan, Ramji, Rush, Cynthia, and Samworth, Richard J. 2022c. A Unifying Tutorial on Approximate Message Passing. *Foundations and Trends® in Machine Learning*, **15**, 335–536.

Feng, Oliver Y, Kao, Yu-Chun, Xu, Min, and Samworth, Richard J. 2026. Optimal convex M-estimation via score matching. *Annals of Statistics*, **54**, 408–441.

Fischer, Ernst. 1905. Über quadratische Formen mit reellen Koeffizienten. *Monatshefte für Mathematik und Physik*, **16**, 234–249.

Fix, Evelyn, and Hodges, Joseph Lawson. 1951. Nonparametric discrimination: Consistency properties. *Randolph Field, Texas, Project*, 21–49.

Fix, Evelyn, and Hodges, Joseph Lawson. 1989. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, **57**, 238–247.

Fokianos, Konstantinos, Leucht, Anne, and Neumann, Michael H. 2020. On integrated L_1 convergence rate of an isotonic regression estimator for multivariate observations. *IEEE Transactions on Information Theory*, **66**, 6389–6402.

Folland, Gerald B. 1999. *Real Analysis: Modern Techniques and their Applications*. John Wiley & Sons.

Fraser, Donald AS, and Massam, Hélene. 1989. A mixed primal-dual bases algorithm for regression under inequality constraints. Application to concave regression. *Scandinavian Journal of Statistics*, **16**, 65–74.

Fréchet, Maurice. 1906. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884–1940)*, **22**, 1–72.

Fréchet, Maurice. 1907. Sur les ensembles de fonctions et les opérations linéaires. *Comptes rendus de l'Académie des Sciences (in French)*, **144**, 1414–1416.

Fremlin, David Heaver. 2000. *Measure Theory*. Torres Fremlin.

Fresen, Daniel. 2013. A multivariate Gnedenko law of large numbers. *Annals of Probability*, **41**, 3051–3080.

Friedman, Jerome, Hastie, Trevor, Höfling, Holger, and Tibshirani, Robert. 2007. Pathwise coordinate optimization. *Annals of Applied Statistics*, **1**, 302–332.

Friedman, Jerome, Hastie, Trevor, and Tibshirani, Robert. 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432–441.

Friedman, Jerome, Hastie, Trevor, and Tibshirani, Rob. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1.

Published online by Cambridge University Press

Frisch, Ragnar, and Waugh, Frederick V. 1933. Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, **1**, 387–401.

Fu, Wenjiang J. 1998. Penalized regressions: The Bridge versus the Lasso. *Journal of Computational and Graphical Statistics*, **7**, 397–416.

Fuchs, Jean-Jacques. 2005. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Transactions on Information Theory*, **51**, 3601–3608.

Gauss, Carl Friedrich. 1855. *Méthode des moindres carrés: Mémoires sur la combinaison des observations*. Mallet-Bachelier.

Geisser, Seymour. 1975. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, **70**, 320–328.

Gerchinovitz, Sebastien, Ménard, Pierre, and Stoltz, Gilles. 2020. Fano's inequality for random variables. *Statistical Science*, **35**, 178–201.

Gibbs, Josiah Willard. 1902. *Elementary Principles in Statistical Mechanics*. Yale University Press.

Gibbs, Mark N, and MacKay, David JC. 2000. Variational Gaussian process classifiers. *IEEE Transactions on Neural Networks*, **11**, 1458–1464.

Gilbert, Edgar N. 1952. A comparison of signalling alphabets. *Bell System Technical Journal*, **31**, 504–522.

Ginsberg, William. 1974. The multiplant firm with increasing returns to scale. *Journal of Economic Theory*, **9**, 283–292.

Giraud, Christophe. 2021. *Introduction to High-Dimensional Statistics*. CRC Press.

Glivenko, Valery. 1933. Sulla determinazione empirica delle leggi di probabilita. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 92–99.

Goeman, Jelle J, and Solari, Aldo. 2011. Multiple testing for exploratory research. *Statistical Science*, **26**, 584–597.

Goldenshluger, Alexander, and Lepski, Oleg. 2014. On adaptive minimax density estimation on R^d. *Probability Theory and Related Fields*, **159**, 479–543.

Golub, Gene H, and Van Loan, Charles F. 2013. *Matrix Computations*. Johns Hopkins University Press.

Golub, Gene H, Heath, Michael, and Wahba, Grace. 1979. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, **21**, 215–223.

Gray, JD. 1984. The shaping of the Riesz representation theorem: A chapter in the history of analysis. *Archive for History of Exact Sciences*, **31**, 127–187.

Green, Peter J, and Silverman, Bernard W. 1994. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. CRC Press.

Greenshtein, Eitan, and Ritov, Ya'Acov. 2004. Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli*, **10**, 971–988.

Grenander, Ulf. 1956. On the theory of mortality measurement: Part ii. *Scandinavian Actuarial Journal*, **39**, 125–153.

Grimmett, Geoffrey, and Stirzaker, David. 2020. *Probability and Random Processes (4th Edition)*. Oxford University Press.

Groeneboom, Piet. 1985. Estimating a monotone density. Pages 539–555 of: Le Cam, Lucien M, and Olshen, Richard (eds), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Volume II (Berkeley, Calif., 1983)*. Wadsworth.

Groeneboom, Piet, and Jongbloed, Geurt. 2014. *Nonparametric Estimation under Shape Constraints*. Cambridge University Press.

Groeneboom, Piet, and Jongbloed, Geurt. 2015. Nonparametric confidence intervals for monotone functions. *Annals of Statistics*, **43**, 2019–2054.

Groeneboom, Piet, Hooghiemstra, Gerard, and Lopuhaä, Hendrik P. 1999. Asymptotic normality of the L_1 error of the Grenander estimator. *Annals of Statistics*, **27**, 1316–1347.

Groeneboom, Piet, Jongbloed, Geurt, and Wellner, Jon A. 2001. Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics*, **29**, 1653–1698.

Published online by Cambridge University Press

Guntuboyina, Adityanand, and Sen, Bodhisattva. 2015. Global risk bounds and adaptation in univariate convex regression. *Probability Theory and Related Fields*, **163**, 379–411.

Guntuboyina, Adityanand, and Sen, Bodhisattva. 2018. Nonparametric shape-restricted regression. *Statistical Science*, **33**, 568–594.

Gut, Allan. 2005. *Probability: A Graduate Course*. Springer.

Györfi, László, Kohler, Michael, Krzyzak, Adam, and Walk, Harro. 2002. *A Distribution-Free Theory of Nonparametric Regression*. Springer.

Hájek, Jaroslav. 1970. A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **14**, 323–330.

Hájek, Jaroslav. 1972. Local asymptotic minimax and admissibility in estimation. Pages 175–194 of: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1.

Hald, Anders. 1998. *A History of Mathematical Statistics from 1750 to 1930*. Wiley Series in Probability and Statistics.

Hall, Peter, and Kang, Kee-Hoon. 2005. Bandwidth choice for nonparametric classification. *Annals of Statistics*, **33**, 284–306.

Hall, Peter, and Marron, James Stephen. 1987. Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probability Theory and Related Fields*, **74**, 567–581.

Halmos, Paul R. 2013. *Measure Theory*. Springer.

Han, Qiyang, and Wellner, Jon A. 2016. Approximation and estimation of s-concave densities via Rényi divergences. *Annals of Statistics*, **44**, 1332–1359.

Han, Qiyang, Wang, Tengyao, Chatterjee, Sabyasachi, and Samworth, Richard J. 2019. Isotonic regression in general dimensions. *Annals of Statistics*, **47**, 2440–2471.

Hansen, Bruce E. 2022. A modern Gauss–Markov theorem. *Econometrica*, **90**, 1283–1294.

Härdle, Wolfgang. 1990. *Applied Nonparametric Regression*. Cambridge University Press.

Hart, Jeffrey. 1997. *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer Science+Business Media.

Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome H, and Friedman, Jerome H. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hastie, Trevor, Tibshirani, Robert, and Wainwright, Martin. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.

Hausdorff, Felix. 1914. *Grundzüge der Mengenlehre*. Leipzig: Veit & Comp.

Hellinger, Ernst. 1909. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die Reine und Angewandte Mathematik*, **136**, 210–271.

Henningsson, Toivo, and Åström, Karl Johan. 2006. Log-concave observers. In: *17th International Symposium on Mathematical Theory of Networks and Systems*.

Hildreth, Clifford. 1954. Point estimates of ordinates of concave functions. *Journal of the American Statistical Association*, **49**, 598–619.

Hiriart-Urruty, Jean-Baptiste, and Lemaréchal, Claude. 2004. *Fundamentals of Convex Analysis*. Springer Science+Business Media.

Hochberg, Yosef. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**, 800–802.

Hoeffding, Wassily. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**, 13–30.

Hoerl, Arthur E, and Kennard, Robert W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.

Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Holst, Ulla, Hössjer, Ola, Björklund, Claes, Ragnarson, Pär, and Edner, Hans. 1996. Locally weighted least squares kernel regression and statistical evaluation of LIDAR measurements. *Environmetrics*, **7**, 401–416.

Published online by Cambridge University Press

Hopkins, Samuel B. 2020. Mean estimation with sub-Gaussian rates in polynomial time. *Annals of Statistics*, **48**, 1193–1213.

Hore, Rohan, Soloff, Jake A, Barber, Rina Foygel, and Samworth, Richard J. 2026+. Testing conditional independence under isotonicity. *arXiv:2501.06133v3*.

Horn, Roger A, and Johnson, Charles R. 2013. *Matrix Analysis*. Cambridge University Press.

Hu, Inchi. 1985. A uniform bound for the tail probability of Kolmogorov–Smirnov statistics. *Annals of Statistics*, **13**, 821–826.

Hunter, David R, Wang, Shaoli, and Hettmansperger, Thomas P. 2007. Inference for mixtures of symmetric distributions. *Annals of Statistics*, **35**, 224–251.

Hyndman, Rob J. 1996. Computing and graphing highest density regions. *The American Statistician*, **50**, 120–126.

Hyvärinen, Aapo, Karhunen, Juha, and Oja, Erkki. 2001. *Independent Component Analysis*. John Wiley & Sons.

Ibragimov, IA, and Has'minskii, RZ. 1981. *Statistical Estimation: Asymptotic Theory*. Springer Science+Business Media.

Ibragimov, Il'dar Abdullovich. 1956. On the composition of unimodal distributions. *Theory of Probability & Its Applications*, **1**, 255–260.

Ising, E. 1925. Beitrag zur Theorie der Ferromagnetismus. *Zeitschrift für Physik*, **31**, 253–258.

James, W, and Stein, Charles. 1961. Estimation with quadratic loss. Pages 361–379 of: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I. University of California Press.

Janková, Jana, Shah, Rajen D, Bühlmann, Peter, and Samworth, Richard J. 2020. Goodness-of-fit testing in high-dimensional generalized linear models. *Journal of the Royal Statistical Society, Series B*, **82**, 773–795.

Jankowski, Hanna. 2014. Convergence of linear functionals of the Grenander estimator under misspecification. *Annals of Statistics*, **42**, 625–653.

Jerrum, Mark. 1992. Large cliques elude the Metropolis process. *Random Structures & Algorithms*, **3**, 347–359.

Jones, Frank. 2001. *Lebesgue Integration on Euclidean Space*. Jones & Bartlett Learning.

Jordan, Camille. 1874. Mémoire sur les formes bilinéaires. *Journal de Mathématiques Pures et Appliquées*, **19**, 35–54.

Kappel, Franz, and Kuntsevich, Alexei V. 2000. An implementation of Shor's r-algorithm. *Computational Optimization and Applications*, **15**, 193–205.

Katkovnik, V Ya. 1979. Linear and nonlinear methods of nonparametric regression analysis. *Soviet Automatic Control*, **5**, 25–34.

Keener, Robert W. 2010. *Theoretical Statistics: Topics for a Core Course*. Springer Science+Business Media.

Kelley, John L. 2017. *General Topology*. Courier Dover Publications.

Kemperman, Johannes HB. 1969. On the optimum rate of transmitting information. *Annals of Mathematical Statistics*, **40**, 2156–2177.

Kendall, Maurice George, and Stuart, Alan. 1961. *The Advanced Theory of Statistics: Inference and Relationship, Volume 2*. Griffin, London.

Kiefer, Jack. 1957. Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics*, **28**, 573–601.

Kiefer, Jack, and Wolfowitz, J. 1976. Asymptotically minimax estimation of concave and convex distribution functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **34**, 73–85.

Kim, Arlene KH. 2020. Obtaining minimax lower bounds: a review. *Journal of the Korean Statistical Society*, **49**, 673–701.

Kim, Arlene KH, and Samworth, Richard J. 2016. Global rates of convergence in log-concave density estimation. *Annals of Statistics*, **44**, 2756–2779.

Published online by Cambridge University Press

Kim, Arlene KH, Guntuboyina, Adityanand, and Samworth, Richard J. 2018. Adaptation in log-concave density estimation. *Annals of Statistics*, **46**, 2279–2306.

Kimeldorf, George, and Wahba, Grace. 1971. Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**, 82–95.

Kline, Morris. 1990. *Mathematical Thought from Ancient to Modern Times: Volume 2*. Oxford University Press.

Koenker, Roger, and Mizera, Ivan. 2010. Quasi-concave density estimation. *Annals of Statistics*, **38**, 2998–3027.

Kolmogorov, Andrey. 1933. *Grundbegriffe der wahrscheinlichkeitsrechnung*. Julius Springer, Berlin.

Koltchinskii, Vladimir. 2006. Local Rademacher complexities and oracle inequalities in risk minimization. *Annals of Statistics*, **34**, 2593–2656.

Korostelev, Aleksandr P, and Tsybakov, Alexandre B. 1993. *Minimax Theory of Image Reconstruction*. Springer.

Krishnapuram, Balaji, Carin, Lawrence, Figueiredo, Mário AT, and Hartemink, Alexander J. 2005. Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 957–968.

Kučera, Luděk. 1995. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, **57**, 193–212.

Kulikov, Vladimir N, and Lopuhaä, Hendrik P. 2005. Asymptotic normality of the L_k-error of the Grenander estimator. *Annals of Statistics*, **33**, 2228–2255.

Kullback, Solomon. 1967. A lower bound for discrimination information in terms of variation. *IEEE Transactions on Information Theory*, **13**, 126–127.

Kullback, Solomon, and Leibler, Richard A. 1951. On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Kur, Gil, Dagan, Yuval, and Rakhlin, Alexander. 2019. Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. *arXiv:1903.05315*.

Kur, Gil, Gao, Fuchang, Guntuboyina, Adityanand, and Sen, Bodhisattva. 2024. Convex regression in multidimensions: Suboptimality of least squares estimators. *Annals of Statistics*, **52**, 2791–2815.

Lacour, Claire, Massart, Pascal, and Rivoirard, Vincent. 2017. Estimator selection: a new method with applications to kernel density estimation. *Sankhyā A*, **79**, 298–335.

Laplace, Pierre Simon. 1812. *Théorie Analytique des Probabilités*. Courcier, Paris.

Lauritzen, SL, and Wermuth, N. 1989. Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, **17**, 31–57.

Lauritzen, SL, Dawid, AP, Larsen, BN, and Leimer, HG. 1990. Independence properties of directed Markov fields. *Networks*, **20**, 491–505.

Lauritzen, Steffen L. 1996. *Graphical Models*. Oxford University Press.

Lauritzen, Steffen L. 2020. *Lectures on Graphical Models*. University of Copenhagen.

Lauritzen, Steffen L, and Spiegelhalter, David J. 1988. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, **50**, 157–194.

Lax, Peter D. 2002. *Functional Analysis*. John Wiley & Sons.

Le Cam, Lucien. 1953. *On some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates*. Ph.D. thesis, University of California, Berkeley.

Le Cam, Lucien. 1973. Convergence of estimates under dimensionality restrictions. *Annals of Statistics*, **1**, 38–53.

Le Cam, Lucien. 1986. The central limit theorem around 1935. *Statistical Science*, **1**, 78–91.

Lei, Jing. 2020a. Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli*, **26**, 767–798.

Lei, Jing. 2020b. Cross-validation with confidence. *Journal of the American Statistical Association*, **115**, 1978–1997.

Published online by Cambridge University Press

Lei, Lihua, and Wooldridge, Jeffrey. 2022. What estimators are unbiased for linear models? *arXiv:2212.14185*.

Lepskiĭ, Oleg V. 1990. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability & Its Applications*, **35**, 454–466.

Lepskiĭ, Oleg V. 1991. Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory of Probability & Its Applications*, **36**, 682–697.

Leskelä, Lasse, and Välimaa, Ian. 2025. Sub-Poisson distributions: Concentration inequalities, optimal variance proxies, and closure properties. *arXiv:2508.12103*.

Li, Chi-Kwong, and Mathias, Roy. 1999. The Lidskii–Mirsky–Wielandt theorem – additive and multiplicative versions. *Numerische Mathematik*, **81**, 377–413.

Liao, Xiyue, and Meyer, Mary C. 2017. Change-point estimation using shape-restricted regression splines. *Journal of Statistical Planning and Inference*, **188**, 8–21.

Liapunov, Aleksandr Michajlovič. 1901. Nouvelle forme du théorème sur la limite de probabilité. *Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg*, **12**, 1–24.

Lidskii, VB. 1950. On the characteristic numbers of the sum and product of symmetric matrices. *Doklady Akademii Nauk SSSR (NS)*, **75**, 769–772.

Lindeberg, Jarl Waldemar. 1922. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **15**, 211–225.

Lindley, Dennis V. 1956. On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, **27**, 986–1005.

Liu, Yu, and Wang, Yong. 2018. A fast algorithm for univariate log-concave density estimation. *Australian & New Zealand Journal of Statistics*, **60**, 258–275.

Loader, Clive. 1999. *Local Regression and Likelihood*. Springer Science+Business Media.

Loh, Po-Ling, and Wainwright, Martin J. 2015. Regularized M-estimators with nonconvexity: statistical and algorithmic theory for local optima. *Journal of Machine Learning Research*, **16**, 559–616.

Lovász, László, and Vempala, Santosh. 2007. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, **30**, 307–358.

Lovell, Michael C. 1963. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, **58**, 993–1010.

Lovell, Michael C. 2008. A simple proof of the FWL theorem. *Journal of Economic Education*, **39**, 88–91.

Lugosi, Gábor, and Mendelson, Shahar. 2019. Sub-Gaussian estimators of the mean of a random vector. *Annals of Statistics*, **47**, 783–794.

Lundborg, Anton Rask, Shah, Rajen D, and Peters, Jonas. 2022. Conditional independence testing in Hilbert spaces with applications to functional data analysis. *Journal of the Royal Statistical Society, Series B*, **84**, 1821–1850.

Lundborg, Anton Rask, Kim, Ilmun, Shah, Rajen D, and Samworth, Richard J. 2024. The projected covariance measure for assumption-lean variable significance testing. *Annals of Statistics*, **52**, 2851–2878.

Luss, Ronny, Rosset, Saharon, and Shahar, Moni. 2012. Efficient regularized isotonic regression with application to gene–gene interaction search. *Annals of Applied Statistics*, **6**, 253–283.

Ma, Tianyi, Verchand, Kabir A, Berrett, Thomas B, Wang, Tengyao, and Samworth, Richard J. 2026+a. Estimation beyond missing (completely) at random. *Annals of Statistics, to appear. arXiv:2410.10704*.

Ma, Tianyi, Kabir, Verchand, and Samworth, Richard J. 2026+b. High-probability minimax lower bounds. *Statistical Science, to appear. arXiv:2406.13447*.

Ma, Zongming, and Wu, Yihong. 2015. Computational barriers in minimax submatrix detection. *Annals of Statistics*, **43**, 1089–1116.

Mammen, Enno. 1991. Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics*, **19**, 741–759.

Published online by Cambridge University Press

Marcus, Ruth, Eric, Peritz, and Gabriel, K Ruben. 1976. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, **63**, 655–660.

Mardia, Kanti V, Kent, John T, and Bibby, John M. 1980. *Multivariate Analysis*. Academic Press.

Markov, Andre A. 1912. *Wahrscheinlichkeitsrechnung (translated by H. Liebmann)*. B. G. Teubner.

Marron, JS, and Wand, MP. 1992. Exact mean integrated squared error. *Annals of Statistics*, **20**, 712–736.

Marshall, AW. 1970. Discussion of Barlow and van Zwet's paper. Pages 174–176 of: Puri, ML (ed), *Nonparametric Techniques in Statistical Inference. Proceedings of the First International Symposium on Nonparametric Techniques held at Indiana University, June 1969*. Cambridge University Press.

Massart, Pascal. 1990. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Annals of Probability*, **18**, 1269–1283.

Matioli, LC, Santos, SR, Kleina, M, and Leite, EA. 2018. A new algorithm for clustering based on kernel density estimation. *Journal of Applied Statistics*, **45**, 347–366.

Matzkin, Rosa L. 1991. Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. *Econometrica*, **59**, 1315–1327.

Mazumder, Rahul, Choudhury, Arkopal, Iyengar, Garud, and Sen, Bodhisattva. 2019. A computational framework for multivariate convex regression and its variants. *Journal of the American Statistical Association*, **114**, 318–331.

Meinshausen, Nicolai. 2007. Relaxed Lasso. *Computational Statistics and Data Analysis*, **52**, 374–393.

Meinshausen, Nicolai. 2008. Hierarchical testing of variable importance. *Biometrika*, **95**, 265–278.

Meinshausen, Nicolai, and Bühlmann, Peter. 2006. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, **34**, 1436–1462.

Meinshausen, Nicolai, and Bühlmann, Peter. 2010. Stability selection. *Journal of the Royal Statistical Society, Series B (with discussion)*, **72**, 417–473.

Mendelson, Shahar. 2002. Geometric parameters of kernel machines. Pages 29–43 of: *International Conference on Computational Learning Theory*. Springer.

Mercer, James. 1909. XVI. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A*, **209**, 415–446.

Meyer, Mary, and Woodroofe, Michael. 2000. On the degrees of freedom in shape-restricted regression. *Annals of Statistics*, **28**, 1083–1104.

Meyer, Mary C. 1999. An extension of the mixed primal–dual bases algorithm to the case of more constraints than dimensions. *Journal of Statistical Planning and Inference*, **81**, 13–31.

Mirsky, Leon. 1960. Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, **11**, 50–59.

Moonen, Marc, and De Moor, Bart. 1995. *SVD and Signal Processing, III: Algorithms, Architectures and Applications*. Elsevier.

Moore, Eliakim H. 1920. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, **26**, 394–395.

Moore, Eliakim H. 1935. *General Analysis, Part I*. Memoirs of the American Philosophical Society. American Philosophical Society. Edited by RW Barnard.

Moore, Eliakim H. 1939. *General Analysis, Part II*. Memoirs of the American Philosophical Society. American Philosophical Society. Edited by RW Barnard.

Muirhead, Robb J. 2009. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons.

Müller, Manuel M, Reeve, Henry WJ, Cannings, Timothy I, and Samworth, Richard J. 2025. Isotonic subgroup selection. *Journal of the Royal Statistical Society, Series B*, **87**, 132–156.

Published online by Cambridge University Press

Müller, Samuel, and Rufibach, Kaspar. 2009. Smooth tail-index estimation. *Journal of Statistical Computation and Simulation*, **79**, 1155–1167.
Munkres, James R. 2000. *Topology*. Prentice Hall.
Nadaraya, Elizbar A. 1964. On estimating regression. *Theory of Probability & Its Applications*, **9**, 141–142.
Neapolitan, Richard E. 1990. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. John Wiley & Sons.
Neykov, Matey, Balakrishnan, Sivaraman, and Wasserman, Larry. 2021. Minimax optimal conditional independence testing. *Annals of Statistics*, **49**, 2151–2177.
Neyman, J. 1923. On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (translated). *Statistical Science*, **5**, 465–480.
Ning, Yang, and Liu, Han. 2017. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Annals of Statistics*, **45**, 158–195.
Niu, Ziang, Chakraborty, Abhinav, Dukes, Oliver, and Katsevich, Eugene. 2024. Reconciling model-X and doubly robust approaches to conditional independence testing. *The Annals of Statistics*, **52**, 895–921.
Osborne, Michael R, Presnell, Brett, and Turlach, Berwin A. 2000a. A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, **20**, 389–403.
Osborne, Michael R, Presnell, Brett, and Turlach, Berwin A. 2000b. On the lasso and its dual. *Journal of Computational and Graphical Statistics*, **9**, 319–337.
O'Searcoid, Mícheál. 2006. *Metric Spaces*. Springer Science+Business Media.
Pananjady, Ashwin, and Samworth, Richard J. 2022. Isotonic regression with unknown permutations: Statistics, computation and adaptation. *Annals of Statistics*, **50**, 324–350.
Pardo, Leandro. 2006. *Statistical Inference based on Divergence Measures*. Chapman and Hall/CRC.
Parzen, Emanuel. 1962. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**, 1065–1076.
Patilea, Valentin. 2001. Convex models, MLE and misspecification. *Annals of Statistics*, **29**, 94–123.
Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*. Morgan Kaufmann.
Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
Pearl, Judea, and Paz, Azaria. 1985. *Graphoids: A graph-based logic for reasoning about relevancy relations*. Tech. rept. CSD-850038. Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles.
Penrose, Roger. 1955. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, **51**, 406–413.
Peters, Jonas, Janzing, Dominik, and Schölkopf, Bernhard. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning. MIT Press.
Petrov, Valentin V. 1995. *Limit Theorems of Probability Theory*. Oxford University Press.
Pinsker, Mark S. 1964. *Information and Information Stability of Random Variables and Processes*. Holden-Day.
Plackett, Robin L. 1949. A historical note on the method of least squares. *Biometrika*, **36**, 458–460.
Plancherel, Michel. 1910. Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales définies. *Rendiconti del Circolo Matematico di Palermo*, **30**, 289–335.
Polyanskiy, Yury, and Wu, Yihong. 2023. *Information Theory: From Coding to Learning*. Cambridge University Press.

Published online by Cambridge University Press

Pötscher, Benedikt M, and Preinerstorfer, David. 2022. A Modern Gauss–Markov Theorem? Really? *arXiv:2203.01425*.

Prakasa Rao, BLS. 1969. Estimation of a unimodal density. *Sankhyā: The Indian Journal of Statistics, Series A*, **31**, 23–36.

Prékopa, András. 1973. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, **34**, 335–343.

Prékopa, András. 1980. Logarithmic concave measures and related topics. In: *Stochastic Programming*. Citeseer.

Pya, Natalia. 2022. scam: Shape Constrained Additive Models. *R package version 1.2-13*.

Pya, Natalya, and Wood, Simon N. 2015. Shape constrained additive models. *Statistics and Computing*, **25**, 543–559.

Quenouille, Maurice H. 1949. Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society, Series B*, **11**, 68–84.

Rado, Richard. 1956. Note on generalized inverses of matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, **52**, 600–601.

Rahimi, Ali, and Recht, Benjamin. 2007. Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, **20**.

Raman-Sundström, Manya. 2015. A pedagogical history of compactness. *The American Mathematical Monthly*, **122**, 619–635.

Rao, Calyampudi Radhakrishna. 1965. *Linear Statistical Inference and is Applications*. John Wiley & Sons.

Rao, Malempati Madhusudana. 2018. *Measure Theory and Integration*. CRC Press.

Raskutti, Garvesh, Wainwright, Martin J, and Yu, Bin. 2010. Restricted eigenvalue properties for correlated Gaussian designs. *Journal of Machine Learning Research*, **11**, 2241–2259.

Rathke, Fabian, and Schnörr, Christoph. 2019. Fast multivariate log-concave density estimation. *Computational Statistics & Data Analysis*, **140**, 41–58.

Reeve, Henry WJ. 2024. A short proof of the Dvoretzky–Kiefer–Wolfowitz–Massart inequality. *arXiv:2403.16651*.

Ren, Zhao, Sun, Tingni, Zhang, Cun-Hui, and Zhou, Harrison H. 2015. Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Annals of Statistics*, **43**, 991–1026.

Richardson, Thomas S, and Robins, James M. 2013. *Single World Intervention Graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality*. CSSS Working Paper 128. Center for Statistics and the Social Sciences, University of Washington.

Riesz, Frigyes. 1907. Sur une espèce de géométrie analytique des systèmes de fonctions sommables. *Comptes rendus de l'Académie des Sciences*, **144**, 1409–1411.

Roberts, A Wayne, and Varberg, Dale E. 1974. *Convex Functions*. Academic Press.

Robertson, Tim, Wright, FT, and Dykstra, RL. 1988. *Order Restricted Statistical Inference*. John Wiley & Sons.

Rockafellar, R Tyrrell. 1997. *Convex Analysis*. Princeton University Press.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 832–837.

Rosset, Saharon, Zhu, Ji, and Hastie, Trevor. 2004. Boosting as a regularized path to a maximum margin classifier. *Journal of Machine Learning Research*, **5**, 941–973.

Rothman, Adam J, Bickel, Peter J, Levina, Elizaveta, and Zhu, Ji. 2008. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, **2**, 494–515.

Royden, Halsey Lawrence, and Fitzpatrick, Patrick. 1988. *Real Analysis*. Macmillan.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**, 688–701.

Rudelson, Mark, and Zhou, Shuheng. 2013. Reconstruction from anisotropic random measurements. *IEEE Transactions on Information Theory*, **59**, 3434–3447.

Published online by Cambridge University Press

Rudemo, Mats. 1982. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65–78.

Rudin, Walter. 1987. *Real and Complex Analysis*. 3rd edn. McGraw-Hill.

Rudin, Walter. 1991. *Functional Analysis*. McGraw-Hill.

Rufibach, Kaspar, and Dümbgen, Lutz. 2010. logcondens: Estimate a log-concave probability density from iid observations. *R package version 2.1.8*.

Salsburg, David. 2001. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. W. H. Freeman and Co.

Samworth, Richard. 2005. Small confidence sets for the mean of a spherically symmetric distribution. *Journal of the Royal Statistical Society, Series B*, **67**, 343–361.

Samworth, Richard J. 2012. Stein's paradox. *Eureka*, **62**, 38–41.

Samworth, Richard J. 2018. Recent progress in log-concave density estimation. *Statistical Science*, **33**, 493–509.

Samworth, Richard J. 2026+. Nonparametric inference under shape constraints: past, present and future. *Proceedings of the International Congress of Mathematicians 2026, to appear. arXiv:2509.26040*.

Samworth, Richard J, and Sen, Bodhisattva. 2018. Editorial: Special Issue on "Nonparametric Inference under Shape Constraints". *Statistical Science*, **33**, 469–472.

Samworth, Richard J, and Wand, Matthew P. 2010. Asymptotics and optimal bandwidth selection for highest density region estimation. *Annals of Statistics*, **38**, 1767–1792.

Samworth, Richard J, and Yuan, Ming. 2012. Independent component analysis via nonparametric maximum likelihood estimation. *Annals of Statistics*, **40**, 2973–3002.

Santosa, Fadil, and Symes, William W. 1986. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, **7**, 1307–1330.

Saulis, Leonas, and Statulevicius, VA. 1991. *Limit Theorems for Large Deviations*. Springer Science+Business Media.

Saumard, Adrien, and Wellner, Jon A. 2014. Log-concavity and strong log-concavity: a review. *Statistics Surveys*, **8**, 45–114.

Schechter, Samuel. 1959. On the inversion of certain matrices. *Mathematical Tables and Other Aids to Computation*, **13**, 73–77.

Scheidegger, Cyrill, Hörrmann, Julia, and Bühlmann, Peter. 2022. The weighted generalised covariance measure. *Journal of Machine Learning Research*, **23**, 1–68.

Schell, Michael J, and Singh, Bahadur. 1997. The reduced monotonic regression method. *Journal of the American Statistical Association*, **92**, 128–135.

Schneider, Rolf. 2014. *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge University Press.

Schumaker, Larry. 2007. *Spline Functions: Basic Theory*. Cambridge University Press.

Scott, David W. 2015. *Multivariate Density Estimation: Theory, Practice, and Visualization (2nd edition)*. John Wiley & Sons.

Seal, Hilary L. 1967. Studies in the History of Probability and Statistics. XV The historical development of the Gauss linear model. *Biometrika*, **54**, 1–24.

Seijo, Emilio, and Sen, Bodhisattva. 2011. Nonparametric least squares estimation of a multivariate convex regression function. *Annals of Statistics*, **39**, 1633–1657.

Sen, Bodhisattva, Banerjee, Moulinath, and Woodroofe, Michael. 2010. Inconsistency of bootstrap: The Grenander estimator. *Annals of Statistics*, **38**, 1953–1977.

Seregin, Arseni, and Wellner, Jon A. 2010. Nonparametric estimation of multivariate convex-transformed densities. *Annals of Statistics*, **38**, 3751–3781.

Shah, Rajen D, and Bühlmann, Peter. 2018. Goodness-of-fit tests for high dimensional linear models. *Journal of the Royal Statistical Society, Series B*, **80**, 113–135.

Shah, Rajen D, and Bühlmann, Peter. 2023. Double-estimation-friendly inference for high-dimensional misspecified models. *Statistical Science*, **38**, 68–91.

Published online by Cambridge University Press

Shah, Rajen D, and Meinshausen, Nicolai. 2018. On b-bit min-wise hashing for large-scale regression and classification with sparse data. *Journal of Machine Learning Research*, **18**, 1–42.

Shah, Rajen D, and Peters, Jonas. 2020. The hardness of conditional independence testing and the generalised covariance measure. *Annals of Statistics*, **48**, 1514–1538.

Shah, Rajen D, and Samworth, Richard J. 2013a. Discussion of 'Correlated variables in regression: Clustering and sparse estimation' by Peter Bühlmann, Philipp Rütimann, Sara van de Geer and Cun-Hui Zhang. *Journal of Statistical Planning and Inference*, **143**, 1866–1868.

Shah, Rajen D, and Samworth, Richard J. 2013b. Variable selection with error control: another look at stability selection. *Journal of the Royal Statistical Society, Series B*, **75**, 55–80.

Shannon, Claude Elwood. 1948. A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423.

Shao, Peter Yi-Shi, and Strawderman, William E. 1994. Improving on the James–Stein positive-part estimator. *Annals of Statistics*, **22**, 1517–1538.

Shevade, Shirish Krishnaj, and Keerthi, S Sathiya. 2003. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, **19**, 2246–2253.

Shevtsova, Irina. 2011. On the absolute constants in the Berry–Esseen type inequalities for identically distributed summands. *arXiv:1111.6554*.

Shirali, Satish, and Vasudeva, Harkrishan Lal. 2005. *Metric Spaces*. Springer Science+Business Media.

Shorack, Galen R, and Wellner, Jon A. 1986. *Empirical Processes with Applications to Statistics*. Wiley-Interscience.

Silverman, Bernard W. 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B*, **47**, 1–21.

Silverman, Bernard W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Smirnov, Nikolai Vasilyevich. 1944. An approximation to the distribution laws of random quantities determined by empirical data. *Uspekhi Matematicheskikh Nauk*, **10**, 179–206.

Smola, Alex J, and Schölkopf, Bernhard. 1998. *Learning with Kernels*. MIT Press.

Snedecor, George W. 1934. *Calculation and Interpretation of Analysis of Variance and Covariance*. Collegiate Press.

Spirtes, P., Glymour, C., and Scheines, R. 2000. *Causation, Prediction, and Search*. 2nd edn. MIT Press.

Steen, Lynn Arthur. 1973. Highlights in the history of spectral theory. *The American Mathematical Monthly*, **80**, 359–381.

Stein, Charles. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Pages 197–206 of: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955*. University of California Press.

Stein, Charles. 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Pages 583–603 of: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. University of California Press.

Stein, Elias M, and Shakarchi, Rami. 2009. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press.

Stewart, G W, and Sun, Ji-guang. 1991. *Matrix Perturbation Theory*. Academic Press.

Stigler, Stephen M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press.

Stigler, Stephen M. 1999. *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press.

Published online by Cambridge University Press

Stokell, Benjamin G, and Shah, Rajen D. 2022. High-dimensional regression with potential prior information on variable importance. *Statistics and Computing*, **32**, 1–12.

Stokell, Benjamin G, Shah, Rajen D, and Tibshirani, Ryan J. 2021. Modelling high-dimensional categorical data using nonconvex fusion penalties. *Journal of the Royal Statistical Society, Series B*, **83**, 579–611.

Stone, Charles J. 1977. Consistent nonparametric regression. *Annals of Statistics*, **5**, 595–620.

Stone, Charles J. 1980. Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, **8**, 1348–1360.

Stone, Charles J. 1984. An asymptotically optimal window selection rule for kernel density estimates. *Annals of Statistics*, **12**, 1285–1297.

Stone, James V. 2004. *Independent Component Analysis: A Tutorial Introduction*. MIT Press.

Stone, Mervyn. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B*, **36**, 111–133.

Strassen, Volker. 1965. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, **36**, 423–439.

Student. 1908. The probable error of a mean. *Biometrika*, **6**, 1–25.

Studeny, Milan. 2006. *Probabilistic Conditional Independence Structures*. Springer Science+Business Media.

Sun, Tingni, and Zhang, Cun-Hui. 2012. Scaled sparse linear regression. *Biometrika*, **99**, 879–898.

Sutherland, Wilson A. 2009. *Introduction to Metric and Topological Spaces*. Oxford University Press.

Tao, Terence. 2011. *An Introduction to Measure Theory*. American Mathematical Society.

Taylor, Brook. 1715. *Methodus Incrementorum Directa et Inversa*. London.

Taylor, Howard L, Banks, Stephen C, and McCoy, John F. 1979. Deconvolution with the l_1 norm. *Geophysics*, **44**, 39–52.

Tibshirani, Robert. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.

Tibshirani, Robert, Saunders, Michael, Rosset, Saharon, Zhu, Ji, and Knight, Keith. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society, Series B*, **67**, 91–108.

Tibshirani, Ryan J. 2013. The lasso problem and uniqueness. *Electronic Journal of Statistics*, **7**, 1456–1490.

Tikhonov, Andrey Nikolayevich. 1943. On the stability of inverse problems. *Proceedings of the USSR Academy of Sciences*, **39**, 195–198.

Tong, Yung Liang. 2012. *The Multivariate Normal Distribution*. Springer Science+Business Media.

Tropp, Joel A. 2006. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, **52**, 1030–1051.

Tropp, Joel A. 2012. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, **12**, 389–434.

Trotter, HF. 1959. An elementary proof of the central limit theorem. *Archiv der Mathematik*, **10**, 226–234.

Tseng, Paul. 2001. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, **109**, 475–494.

Tsiatis, Anastasios A. 2006. *Semiparametric Theory and Missing Data*. Springer.

Tsybakov, Alexandre B. 1998. Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Annals of Statistics*, **26**, 2420–2469.

Tsybakov, Alexandre B. 2009. *Introduction to Nonparametric Estimation*. Springer Science+Business Media.

Published online by Cambridge University Press

Tukey, John. 1953. Multiple comparisons. *Journal of the American Statistical Association*, **48**, 624–625.

Tychonoff, Andrey. 1930. Über die topologische Erweiterung von Räumen. *Mathematische Annalen*, **102**, 544–561.

van de Geer, Sara. 2000. *Empirical Processes in M-Estimation*. Cambridge University Press.

van de Geer, Sara. 2007. The deterministic Lasso. In: *Proceedings of Joint Statistical Meetings*. American Statistical Association.

van de Geer, Sara A. 2008. High-dimensional generalized linear models and the lasso. *Annals of Statistics*, **36**, 614–645.

van de Geer, Sara A, and Bühlmann, Peter. 2009. On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, **3**, 1360–1392.

van der Kooij, Anita J. 2007. *Prediction Accuracy and Stability of Regression with Optimal Scaling Transformations*. Ph.D. thesis, Leiden University.

van der Vaart, Aad W. 1998. *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, Aad W, and Wellner, Jon A. 1996. *Weak Convergence*. Springer.

van Eeden, C. 1956. Maximum likelihood estimation of ordered probabilities. *Indagationes Mathematicae*, **18**, 444–455.

van Genuchten, Martinus Th., and Gupta, SK. 1993. A reassessment of the crop tolerance response function. *Journal of the Indian Society of Soil Science*, **41**, 730–730.

Varshamov, Rom Rubenovich. 1957. Estimate of the number of signals in error correcting codes. *Doklady Akademii Nauk SSSR*, **117**, 739–741.

Verma, Thomas, and Pearl, Judea. 1990a. Causal networks: Semantics and expressiveness. *Machine Intelligence and Pattern Recognition*, **9**, 69–76.

Verma, Thomas, and Pearl, Judea. 1990b. Equivalence and synthesis of causal models. Page 255–270 of: *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*. Elsevier.

Vershynin, Roman. 2018. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press.

von Neumann, John. 1927. Mathematische begründung der quantenmechanik. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, **1927**, 1–57.

von Neumann, John. 1929. Allgemeine Eigenwerttheorie Hermitescher Funktionaloperation. *Mathematische Annalen*, **102**, 49–131.

Wahba, Grace. 1990. *Spline Models for Observational Data*. SIAM.

Wainwright, Martin J. 2009a. Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Transactions on Information Theory*, **55**, 5728–5741.

Wainwright, Martin J. 2009b. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory*, **55**, 2183–2202.

Wainwright, Martin J. 2014. Constrained forms of statistical minimax: Computation, communication and privacy. In: *Proceedings of the International Congress of Mathematicians*.

Wainwright, Martin J. 2019. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press.

Wald, Abraham. 1939. Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics*, **10**, 299–326.

Wald, Abraham. 1949. Statistical decision functions. *Annals of Mathematical Statistics*, **20**, 165–205.

Wald, Abraham. 1950. *Statistical Decision Functions*. John Wiley & Sons.

Walther, Guenther. 2002. Detecting the presence of mixing with multiscale maximum likelihood. *Journal of the American Statistical Association*, **97**, 508–513.

Published online by Cambridge University Press

Walther, Guenther. 2009. Inference and modeling with log-concave distributions. *Statistical Science*, **24**, 319–327.

Wand, MP, and Jones, MC. 1995. *Kernel Smoothing*. CRC Press.

Wang, Hongjian, and Ramdas, Aaditya. 2024. Anytime-valid t-tests and confidence sequences for Gaussian means with unknown variance. *Sequential Analysis*, **44**, 56–110.

Wang, Tengyao. 2016. *Spectral Methods and Computational Trade-offs in High-dimensional Statistical Inference*. Ph.D. thesis, University of Cambridge.

Wang, Tengyao, Berthet, Quentin, and Samworth, Richard J. 2016. Statistical and computational trade-offs in estimation of sparse principal components. *Annals of Statistics*, **44**, 1896–1930.

Wasserman, Larry. 2004. *All of Statistics: A Concise Course in Statistical Inference*. Springer Science+Business Media.

Wasserman, Larry. 2006. *All of Nonparametric Statistics*. Springer Science+Business Media.

Watson, Geoffrey S. 1964. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, **26**, 359–372.

Wedin, Per-Åke. 1972. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, **12**, 99–111.

Wendland, Holger. 2005. *Scattered Data Approximation*. Cambridge University Press.

Wermuth, Nanny, and Lauritzen, Steffen L. 1983. Graphical and Recursive Models for Contingency Tables. *Biometrika*, **70**, 537–552.

Weyl, Hermann. 1912. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen. *Mathematische Annalen*, **71**, 441–479.

Widom, Harold. 1964. Asymptotic behavior of the eigenvalues of certain integral equations. II. *Archive for Rational Mechanics and Analysis*, **17**, 215–229.

Wilks, Samuel S. 1962. *Mathematical Statistics*. John Wiley & Sons.

Williams, David. 1991. *Probability with Martingales*. Cambridge University Press.

Wolpert, David H. 1992. Stacked generalization. *Neural Networks*, **5**, 241–259.

Woodroofe, Michael, and Sun, Jiayang. 1993. A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statistica Sinica*, **3**, 501–515.

Wright, JD Maitland. 1973. All operators on a Hilbert space are bounded. *Bulletin of the American Mathematical Society*, **79**, 1247–1250.

Wright, Sewall. 1921. Correlation and causation. *Journal of Agricultural Research*, **20**, 557.

Wright, Sewall. 1934. The method of path coefficients. *Annals of Mathematical Statistics*, **5**, 161–215.

Wu, Tong Tong, and Lange, Kenneth. 2008. Coordinate descent algorithms for lasso penalized regression. *Annals of Applied Statistics*, **2**, 224–244.

Xu, Min, and Samworth, Richard J. 2021. High-dimensional nonparametric density estimation via symmetry and shape constraints. *Annals of Statistics*, **49**, 650–672.

Yang, Fan, and Barber, Rina Foygel. 2019. Contraction and uniform convergence of isotonic regression. *Electronic Journal of Statistics*, **13**, 646–677.

Yang, Yuhong, and Barron, Andrew. 1999. Information-theoretic determination of minimax rates of convergence. *Annals of Statistics*, **27**, 1564–1599.

Ye, Fei, and Zhang, Cun-Hui. 2010. Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *Journal of Machine Learning Research*, **11**, 3519–3540.

Yeh, James. 2014. *Real Analysis: Theory of Measure and Integration*. World Scientific Publishing Company.

Young, Elliot H, Shah, Rajen D, and Samworth, Richard J. 2026+. Outrigger local polynomial regression. *arXiv:2603.11282*.

Young, Nicholas. 1988. *An Introduction to Hilbert Space*. Cambridge University Press.

Yu, Bin. 1997. Assouad, Fano, and Le Cam. Pages 423–435 of: *Festschrift for Lucien Le Cam*. Springer.

Published online by Cambridge University Press

Yu, Yi, Wang, Tengyao, and Samworth, Richard J. 2015. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, **102**, 315–323.

Yuan, Ming, and Lin, Yi. 2007a. Model selection and estimation in the Gaussian graphical model. *Biometrika*, **94**, 19–35.

Yuan, Ming, and Lin, Yi. 2007b. On the non-negative garrotte estimator. *Journal of the Royal Statistical Society, Series B*, **69**, 143–161.

Zhang, Cun-Hui. 2002. Risk bounds in isotonic regression. *Annals of Statistics*, **30**, 528–555.

Zhang, Cun-Hui. 2010. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, **38**, 894–942.

Zhang, Yuchen, Duchi, John, Jordan, Michael I, and Wainwright, Martin J. 2013. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. *Advances in Neural Information Processing Systems*, **26**.

Zhao, Peng, and Yu, Bin. 2006. On model selection consistency of Lasso. *Journal of Machine Learning Research*, **7**, 2541–2563.

Zou, Hui. 2006. The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418–1429.

Zou, Hui, and Hastie, Trevor. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, **67**, 301–320.

Published online by Cambridge University Press