References

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

AUDRINO, F. and BÜHLMANN, P. (2003). Volatility estimation with functional gradient descent for very high-dimensional financial time series. *Journal of Computational Finance* **6** 65–89.

BACH, F. (2008). Consistency of the group Lasso and multiple kernel learning. *Journal of Machine Learning Research* **9** 1179–1225.

BALL, K. and PAJOR, A. (1990). The entropy of convex bodies with few extreme points. *London Mathematical Society Lecture Note Series* **158** 25–32.

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** 485–516.

BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields* **113** 301–413.

BARTLETT, P., MENDELSON, S. and NEEMAN, J. (2009). ℓ_1-regularized regression: persistence and oracle inequalities. Manuscript.

BEER, M. and TAVAZOIE, S. (2004). Predicting gene expression from sequence. *Cell* **117** 185–198.

BENJAMINI, Y. and HOCHBERG, Y. (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, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29** 1165–1188.

BENNET, G. (1962). Probability inequalities for sums of independent random variables. *Journal of the American Statistical Association* **57** 33–45.

BERTSEKAS, D. (1995). *Nonlinear Programming.* Athena Scientific, Belmont, MA.

BICKEL, P., RITOV, Y. and TSYBAKOV, A. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics* **37** 1705–1732.

BISSANTZ, N., HOHAGE, T., MUNK, A. and RUYMGAART, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM Journal of Numerical Analysis* **45** 2610–2636.

BLANCHARD, G. and ROQUAIN, E. (2008). Two simple sufficient conditions for FDR control. *Electronic Journal of Statistics* **2** 963–992.

BOUSQUET, O. (2002). A Bennet concentration inequality and its application to suprema of empirical processes. *Comptes Rendus de l'Académie des Sciences, Paris* **334** 495–550.

BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge University Press, New York.

BREIMAN, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics* **37** 373–384.

BREIMAN, L. (1996). Bagging predictors. *Machine Learning* **24** 123–140.

BREIMAN, L. (1998). Arcing classifiers (with discussion). *Annals of Statistics* **26** 801–849.

BREIMAN, L. (1999). Prediction games and arcing algorithms. *Neural Computation* **11** 1493–1517.

BREIMAN, L. (2001). Random forests. *Machine Learning* **45** 5–32.

BREIMAN, L., FRIEDMAN, J., OLSHEN, R. and STONE, C. (1984). *Classification and Regression Trees*. Wadsworth.

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

BÜHLMANN, P. (2004). Bagging, boosting and ensemble methods. In *Handbook of Computational Statistics: Concepts and Methods* (J. Gentle, W. Härdle and Y. Mori, eds.). Springer, 877–907.

BÜHLMANN, P. (2006). Boosting for high-dimensional linear models. *Annals of Statistics* **34** 559–583.

BÜHLMANN, P. and HOTHORN, T. (2007). Boosting algorithms: regularization, prediction and model fitting (with discussion). *Statistical Science* **22** 477–505.

BÜHLMANN, P., KALISCH, M. and MAATHUIS, M. (2010). Variable selection in high-dimensional linear models: partially faithful distributions and the PC-simple algorithm. *Biometrika* **97** 261–278.

BÜHLMANN, P. and MEIER, L. (2008). Discussion of "One-step sparse estimates in nonconcave penalized likelihood models" (auths H. Zou and R. Li). *Annals of Statistics* **36** 1534–1541.

BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Annals of Statistics* **30** 927–961.

BÜHLMANN, P. and YU, B. (2003). Boosting with the L_2 loss: regression and classification. *Journal of the American Statistical Association* **98** 324–339.

BÜHLMANN, P. and YU, B. (2006). Sparse boosting. *Journal of Machine Learning Research* **7** 1001–1024.

BUNEA, F. (2008). Honest variable selection in linear and logistic regression models via ℓ_1 and $\ell_1 + \ell_2$ penalization. *Electronic Journal of Statistics* **2** 1153–1194.

BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2006). Aggregation and sparsity via ℓ_1-penalized least squares. In *Proceedings of 19th Annual Conference on Learning Theory, COLT 2006. Lecture Notes in Artificial Intelligence*. Springer Verlag.

BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2007a). Aggregation for Gaussian regression. *Annals of Statistics* **35** 1674.

BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2007b). Sparse density estimation with ℓ_1 penalties. In *Proceedings of 20th Annual Conference on Learning Theory, COLT 2007. Lecture Notes in Artificial Intelligence*. Springer.

BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2007c). Sparsity oracle inequalities for the Lasso. *Electronic Journal of Statistics* **1** 169–194.

BURGE, C. (1998). Modeling dependencies in pre-mRNA splicing signals. In *Computational Methods in Molecular Biology* (S. Salzberg, D. Searls and S. Kasif, eds.), chap. 8. Elsevier Science, 129–164.

BURGE, C. and KARLIN, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268** 78–94.

CANDÈS, E., ROMBERG, J. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* **59** 1207–1223.

CANDÈS, E. and TAO, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory* **51** 4203–4215.

CANDÈS, E. and TAO, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *Annals of Statistics* **35** 2313–2351.

CARUANA, R. (1997). Multitask learning. *Machine Learning* **28** 41–75.

CHESNEAU, C. and HEBIRI, M. (2008). Some theoretical results on the grouped variable Lasso. *Mathematical Methods of Statistics* **17** 317–326.

CONLON, E., LIU, X., LIEB, J. and LIU, J. (2003). Integrating regulatory motif discovery and genome-wide expression analysis. *Proceedings of the National Academy of Sciences* **100** 3339–3344.

DAHINDEN, C., KALISCH, M. and BÜHLMANN, P. (2010). Decomposition and model selection for large contingency tables. *Biometrical Journal* **52** 233–252.

DAHINDEN, C., PARMIGIANI, G., EMERICK, M. and BÜHLMANN, P. (2007). Penalized likelihood for sparse contingency tables with an application to full-length cDNA libraries. *BMC Bioinformatics* **8** 1–11.

DEMPSTER, A., LAIRD, N. and RUBIN, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* **39** 1–38.

DEVORE, R. and TEMLYAKOV, V. (1996). Some remarks on greedy algorithms. *Advances in Computational Mathematics* **5** 173–187.

DIMARZIO, M. and TAYLOR, C. (2008). On boosting kernel regression. *Journal of Statistical Planning and Inference* **138** 2483–2498.

DONOHO, D. (2006). For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* **59** 797–829.

DONOHO, D. and ELAD, M. (2003). Uncertainty principles and ideal atomic decomposition. *Proceedings of the National Academy of Sciences* **100** 2197–2202.

DONOHO, D. and HUO, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory* **47** 2845–2862.

DONOHO, D. and JOHNSTONE, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.

DUDLEY, R. (1987). Universal Donsker classes and metric entropy. *Annals of Probability* **15** 1306–1326.

DUDOIT, S., SHAFFER, J. and BOLDRICK, J. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science* **18** 71–103.

DÜMBGEN, L., VAN DE GEER, S., VERAAR, M. and WELLNER, J. (2010). Nemirovski's inequalities revisited. *The American Mathematical Monthly* **117** 138–160.

EDWARDS, D. (2000). *Introduction to Graphical Modelling*. 2nd ed. Springer Verlag.

EFRON, B. (2004). The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association* **99** 619–632.

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression (with discussion). *Annals of Statistics* **32** 407–451.

FAN, J., FENG, Y. and WU, Y. (2009a). Network exploration via the adaptive Lasso and SCAD penalties. *Annals of Applied Statistics* **3** 521–541.

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96** 1348–1360.

FAN, J. and LV, J. (2008). Sure independence screening for ultra-high dimensional feature space (with discussion). *Journal of the Royal Statistical Society Series B* **70** 849–911.

FAN, J., SAMWORTH, R. and WU, Y. (2009b). Ultrahigh dimensional variable selection: beyond the linear model. *Journal of Machine Learning Research* **10** 1989–2014.

FISHER, R. (1924). The distribution of the partial correlation coefficient. *Metron* **3** 329–332.

FREEDMAN, D. (1977). A remark on the difference between sampling with and without replacement. *Journal of the American Statistical Association* **72** 681.

FREUND, Y. and SCHAPIRE, R. (1996). Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA.

FREUND, Y. and SCHAPIRE, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* **55** 119–139.

FRIEDMAN, J. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29** 1189–1232.

FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007a). Pathwise coordinate optimization. *Annals of Applied Statistics* **1** 302–332.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion). *Annals of Statistics* **28** 337–407.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2007b). Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics* **9** 432–441.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularized paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33** 1–22.

FU, W. (1998). Penalized regressions: the Bridge versus the Lasso. *Journal of Computational and Graphical Statistics* **7** 397–416.

FUCHS, J. (2004). On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory* **50** 1341–1344.

GATU, C., YANEV, P. and KONTOGHIORGHES, E. (2007). A graph approach to generate all possible regression submodels. *Computational Statistics & Data Analysis* **52** 799–815.

GREEN, P. and SILVERMAN, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall, New York.

GREENSHTEIN, E. (2006). Best subset selection, persistence in high dimensional statistical learning and optimization under ℓ_1 constraint. *Annals of Statistics* **34** 2367–2386.

GREENSHTEIN, E. and RITOV, Y. (2004). Persistence in high-dimensional predictor selection and the virtue of over-parametrization. *Bernoulli* **10** 971–988.

GUÉDON, O., MENDELSON, S., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2007). Subspaces and orthogonal decompositions generated by bounded orthogonal systems. *Positivity* **11** 269–283.

HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models*. Chapman & Hall, London.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. Springer, New York.

HEBIRI, M. and VAN DE GEER, S. (2010). The smooth Lasso and other $\ell_1 + \ell_2$-penalized methods. ArXiv:1003.4885.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded variables. *Journal of the American Statistical Association* **58** 13–30.

HOFMANN, M., GATU, C. and KONTOGHIORGHES, E. (2007). Efficient algorithms for computing the best subset regression models for large-scale problems. *Computational Statistics & Data Analysis* **52** 16–29.

HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society Series B* **15** 193–232.

HUANG, J., MA, S. and ZHANG, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica* **18** 1603–1618.

HUANG, J. and ZHANG, T. (2010). The benefit of group sparsity. *Annals of Statistics* **38** 1978–2004.

KALISCH, M. and BÜHLMANN, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research* **8** 613–636.

KING, G. and ZENG, L. (2001). Logistic regression in rare events data. *Political Analysis* **9** 137–163.

KOLTCHINSKII, V. (2009a). Sparsity in penalized empirical risk minimization. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **45** 7–57.

KOLTCHINSKII, V. (2009b). The Dantzig selector and sparsity oracle inequalities. *Bernoulli* **15** 799–828.

KOLTCHINSKII, V., TSYBAKOV, A. and LOUNICI, K. (2010b). Nuclear norm penalization and optimal rates for noisy low rank matrix completion. ArXiv:1011.6256.

KOLTCHINSKII, V. and YUAN, M. (2008). Sparse recovery in large ensembles of kernel machines. In *Proceedings of the 21st Annual Conference on Learning Theory, COLT 2008. Lecture Notes in Artificial Intelligence*. Springer.

KOLTCHINSKII, V. and YUAN, M. (2010). Sparsity in multiple kernel learning. *Annals of Statistics* **38** 3660–3695.

LAURITZEN, S. (1996). *Graphical Models*. Oxford University Press.

LEDOUX, M. (1996). Talagrand deviation inequalities for product measures. *ESAIM: Probability and Statistics* **1** 63–87.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Verlag, New York.

LEEB, H. and PÖTSCHER, B. (2003). The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory* **19** 100–142.

LEEB, H. and PÖTSCHER, B. (2005). Model selection and inference: facts and fiction. *Econometric Theory* **21** 21–59.

LIU, X., BRUTLAG, D. and LIU, J. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nature Biotechnology* **20** 835–839.

LOUBES, J.-M. and VAN DE GEER, S. (2002). Adaptive estimation in regression, using soft thresholding type penalties. *Statistica Neerlandica* **56** 453–478.

LOUNICI, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electronic Journal of Statistics* **2** 90–102.

LOUNICI, K., PONTIL, M., TSYBAKOV, A. and VAN DE GEER, S. (2009). Taking advantage of sparsity in multi-task learning. In *Proceedings of 22th Annual Conference on Learning Theory, COLT 2009. Lecture Notes in Artificial Intelligence*. Springer.

LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. (2010). Oracle inequalities and optimal inference under group sparsity. ArXiv:1007.1771.

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

MASSART, P. (2000a). About the constants in Talagrand's concentration inequalities for empirical processes. *Annals of Probability* **28** 863–884.

MASSART, P. (2000b). Some applications of concentration inequalities to statistics. *Annales de la Faculté des Sciences de Toulouse* **9** 245–303.

MCCULLAGH, P. and NELDER, J. (1989). *Generalized Linear Models*. 2nd ed. Chapman & Hall, London.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. Wiley, New York.

MEIER, L. and BÜHLMANN, P. (2007). Smoothing ℓ_1-penalized estimators for high-dimensional time-course data. *Electronic Journal of Statistics* **1** 597–615.

MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2008). The group Lasso for logistic regression. *Journal of the Royal Statistical Society Series B* **70** 53–71.

MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Annals of Statistics* **37** 3779–3821.

MEINSHAUSEN, N. (2007). Relaxed Lasso. *Computational Statistics & Data Analysis* **52** 374–393.

MEINSHAUSEN, N. (2008). A note on the Lasso for graphical Gaussian model selection. *Statistics and Probability Letters* **78** 880–884.

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics* **34** 1436–1462.

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection (with discussion). *Journal of the Royal Statistical Society Series B* **72** 417–473.

MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). P-values for high-dimensional regression. *Journal of the American Statistical Association* **104** 1671–1681.

MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Annals of Statistics* **37** 246–270.

NARDI, Y. and RINALDO, A. (2008). On the asymptotic properties of the group Lasso estimator for linear models. *Electronic Journal of Statistics* **2** 605–633.

OSBORNE, M., PRESNELL, B. and TURLACH, B. (2000). A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis* **20** 389–403.

PINHEIRO, J. and BATES, D. (2000). *Mixed-Effects Models in S and S-Plus*. Springer, New York.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. IMS Lecture Notes.

RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009a). Sparse additive models. *Journal of the Royal Statistical Society Series B* **71** 1009–1030.

RAVIKUMAR, P., WAINWRIGHT, M. and LAFFERTY, J. (2009b). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Annals of Statistics* **38** 1287–1319.

RAVIKUMAR, P., WAINWRIGHT, M., RASKUTTI, G. and YU, B. (2008). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. ArXiv:0811.3628.

RIDGEWAY, G. (1999). The state of boosting. *Computing Science and Statistics* **31** 172–181.

ROSSET, S. and ZHU, J. (2007). Piecewise linear regularized solution paths. *Annals of Statistics* **35** 1012–1030.

ROTHMAN, A., BICKEL, P., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* **2** 494–515.

SCHAPIRE, R. (2002). The boosting approach to machine learning: an overview. In *MSRI Workshop on Nonlinear Estimation and Classification* (D. Denison, M. Hansen, C. Holmes, B. Mallick and B. Yu, eds.). Springer.

SCHELLDORFER, J., BÜHLMANN, P. and VAN DE GEER, S. (2011). Estimation for high-dimensional linear mixed-effects models using ℓ_1-penalization. *Scandinavian Journal of Statistics* (to appear).

SCHMID, M. and HOTHORN, T. (2008). Boosting additive models using component-wise p-splines as base-learners. *Computational Statistics & Data Analysis* **53** 298–311.

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

SHAWE-TAYLOR, J. and SUN, S. (2010). Discussion of "Stability selection" (auths N. Meinshausen and P. Bühlmann). *Journal of the Royal Statistical Society Series B* **72** 451–453.

SOUTHWELL, R. (1946). *Relaxation Methods in Theoretical Physics*. Oxford University Press.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*. 2nd ed. MIT Press.

STÄDLER, N., BÜHLMANN, P. and VAN DE GEER, S. (2010). ℓ_1-penalization for mixture regression models (with discussion). *Test* **19** 209–285.

SUN, T. and ZHANG, C.-H. (2010). Discussion of "ℓ_1-penalization for mixture regression models" (auths N. Städler, P. Bühlmann and S. van de Geer). *Test* **19** 270–275.

TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'IHES* **81** 73–205.

TALAGRAND, M. (2005). *The generic chaining: upper and lower bounds of stochastic processes*. Springer Verlag.

TARIGAN, B. and VAN DE GEER, S. (2006). Classifiers of support vector machine type with ℓ_1 complexity regularization. *Bernoulli* **12** 1045–1076.

TEMLYAKOV, V. (2000). Weak greedy algorithms. *Advances in Computational Mathematics* **12** 213–227.

TEMLYAKOV, V. (2008). Nonlinear methods of approximation. *Foundations of Computational Mathematics* **3** 33–107.

TIBSHIRANI, R. (1996). Regression analysis and selection via the Lasso. *Journal of the Royal Statistical Society Series B* **58** 267–288.

TROPP, J. (2004). Greed is good: algorithmic results for sparse approximation. *IEEE Transactions on Information Theory* **50** 2231–2242.

TROPP, J. and GILBERT, A. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory* **53** 4655–4666.

TSENG, P. (2001). Convergence of a block coordinate descent method for nonsmooth separable minimization. *Journal of Optimization Theory and Applications* **109** 475–494.

TSENG, P. and YUN, S. (2009). A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming, Series B* **117** 387–423.

TSYBAKOV, A. (2004). Optimal aggregation of classifiers in statistical learning. *Annals of Statistics* **32** 135–166.

TUKEY, J. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

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

VAN DE GEER, S. (2001). Least squares estimation with complexity penalties. *Mathematical Methods of Statistics* **10** 355–374.

VAN DE GEER, S. (2002). M-estimation using penalties or sieves. *Journal of Statistical Planning and Inference* **32** 55–69.

VAN DE GEER, S. (2007). The deterministic Lasso. In *JSM proceedings, 2007, 140*. American Statistical Association.

VAN DE GEER, S. (2008). High-dimensional generalized linear models and the Lasso. *Annals of Statistics* **36** 614–645.

VAN DE GEER, S. (2010). The Lasso with within group structure. In *IMS Collections: Festschrift for Jana Jurečková*. IMS, 235–244.

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 DE GEER, S., BÜHLMANN, P. and ZHOU, S. (2010). The adaptive and thresholded Lasso for potentially misspecified models. ArXiv:1001.5176.

VAN DE WIEL, M., BERKHOF, J. and VAN WIERINGEN, W. (2009). Testing the prediction error difference between two predictors. *Biostatistics* **10** 550–560.

VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics, Springer-Verlag, New York.

VAPNIK, V. (2000). *The Nature of Statistical Learning Theory*. 2nd ed. Statistics for Engineering and Information Science, Springer-Verlag, New York.

VIENS, F. and VIZCARRA, A. (2007). Supremum concentration inequality and modulus of continuity for sub-nth chaos processes. *Journal of Functional Analysis* **248** 1–26.

WAHBA, G. (1990). *Spline models for observational data*. CBMS-NSF Regional Conference Series in Applied Mathematics, 59, SIAM.

WAINWRIGHT, M. (2007). Information-theoretic limitations on sparsity recovery in the high-dimensional and noisy setting. *IEEE Transactions on Information Theory* **55** 5728–5741.

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

WALLACE, D. (1959). Bounds for normal approximations of Student's t and the chi-square distributions. *Annals of Mathematical Statistics* **30** 1121–1130.

WASSERMAN, L. and ROEDER, K. (2009). High dimensional variable selection. *Annals of Statistics* **37** 2178–2201.

WEST, M., BLANCHETTE, C., DRESSMAN, H., HUANG, E., ISHIDA, S., SPANG, R., ZUZAN, H., OLSON, J., MARKS, J. and NEVINS, J. (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proceedings of the National Academy of Sciences* **98** 11462–11467.

WILLE, A. and BÜHLMANN, P. (2006). Low-order conditional independence graphs for inferring genetic networks. *Statistical Applications in Genetics and Molecular Biology* **5** 1–32.

WILLE, A., ZIMMERMANN, P., VRANOVÁ, E., FÜRHOLZ, A., LAULE, O., BLEULER, S., HENNIG, L., PRELIĆ, A., VON ROHR, P., THIELE, L., ZITZLER, E., GRUISSEM, W. and BÜHLMANN, P. (2004). Sparse graphical Gaussian modeling for genetic regulatory network inference. *Genome Biology* **5** R92, 1–13.

WOOD, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall.

WU, C. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics* **11**.

YAO, Y., ROSASCO, L. and CAPONNETTO, A. (2007). On early stopping in gradient descent learning. *Constructive Approximation* **26** 289–315.

YEO, G. and BURGE, C. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology* **11** 475–494.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B* **68** 49.

ZHANG, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics* **38** 894–942.

ZHANG, C. and HUANG, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *Annals of Statistics* **36** 1567–1594.

ZHANG, T. (2009a). On the consistency of feature selection using greedy least squares regression. *Journal of Machine Learning Research* **10** 555–568.

ZHANG, T. (2009b). Some sharp performance bounds for least squares regression with L1 regularization. *Annals of Statistics* **37** 2109–2144.

ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research* **7** 2541–2563.

ZHOU, S. (2009a). Restricted eigenvalue conditions on subgaussian random matrices. ArXiv:0912.4045.

ZHOU, S. (2009b). Thresholding procedures for high dimensional variable selection and statistical estimation. In *Advances in Neural Information Processing Systems 22*. MIT Press.

ZHOU, S. (2010). Thresholded Lasso for high dimensional variable selection and statistical estimation. ArXiv:1002.1583.

ZHOU, S., RÜTIMANN, P., XU, M. and BÜHLMANN, P. (2010). High-dimensional covariance estimation based on Gaussian graphical models. ArXiv:1009.0530.

ZOU, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* **101** 1418–1429.

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society Series B* **67** 301–320.

ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the "degrees of freedom" of the Lasso. *Annals of Statistics* **35** 2173–2192.

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models (with discussion). *Annals of Statistics* **36** 1509–1566.