

# References

- F. Bach, Structured sparsity-inducing norms through submodular functions, in *Advances in Neural Information Processing Systems (NIPS)*, vol. 23 (2010)
- F. Bach, R. Jenatton, J. Mairal, G. Obozinski, Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.* **4**, 1–106 (2012)
- A. Belloni, V. Chernozhukov, L. Wang, Square-root Lasso: pivotal recovery of sparse signals via conic programming. *Biometrika* **98**, 791–806 (2011)
- Q. Berthet, P. Rigollet, Complexity theoretic lower bounds for sparse principal component detection, in *Conference on Learning Theory* (2013)
- P. Bickel, C. Klaassen, Y. Ritov, J. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models* (Johns Hopkins University, Press Baltimore, 1993)
- P. Bickel, Y. Ritov, A. Tsybakov, Simultaneous analysis of Lasso and Dantzig selector. *Ann. Stat.* **37**, 1705–1732 (2009)
- L. Birgé, P. Massart, Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329–375 (1998)
- M. Bogdan, E. van den Berg, W. Su, E. Candes, Statistical estimation and testing via de sorted l1 norm. *ArXiv:1310.1969* (2013)
- C. Borell, The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30**, 207–216 (1975)
- S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Oxford University Press, Oxford, 2013)
- O. Bousquet, A Bennet concentration inequality and its application to suprema of empirical processes. *C. R. Acad. Sci. Paris* **334**, 495–550 (2002)
- P. Bühlmann, S. van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications* (Springer, Heidelberg, 2011)
- F. Bunea, J. Lederer, Y. She, The group square-root lasso: theoretical properties and fast algorithms. *IEEE Trans. Inf. Theory* **2**, 1313–1325 (2014)
- B. Carl, Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in banach spaces. *Ann. Inst. Fourier* **35**, 79–118 (1985)
- G. Chen, M. Teboulle, Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.* **3**, 538–543 (1993)
- S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20**, 33–61 (1998)
- R. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1**, 290–330 (1967)
- L. Dümbgen, S. van de Geer, M. Veraar, J. Wellner, Nemirovski’s inequalities revisited. *Am. Math. Mon.* **117**, 138–160 (2010)

- J. Friedman, T. Hastie, R. Tibshirani, Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics* **9**, 432–441 (2008)
- O. Güler, On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.* **29**, 403–419 (1991)
- T. Hastie, R. Tibshirani, M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations* (CRC Press, Boca Raton, 2015)
- W. Hoeffding, Probability inequalities for sums of bounded variables. *J. Am. Stat. Assoc.* **58**, 13–30 (1963)
- J. Janková, S. van de Geer, Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Stat.* **9**, 1205–1229 (2015a)
- J. Janková, S. van de Geer, Honest confidence regions and optimality in high-dimensional precision matrix estimation. *ArXiv:1507.02061* (2015b)
- A. Javanmard, A. Montanari, Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15**, 2869–2909 (2014)
- R. Jenatton, J.-Y. Audibert, F. Bach, Structured variable selection with sparsity inducing norms. *J. Mach. Learn. Res.* **12**, 2777–2824 (2011)
- V. Koltchinskii, *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: École d'Été de Probabilités de Saint-Flour XXXVIII-2008*, vol. 38 (Springer, Heidelberg, 2011)
- V. Koltchinskii, K. Lounici, A. Tsybakov, Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Stat.* **39**, 2302–2329 (2011)
- J. Lafond, Low rank matrix completion with exponential family noise, in *Conference on Learning Theory* (2015)
- B. Laurent, P. Massart, Adaptive estimation of a quadratic functional by model selection. *Ann. Stat.* **28**, 1302–1338 (2000)
- G. Lecué, S. Mendelson, Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc.* To appear, available at arXiv preprint. *arXiv:1401.2188* (2014)
- M. Ledoux, Talagrand deviation inequalities for product measures. *ESAIM Probab. Stat.* **1**, 63–87 (1996)
- M. Ledoux, *The Concentration of Measure Phenomenon*, vol. 89 (American Mathematical Society, Providence, 2005)
- M. Ledoux, M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes* (Springer, New York, 1991)
- K. Lounici, Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electron. J. Stat.* **2**, 90–102 (2008)
- K. Lounici, M. Pontil, S. van de Geer, A. Tsybakov, Oracle inequalities and optimal inference under group sparsity. *Ann. Stat.* **39**, 2164–2204 (2011)
- P. Massart, About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28**, 863–884 (2000)
- A. Maurer, M. Pontil, Structured sparsity and generalization. *J. Mach. Learn. Res.* **13**, 671–690 (2012)
- N. Meinshausen, P. Bühlmann, High-dimensional graphs and variable selection with the Lasso. *Ann. Stat.* **34**, 1436–1462 (2006)
- S. Mendelson, Learning without concentration. *J. ACM* **62**, 21 (2015)
- C. Micchelli, J. Morales, M. Pontil, A family of penalty functions for structured sparsity, in *Advances in Neural Information Processing Systems, NIPS 2010*, vol. 23 (2010)
- R.I. Oliveira, Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab.* **15**, 26 (2010)
- D. Pollard, *Convergence of Stochastic Processes* (Springer, New York, 1984)
- P. Ravikumar, M. Wainwright, G. Raskutti, B. Yu, High-dimensional covariance estimation by minimizing  $\ell_1$ -penalized log-determinant divergence. *Electron. J. Stat.* **5**, 935–980 (2011)
- Z. Ren, T. Sun, C.-H. Zhang, H. Zhou, Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Stat.* **43**, 991–1026 (2015)

- A. Rothmann, P. Bickel, E. Levina, J. Zhu, Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2**, 494–515 (2008)
- J. Schauder, Der Fixpunktsatz in funktionalräumen. *Studia* **2**, 171–180 (1930)
- N. Städler, P. Bühlmann, S. van de Geer,  $\ell_1$ -penalization for mixture regression models. *Test* **19**, 209–256 (2010)
- N. Städler, P. Bühlmann, S. van de Geer, Rejoinder  $\ell_1$ -penalization in mixture regression models. *Test* **19**, 280–285 (2010)
- B. Stucky, S. van de Geer, Sharp oracle inequalities for square root regularization. *ArXiv:1509.04093* (2015)
- T. Sun, C.-H. Zhang, Comments on:  $\ell_1$ -penalization in mixture regression models. *Test* **19**, 270–275 (2010)
- T. Sun, C.-H. Zhang, Scaled sparse linear regression. *Biometrika* **99**, 879–898 (2012)
- T. Sun, C.-H. Zhang, Sparse matrix inversion with scaled lasso. *J. Mach. Learn. Res.* **14**, 3385–3418 (2013)
- M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. IHES* **81**, 73–205 (1995)
- M. Talagrand, *The Generic Chaining* (Springer, Heidelberg, 2005)
- R. Tibshirani, Regression analysis and selection via the Lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996)
- J. Tropp, User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12**, 389–434 (2012)
- J.A. Tropp, An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8**, 1–230 (2015)
- S. van de Geer, *Empirical Processes in M-Estimation* (Cambridge University Press, Cambridge, 2000)
- S. van de Geer, Least squares estimation with complexity penalties. *Math. Methods Stat.* **10**, 355–374 (2001)
- S. van de Geer, The deterministic Lasso, in *JSM Proceedings, 2007*, 140 (American Statistical Association, Alexandria, 2007)
- S. van de Geer, Weakly decomposable regularization penalties and structured sparsity. *Scand. J. Stat.* **41**, 72–86 (2014)
- S. van de Geer, P. Bühlmann, On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3**, 1360–1392 (2009)
- S. van de Geer, J. Janková, Semi-parametric efficiency bounds and efficient estimation for high-dimensional models. *ArXiv:1601.00815* (2016)
- S. van de Geer, A. Muro, On higher order isotropy conditions and lower bounds for sparse quadratic forms. *Electron. J. Stat.* **8**, 3031–3061 (2014)
- S. van de Geer, P. Bühlmann, Y. Ritov, R. Dezeure, On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Stat.* **42**, 1166–1202 (2014)
- R. van Handel, Chaining, interpolation and convexity. *ArXiv:1508.05906* (2015)
- A. van der Vaart, *Asymptotic Statistics*, vol. 3 (Cambridge University Press, Cambridge, 2000)
- A.W. van der Vaart, J.A. Wellner, *Weak Convergence and Empirical Processes*. Springer Series in Statistic (Springer, New York, 1996)
- G. Watson, Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl.* **170**, 33–45 (2015)
- M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* **68**, 49 (2006)
- X. Zeng, A. Mario, The ordered weighted  $\ell_1$  norm: Atomic formulation, dual norm, and projections. *ArXiv:1409.4271* (2014)
- C.-H. Zhang, S.S. Zhang, Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **76**, 217–242 (2014)

# Author Index

- Audibert, J.-Y., 75
- Bühlmann, P., 10, 11, 28, 53, 54, 61, 63, 125, 257
- Bach, F., 75, 76
- Belloni, A., 27
- Berthet, Q., 191
- Bickel, P., 24, 56, 211, 216
- Birgé, L., 125
- Bogdan, M., 97
- Borell, C., 121, 131
- Boucheron, S., 235, 239
- Bousquet, O., 227, 233, 235–237
- Bunea, F., 89
- Candès, E., 97
- Carl, B., 260
- Chen, G., 10
- Chen, S.S., 7
- Chernozhukov, V., 27
- Dümbgen, L., 196, 204, 221
- Dezeure, R., 63
- Donoho, D.L., 7
- Dudley, R., 252
- Friedman, J., 216
- Güler, O., 10
- Hastie, T., 2, 216
- Hoeffding, W., 241
- Janková, J., 57, 215, 217, 218
- Javanmard, A., 63
- Jenatton, R., 75
- Klaassen, C., 56
- Koltchinskii, V., 12, 135, 182
- Lafond, J., 187
- Laurent, B., 123, 126, 128
- Lecué, G., 152, 230
- Lederer, J., 89
- Ledoux, M., 233, 235, 239
- Levina, E., 211, 216
- Lin, Y., 89
- Lounici, K., 12, 46, 135
- Lugosi, G., 235, 239
- Mario, A., 97
- Massart, P., 123, 125, 126, 128, 233, 235, 236, 239
- Maurer, A., 75, 124
- Meinshausen, N., 61
- Mendelson, S., 152, 230
- Michelli, C., 75, 86
- Montanari, A., 63
- Morales, J., 75, 86
- Muro, A., 225

Oliviera, R.I., 135

Pollard, D., 234

Pontil, M., 75, 86, 124

Raskutti, G., 199, 211, 216

Ravikumar, P., 199, 211, 216

Ren, Z., 215

Rigollet, P., 191

Ritov, Y., 24, 56, 63

Rothmann, A., 211, 216

Saunders, M.A., 7

Schauder, J., 199

She, Y., 89

Städler, N., 28

Stucky, B., 76

Su, W., 97

Sun, T., 28, 33, 35, 215

Talagrand, M., 235, 240, 251, 256, 259

Teboulle, M., 10

Tibshirani, R., 2, 8, 216

Tropp, J.A., 133–135

Tsybakov, A., 12, 24, 135

van de Geer, S., 10, 11, 28, 53, 54, 57, 63, 76,  
80, 118, 125, 150, 196, 204, 215,  
217, 218, 221, 225, 234, 257

van den Berg, E., 97

van der Vaart, A., 56, 233

van Handel, R., 256

Veraar, M., 196, 204, 221

Wainwright, M., 2, 199, 211, 216

Wang, L., 27

Watson, G., 44, 184

Wellner, J., 56, 196, 204, 221, 233

Yu, B., 199, 211, 216

Yuan, M., 89

Zeng, X., 97

Zhang, C.-H., 28, 33, 35, 63, 215

Zhang, S.S., 63

Zhou, H., 215

Zhu, J., 211, 216

# Index

- $\ell_1$ -operator norm, 44
- active set, 6
- allowed set, 78, 88
- allowed vector, 108
- almost orthogonality, 9, 42, 50, 62
- almost-differentiable, 105
- anti-projection, 42, 54
- approximation error, 2, 173
- asymptotic linearity, 65, 218, 220
- basis pursuit, 7
- Bernstein's inequality, 124, 133
- bias, 44
- Brouwer's fixed point, 199
- candidate oracle, 13, 16, 82, 110, 113, 185
- chaining, 245
- compatibility constant, 11, 80, 230
- concentration, 235, 244, 250
- confidence interval, 65, 67
- contraction, 234
- convex conjugate, 10, 105
- covering number, 245
- de-selection, 54, 204
- de-sparsifying, 63, 67, 71, 93, 217
- decomposable, 10
- density estimation, 140, 155, 172
- dual norm, 76, 89, 183
- dual norm inequality, 10, 44
- effective sparsity, 11, 17, 81, 108, 229
- empirical process, 139, 239
- empirical risk, 1, 103
- entropy, 245
- entropy of a vector, 261
- estimation error, 33
- exponential family, 140, 155, 209
- family tree, 246
- fixed design, 7
- Frobenius norm, 182
- generalized linear model, 146, 160
- generic chaining, 246
- Gram matrix, 8, 61, 143, 223
- graphical Lasso, 210, 216
- group Lasso, 89, 95, 101, 170
- high-dimensional, 1, 7, 103
- Hoeffding's inequality, 133, 241
- Huber loss, 146
- inverse, 48
- irrepresentable condition, 53, 205, 214
- isotropy, 224
- KKT-conditions, 9, 28, 37, 41, 69, 77, 93, 101, 200, 217, 219

- Lasso, 8, 41, 167
- least absolute deviation regression, 146
- least squares estimator, 5
- linear contrast, 67
- linear model, 5, 70
- local neighbourhood, 104, 111, 151, 199
- logistic regression, 141, 174
- loss function, 103
  
- matrix Bernstein, 134
- matrix completion, 136, 187
- matrix Hoeffding, 134
- misspecified, 6, 7, 173
- mixture model, 148, 164
- multiple regression model, 100, 130, 196
  
- node-wise Lasso, 61, 218
- norm generated from cone, 86
- nuclear norm, 37, 44, 182
  
- one point margin, 113
- oracle, 13, 17, 33, 82, 111
- Orlicz norm, 135
- overfitting, 7, 28, 29, 84
  
- Poisson regression, 141
- precision matrix, 42, 210, 215
- prediction error, 6, 16
- projection estimators, 142, 158, 172
  
- quantile regression, 146
  
- Rademacher sequence, 135, 146, 225, 233, 241
- regression, 157
- robust loss, 146
  
- scale parameter, 35
  
- scaled Lasso, 28, 35
- Schauder's fixed point, 199
- semi-parametric, 56
- sequence space model, 23, 39
- Sharp Oracle Inequality, 81, 85
- sharp oracle inequality, 2, 12, 20, 32, 33, 110, 195
- singular value decomposition, 183, 184
- smallish coefficients, 16
- sorted  $\ell_1$ -norm, 97
- sparse, 6, 7, 9, 13
- sparse PCA, 191
- sparsity inducing norm, 75
- spectral decomposition, 192
- spikiness condition, 192
- square-root Lasso, 27, 36, 168
- stretching factor, 11, 80, 108
- strong sparsity, 6, 19
- strongly asymptotically unbiased, 56
- structured sparsity estimator, 76, 81, 101, 169
- structured sparsity M-estimator, 104, 200
- sub-differential, 9, 76
- sub-Gaussian vector, 224
- surrogate anti-projection, 47, 49, 68, 93
- surrogate projection, 9, 47, 49, 68, 93
- symmetrization, 233
  
- target, 1, 6, 104, 114, 119, 160
- Taylor expansion, 153
- theoretical risk, 103
- trace regression, 182
- triangle property, 10, 107
- tuning parameter, 9, 83, 104
- two point inequality, 9, 77, 105
- two point margin, 9, 106, 151
  
- weak sparsity, 19, 189
- weakly decomposable norm, 78, 88, 109
- wedge penalty, 90, 95, 129, 171
- worst possible sub-direction, 45, 48, 49, 56, 57

Editors in Chief: J.-M. Morel, B. Teissier;

### Editorial Policy

1. Lecture Notes aim to report new developments in all areas of mathematics and their applications – quickly, informally and at a high level. Mathematical texts analysing new developments in modelling and numerical simulation are welcome.

Manuscripts should be reasonably self-contained and rounded off. Thus they may, and often will, present not only results of the author but also related work by other people. They may be based on specialised lecture courses. Furthermore, the manuscripts should provide sufficient motivation, examples and applications. This clearly distinguishes Lecture Notes from journal articles or technical reports which normally are very concise. Articles intended for a journal but too long to be accepted by most journals, usually do not have this “lecture notes” character. For similar reasons it is unusual for doctoral theses to be accepted for the Lecture Notes series, though habilitation theses may be appropriate.

2. Besides monographs, multi-author manuscripts resulting from SUMMER SCHOOLS or similar INTENSIVE COURSES are welcome, provided their objective was held to present an active mathematical topic to an audience at the beginning or intermediate graduate level (a list of participants should be provided).

The resulting manuscript should not be just a collection of course notes, but should require advance planning and coordination among the main lecturers. The subject matter should dictate the structure of the book. This structure should be motivated and explained in a scientific introduction, and the notation, references, index and formulation of results should be, if possible, unified by the editors. Each contribution should have an abstract and an introduction referring to the other contributions. In other words, more preparatory work must go into a multi-authored volume than simply assembling a disparate collection of papers, communicated at the event.

3. Manuscripts should be submitted either online at [www.editorialmanager.com/lnm](http://www.editorialmanager.com/lnm) to Springer’s mathematics editorial in Heidelberg, or electronically to one of the series editors. Authors should be aware that incomplete or insufficiently close-to-final manuscripts almost always result in longer refereeing times and nevertheless unclear referees’ recommendations, making further refereeing of a final draft necessary. The strict minimum amount of material that will be considered should include a detailed outline describing the planned contents of each chapter, a bibliography and several sample chapters. Parallel submission of a manuscript to another publisher while under consideration for LNM is not acceptable and can lead to rejection.
4. In general, **monographs** will be sent out to at least 2 external referees for evaluation.

A final decision to publish can be made only on the basis of the complete manuscript, however a refereeing process leading to a preliminary decision can be based on a pre-final or incomplete manuscript.

Volume Editors of **multi-author works** are expected to arrange for the refereeing, to the usual scientific standards, of the individual contributions. If the resulting reports can be



forwarded to the LNM Editorial Board, this is very helpful. If no reports are forwarded or if other questions remain unclear in respect of homogeneity etc, the series editors may wish to consult external referees for an overall evaluation of the volume.

5. Manuscripts should in general be submitted in English. Final manuscripts should contain at least 100 pages of mathematical text and should always include
  - a table of contents;
  - an informative introduction, with adequate motivation and perhaps some historical remarks: it should be accessible to a reader not intimately familiar with the topic treated;
  - a subject index: as a rule this is genuinely helpful for the reader.
  - For evaluation purposes, manuscripts should be submitted as pdf files.
6. Careful preparation of the manuscripts will help keep production time short besides ensuring satisfactory appearance of the finished book in print and online. After acceptance of the manuscript authors will be asked to prepare the final LaTeX source files (see LaTeX templates online: <https://www.springer.com/gb/authors-editors/book-authors-editors/manuscriptpreparation/5636>) plus the corresponding pdf- or zipped ps-file. The LaTeX source files are essential for producing the full-text online version of the book, see <http://link.springer.com/bookseries/304> for the existing online volumes of LNM). The technical production of a Lecture Notes volume takes approximately 12 weeks. Additional instructions, if necessary, are available on request from [lnm@springer.com](mailto:lnm@springer.com).
7. Authors receive a total of 30 free copies of their volume and free access to their book on SpringerLink, but no royalties. They are entitled to a discount of 33.3 % on the price of Springer books purchased for their personal use, if ordering directly from Springer.
8. Commitment to publish is made by a *Publishing Agreement*; contributing authors of multiauthor books are requested to sign a *Consent to Publish form*. Springer-Verlag registers the copyright for each volume. Authors are free to reuse material contained in their LNM volumes in later publications: a brief written (or e-mail) request for formal permission is sufficient.

**Addresses:**

Professor Jean-Michel Morel, CMLA, École Normale Supérieure de Cachan, France  
E-mail: [moreljeanmichel@gmail.com](mailto:moreljeanmichel@gmail.com)

Professor Bernard Teissier, Equipe Géométrie et Dynamique,  
Institut de Mathématiques de Jussieu – Paris Rive Gauche, Paris, France  
E-mail: [bernard.teissier@imj-prg.fr](mailto:bernard.teissier@imj-prg.fr)

Springer: Ute McCrory, Mathematics, Heidelberg, Germany,  
E-mail: [lnm@springer.com](mailto:lnm@springer.com)