

Bibliography:

Linear Algebra and Learning from Data

Printed publications and online references

- [1] C. Andersson and R. Bro, The N -Way Toolbox for MATLAB, www.models.kvl.dk/source **52** (2000) 1-4, **108**.
- [2] F. Bach, *Duality between subgradient and conditional gradient methods*, SIAM Journal of Optimization **25** (2015) 115-129; arXiv : 1211.6302, **343**.
- [3] B. Bader and T. Kolda, MATLAB Tensor Toolbox, version 2.2, <https://old-www.sandia.gov/~tgkolda/TensorToolbox/index-2.2.html> (2007), **108**.
- [4] R. Balestrierio and R. Baraniuk, *Mad Max: Affine spline insights into deep learning*, Proceedings of the IEEE **109** (2021) 704-727; arXiv : 1805.06576, **395**.
- [5] B. Beckermann, *The condition number of real Vandermonde, Krylov, and positive definite Hankel matrices*, Numerische Mathematik **85** (2000) 553-577, **182**.
- [6] B. Beckermann and A. Townsend, *On the singular values of matrices with displacement structure*, SIAM J. Matrix Analysis **38** (2017) 1227-1248; arXiv : 1609.09494, **182**.
- [7] M. Belkin, S. Ma, and S. Mandal, *To understand deep learning we need to understand kernel learning*, Proceedings of Machine Learning Research **80** (2018) 541-549; arXiv :1802.01396, **414**.
- [8] Y. Bengio, *Neural Networks : Tricks of the Trade* (2nd edition), edited by G. Montavon, G. Orr, and K.-R. Müller. Springer (2012), **408**.
- [9] P. Benner, R.-C. Li, and N. Truhar, *On the ADI method for Sylvester equations*, J. Comput. Appl. Math. **233** (2009) 1035-1045, **182**.
- [10] L. Berrada, A. Zisserman, and P. Kumar, *Trusting SVM for piecewise linear CNNs*, 5th Int. Conf. on Learning Representations (2017), 24-26; arXiv : 1611.02185, **394**.
- [11] D. Bertsekas, *Convex Analysis and Optimization*, Athena Scientific (2003), **356**.

- [12] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific (1986), **188**.
- [13] K. Borre and G. Strang, *Algorithms for Global Positioning*, Wellesley-Cambridge Press (2012), **165** and **302**.
- [14] L. Bottou et al, *Advances in Neural Information Processing Systems*, **16** (2004) and **20** (2008), MIT Press, **363**.
- [15] S. Boyd, http://stanford.edu/~boyd/papers/pdf/prox_algs.pdf (book and codes), **191**.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press (2004), **356**.
- [17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Foundations and Trends in Machine Learning, **3(1)** (2011) 1-122, **185**.
- [18] M. D. Buhmann, *Radial basis functions*, Acta Numerica **9** (2000) 1-38, **181**.
- [19] E. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Comp. Math. **9** (2009) 717-736; arXiv : 0805.4471, **198**.
- [20] E. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles : Exact signal reconstruction from highly incomplete Fourier information*, IEEE Transactions on Information Theory **52** (2006) 489-509, **198**.
- [21] E. J. Candès and M. B. Wakin, *An Introduction to Compressive Sampling*, IEEE Signal Processing Magazine **25** (2008) 21-30, **196**.
- [22] P. Combettes and J.- C. Pesquet, *Proximal splitting methods in signal processing*, in *Fixed-Point Algorithms for Inverse Problems*, Springer (2011) 185-212; arXiv : 0912.3522, **191**.
- [23] F. Curtis and K. Scheinberg, *Optimization Methods for Supervised Machine Learning : From Linear Models to Deep Learning*, The Operations Research Revolution (2017) 89-113; arXiv : 1706.10207, **367**.
- [24] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Mathematics of Control, Signals, and Systems **7** (1989) 303-314, **384**.
- [25] C. Davis and W. M. Kahan, *Some new bounds on perturbation of subspaces*, Bull. Amer. Math. Soc. **75** (1969) 863-868, **170**.
- [26] D. Donoho, *Notices of the American Math Society* (January 2018), **196**.
- [27] D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Th. **52** (2006) 1289-1306, **198**.
- [28] J. Duchi, E. Hazan, and Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization*, J. of Machine Learning Research **12** (2011) 2121-2159, **367**.

- [29] J. Eldridge, M. Belkin, and Y. Wang, *Unperturbed: Spectral analysis beyond Davis-Kahan*, Proceedings of Algorithmic Learning Theory, PMLR **83** (2018) 321-358; arXiv: 1706.06516, **170**.
- [30] B. Fornberg and N. Flyer, *A Primer on Radial Basis Functions with Applications to the Geosciences*, SIAM (2015) <https://doi.org/10.1137/1.9781611974041>, **181**.
- [31] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods*, North-Holland (1983-5), **188**.
- [32] D. Fortunato and A. Townsend, Fast Poisson solvers for spectral methods; arXiv: 1710.11259, **182**.
- [33] Y. Gal and Z. Ghahramani, *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, 33rd International Conference on Machine Learning, PMLR **48** (2016) 1050-1059; arXiv: 1506.02142, **409**.
- [34] N. Gillis, *The Why and How of Nonnegative Matrix Factorization*, Chapman & Hall/CRC, (2014) 257-291; arXiv: 1401.5226, **98**.
- [35] T. Goldstein and S. Osher, *The split Bregman method for L^1 regularized problems*, SIAM Journal of Imaging Sciences **2** (2009) 323-343, **193**.
- [36] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 4th Edition, **120**, **258**.
- [37] J. C. Gower, Properties of Euclidean and non-Euclidean distance matrices, *Linear Algebra and Its Applications* **67** (1985) 81-97, **261**.
- [38] J. C. Gower and G. B. Dijkstra, *Procrustes problems*, Oxford University Press (2004), **258**.
- [39] R. M. Gower and P. Richtarik, Randomized iterative methods for linear systems, *SIAM J. Matrix Analysis*. **36** (2015) 1660-1690; arXiv: 1506.03296, **363**.
- [40] P. Grinfeld, *Introduction to Tensor Calculus and the Calculus of Moving Surfaces*, Springer (2013), **108**.
- [41] W. W. Hager, *Updating the Inverse of a Matrix*, SIAM Review **31** (1989) 221-239, **167**.
- [42] D. Hajinezhad and Q. Shi, *ADMM for a class of nonconvex bilinear optimization*, Journal of Global Optimization **70** (2018) 261-288, **188**.
- [43] N. Halko, P.-G. Martinsson, and J. A. Tropp, Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review* **53** (2011) 217-288, **151**.
- [44] B. Hanin and D. Rolnick, *Deep ReLU Networks Have Surprisingly Few Activation Patterns*, Advances in Neural Information Processing Systems (NeurIPS) **32** (2019) 359-368; arXiv: 1906.00904, **383**.

- [45] B. Hanin and D. Rolnick, *How to start training : The effect of initialization and architecture*, Advances in Neural Information Processing Systems (NeurIPS) **31** (2018) 571-581; arXiv : 1803.01719, **378**.
- [46] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems : Numerical Aspects of Linear Inversion*, SIAM (1998), **368**.
- [47] M. Hardt, B. Recht, and Y. Singer, *Train faster, generalize better : Stability of stochastic gradient descent*, The 33rd International Conference on Machine Learning, PMLR **48** (2016) 1225-1234; arXiv : 1509.01240, **367**.
- [48] R. A. Harshman, <http://www.psychology.uwo.ca/faculty/harshman>, **108**
- [49] T. Hastie, B. Mazumder, J. Lee, and R. Zadeh, *Matrix completion and low-rank SVD via fast alternating least squares*, Journal of Machine Learning Research **16** (2015) 3367-3402; arXiv : 1410.2596, **198**.
- [50] K. He, X. Zhang, S. Ren, and J. Sun, *Identity mappings in deep residual networks*, https://doi.org/10.1007/978-3-319-46493-0_38 (2016); arXiv : 1603.05027, **395**.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, *Delving Deep into Rectifiers : Surpassing Human-Level Performance on ImageNet Classification*, IEEE International Conference on Computer Vision (ICCV), (2015) 1026-1034; arXiv : 1502.01852, **378**.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2016) 770-778; arXiv : 1512.03385, **395**.
- [53] C. F. Higham and D. J. Higham, *Deep Learning : An Introduction for Applied Mathematicians*, SIAM Review **61** (2019) 860-891; arXiv : 1801.05894, **402**.
- [54] D. Hilbert, <https://www.cmi.ac.in/smahanta/hilbert.html> (1900), The 13th problem asked a question about *all* continuous functions, **383**.
- [55] C. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, J. ACM **60** (2013) Article 45, **104**.
- [56] T. Hofmann, B. Scholkopf, and A. J. Smola, *Kernel methods in machine learning*, Annals of Statistics **36** (2008) 1171-1220, **181**, **414**.
- [57] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4** (1991) 251-257, **384**.
- [58] P. Izmailov, D. Podoprikin, T. Garipov, D. Vetrov, and A. Gordon Wilson, *Averaging weights leads to wider optima and better generalization*, 34th Conference on Uncertainty in Artificial Intelligence (2018) 876-885; arXiv : 1803.05407, **365**.
- [59] S. Johnson, *Adjoint methods and sensitivity analysis for recurrence relations*, Class notes : <http://math.mit.edu/~stevenj/18.336/recurrence2.pdf>. Also online : *Notes on adjoint methods for 18.335*, **406**.

- [60] R. Kannan and S. Vempala, Randomized algorithms in numerical linear algebra, *Acta Numerica* **26** (2017) 95-135, **151, 155**.
- [61] A. Karpathy, [cs231n.github.io/convolutionalnetworks/](https://github.com/karpathy/convolutionalnetworks) (karpathy@cs.stanford.edu), **393**
- [62] P. Kingma and J. Ba, ADAM: A method for stochastic optimization, ICLR (2015), **367**.
- [63] T. Kolda and B. Bader, *Tensor decompositions and applications*, SIAM Review **52** (2009) 455–500, **108**.
- [64] A. Krizhevsky, I. Sutskever, and G. Hinton, *ImageNet classification with deep convolutional neural networks*, NIPS (2012) 1106–1114, **393**.
- [65] C.-C. J. Kuo, *The CNN as a guided multilayer RECOs transform*, IEEE Signal Proc. Mag. **34** (2017) 81-89; arXiv: 1701.08481, **395**.
- [66] L. De Lathauwer, B. de Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21** (2000) 1253–1278 and 1324–1342 (*and more recent papers on tensors*), **108**.
- [67] Q. Le : *Building high-level features using large scale unsupervised learning* : arXiv : 1112.6209, **415**.
- [68] Y. LeCun and Y. Bengio, *Convolutional networks for images, speech, and time-series*, *Handbook of Brain Theory and Neural Networks*, MIT Press (1998), **393**.
- [69] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature **401** (1999) 788-791, **97, 108**.
- [70] Y. Lei and D.-X. Zhou, *Learning theory of randomized sparse Kaczmarz method*, SIAM J. Imaging Sciences **11** (2018) 547-574, **193**.
- [71] L. Lessard, B. Recht, and A. Packard. *Analysis and design of optimization algorithms via integral quadratic constraints*, SIAM Journal on Optimization **26** (2016) 57-95; arXiv : 1408.3595, **356**.
- [72] G. Lewis-Kraus, Deep learning and the development of Google Translate, New York Times Magazine, 14 December 2016. <https://www.nytimes.com/2016/12/14-magazine/the-great-AI-awakening.html>, **415**
- [73] E. Liberty, F. Woolfe, P. -G. Martinsson, V. Rokhlin, and M. Tygert, *Randomized algorithms for the low-rank approximation of matrices*, PNAS **104** (2007) 51, 20167-20172, **151**.
- [74] A. Ma, D. Needell and A. Ramdas, *Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods*, **36** (2015) 1590-1604; arXiv : 1503.08235, **363**.

- [75] S. Ma, R. Bassily, and M. Belkin, *The power of interpolation : Understanding the effectiveness of SGD in modern over-parametrized learning*, 35th International Conference on Machine Learning, PMLR **80** (2018) 3325-3334; arXiv : 1712.06559, **363**.
- [76] M. Mahoney, M. Maggioni, and P. Drineas, *Tensor-CUR decompositions for tensor-based data*, SIAM J. Matrix Analysis Appl. **30** (2008) 957–987, **108**.
- [77] M. W. Mahoney, *Lecture Notes on Randomized Linear Algebra*, (2016); arXiv :1608.04481, **151**.
- [78] S. Mallat, *Understanding deep convolutional networks*, Phil. Trans. Roy. Soc. **374** (2016); arXiv : 1601.04920, **395**.
- [79] P.-G. Martinsson, 2016 *Course Notes for APPM 5720*, https://amath.colorado.edu/faculty/martinss/Teaching/APPM5720_2016s/index.html **130**.
- [80] P.-G. Martinsson, *Randomized methods for matrix computations*, American Mathematical Society **25** 187-229 (2018); arXiv : 1607.01649, **155**.
- [81] P.-G. Martinsson, *Compressing rank-structured matrices via randomized sampling*, SIAM J. Sci. Comput., **38** A1959-A1986; arXiv :1503.07152, **151**.
- [82] P. G. Martinsson, G. Quintana-Ortí, and N. Heavner, *randUTV : A blocked randomized algorithm for computing a rank-revealing UTV factorization*, ACM Transactions on Mathematical Software **45** (2019) 1-26; arXiv 1703.00998, **67**.
- [83] P.-G. Martinsson and J. A. Tropp, *Randomized numerical linear algebra : Foundations and algorithms*, Acta Numerica **29** (2020) 403-572; arXiv : 2002.01387, **151**.
- [84] H. Mhaskar, Q. Liao, and T. Poggio, *When and why are deep networks better than shallow ones?*, Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (2017) 2343-2349; arXiv : 01603.00988, **384**.
- [85] R. R. Nadakuditi, *Optshrink : An algorithm for improved low-rank signal matrix denoising by optimal, data-driven shrinkage*, IEEE Transactions on Information Theory **60** (May 2014) 3002-3018, **170**.
- [86] R. R. Nadakuditi, *When are the most informative components for inference also the principal components?* arXiv : 1302.1232, **171**.
- [87] Y. Nakatsukasa, *Fast and stable randomized low-rank matrix approximation*, (2020); arXiv :2009.11392, **151**, **155**.
- [88] D. Needell, N. Srebro, and R. Ward, *Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm*, Math. Progr. **155** (2015) 549-573, **363**.
- [89] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Springer (2004), **356**.
- [90] M. E. J. Newman, *Modularity and community structure in networks*, PNAS **103** (2006) 8577-8582, **246**.

- [91] M. Nielsen, *Neural Networks and Deep Learning*, online book at <http://neuralnetworksanddeeplearning.com>, Determination Press (2015), **411**.
- [92] J. Nocedal and S. Wright, *Numerical Optimization*, Springer (1999), **356**.
- [93] C. Olah, *Computational graphs* (colah.github.io) and (<https://distill.pub>).
- [94] I. Oseledets, *Tensor-train decomposition*, SIAM J. Sci. Comp. **33** (2011) 2295-2317, **108**.
- [95] P. Paatero and U. Tapper, *Positive matrix factorization*, Environmetrics **5** (1994) 111-126, **108**.
- [96] N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends in Optimization **1** (2013) 123-321, **191**.
- [97] R. Pascanu, G. Montufar, and Y. Bengio, *On the number of response regions of deep feed forward networks with piece-wise linear activations*, (2013); arXiv 1312.6098, **381**.
- [98] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein, *On the Expressive Power of Deep Neural Networks*, 34th International Conference on Machine Learning, PMLR **70** (2017) 2847-2854; arXiv : 1606.05336, **381**.
- [99] S. Ragnarsson and C. Van Loan, *Block tensor unfoldings*, SIAM J. Matrix Anal. Appl. **33** (2013) 149–169; arXiv : 1101.2005, **108**.
- [100] S. Reddi, M. Zaheer, D. Sachan, S. Kale, and S. Kumar, *Adaptive methods for non-convex optimization*, 32nd International Conference on Neural Information Processing Systems (NeurIPS) **31** (2018) 9815-9825, **367**.
- [101] S. Reddi, S. Kale, and S. Kumar, *On the convergence of ADAM and beyond*, Proc. Intl. Conference on Learning Representations, ICLR (2018); arXiv : 1904.09237, **367**.
- [102] B. Recht, <https://simons.berkeley.edu/talks/ben-recht-2013-09-04>, **356**.
- [103] S. Reddi, S. Sra, B. Póczos, and A. Smola, *Fast stochastic methods for nonsmooth non-convex optimization*, Advances in Neural Information Processing Systems **29** (2016) 1145-1153; arXiv : 1605.06900, **363**.
- [104] H. Robbins and S. Monro, *A Stochastic Approximation Method*, Annals of Mathematical Statistics **22** (1951) 400-407, **361**.
- [105] D. Rolnick and M. Tegmark, *The power of deeper networks for expressing natural functions*, ICLR (2018); arXiv : 1705.05502, **385**.
- [106] S. Ruder, *An overview of gradient descent optimization algorithms*, (2016); arXiv : 1609.04747, **367**.

- [107] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, *Physica D* **60** (1992) 259-268, **193**.
- [108] D. Ruiz-Antolin and A. Townsend, *A nonuniform fast Fourier transform based on low rank approximation*, *SIAM J. Sci. Comp.* **40-1** (2018) A529-A547; arXiv : 1701.04492, **182**.
- [109] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, *ICLR* (2015); arXiv : 1409.1556, **392**, **393**.
- [110] Daniel Smilkov, playground.tensorflow.org (back cover), **386**.
- [111] Z. Song, D. P. Woodruff, and P. Zhong, *Relative Error Tensor Low Rank Approximation*, *Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* (2019) 2772-2789; arXiv : 1704.08246, **108**.
- [112] D. C. Sorensen and M. Embree, *A DEIM induced CUR factorization*, *SIAM J. Scientific Computing* **38** (2016) 1454-1482; arXiv : 1407.5516, **142**.
- [113] N. Srebro, ttic.uchicago.edu/~nati/Publications/thesis.pdf, **75**
- [114] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout”, *Journal of Machine Learning Research* **15** (2014) 1929-1958, **409**.
- [115] G. W. Stewart, *Four algorithms for the efficient computation of truncated pivoted QR approximations to a sparse matrix*, *Numerische Mathematik* **83** (1999) 313-323, **143**.
- [116] T. Strohmer and R. Vershynin, *A randomized Kaczmarz algorithm with exponential convergence*, *J. Fourier Anal. Appl.* **15** (2009) 262-278; arXiv : 0702226, **122**, **363**.
- [117] T. Tao, <https://terrytao.wordpress.com/2010/01/12/254a-notes-3a-eigenvalues-and-sums-of-hermitian-matrices/>, **175**
- [118] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, *Journal of the Royal Statistical Society, Series B* **58** (1996) 267-288, **100**.
- [119] A. Townsend and H. Wilber, *On the singular values of matrices with high displacement rank*, *Linear Alg. Appl.* **548** (2018) 19-41; arXiv :17120.5864, **182**.
- [120] A. Townsend, www.math.cornell.edu/~ajt/presentations/LowRankMatrices.pdf, **182**
- [121] J. A. Tropp, *An Introduction to Matrix Concentration Inequalities*, *Foundations and Trends in Machine Learning* **8** (2015) 1-230; arXiv :1501.01571, **291**.
- [122] M. Udell and A. Townsend, *Nice latent variable models have log-rank*, *SIAM J. Math. of Data Science* **1** (2019) 144-160; arXiv : 1705.07474, **182**.
- [123] C. Van Loan, www.alm.unibo.it/~simoncin/CIME/vanloan1.pdf—[vanloan4.pdf](http://www.alm.unibo.it/~simoncin/CIME/vanloan4.pdf), **108**
- [124] C. Van Loan, *The Kronecker Product*, <https://www.cs.cornell.edu/cv/ResearchPDF/KPhist.pdf>, **226**.

- [125] C. Van Loan, *The ubiquitous Kronecker product*, J. Comp. Appl. Math. **123** (2000) 85–100, **226**.
- [126] A. Veit, M. Wilber, and S. Belongie, *Residual networks behave like ensembles of relatively shallow networks*, Proceedings of the 30th International Conference on Neural Information Processing Systems (2016) 550-558; arXiv : 1605.06431, **395**.
- [127] A. C. Wilson, B. Recht, and M. Jordan, *A Lyapunov analysis of momentum methods in optimization*, Journal of Machine Learning Research **22** (2021) 1-34; arXiv : 1611.02635, **356**.
- [128] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, *The marginal value of adaptive gradient methods in machine learning*, 31st International Conference on Neural Information Processing Systems (2017) 4151-4161; arXiv :1705.08292, **356**, **367**.
- [129] D. P. Woodruff, *Sketching as a tool for numerical linear algebra*. Foundations and Trends in Theoretical Computer Science **10** (2014) 1-157, **151**.
- [130] L. Xiao, Y. Bahri, and J. Sohl-Dickstein, S. Schoenholz, and J. Pennington, *Dynamical isometry and a mean field theory of CNNs: How to train 10,000 layers*, 35th International Conference on Machine Learning, PMLR **80** (2018) 5393-5402 ; arXiv : 1806.05393, **378**.
- [131] L. Xu, B. Yu, and Y. Zhang, *An alternating direction and projection algorithm for structure-enforced matrix factorization*, Computational Optimization Appl. **68** (2017) 333-362, **98**.
- [132] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for ℓ^1 minimization with applications to compressed sensing*, SIAM J. Imaging Sciences **1** (2008) 143-168, **193**.
- [133] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, *Understanding deep learning (still) requires rethinking generalization*, Communications of the ACM **64** (2021) 107-115; arXiv : 1611.03530, **356**.
- [134] D. Zhang, Lu Lu, L. Guo, and G. Em Karniadakis, *Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems*, Journal of Computational Physics **397** (2019); arXiv : 1809.08327, **409**.
- [135] D.-X. Zhou, *Universality of deep convolutional neural networks*, Applied and Computational Harmonic Analysis, **48** (2020) 787-794; arXiv : 1805.10769, **385**.
- [136] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society, Series B **67** (2005) 301-320, **100**.
- [137] H. Zou, T. Hastie, and R. Tibshirani, *Sparse principal component analysis*, Journal of Computational and Graphical Statistics **15** (2006) 265-286, **99**.