

Bibliography

- [1] D. Achlioptas and F. McSherry. On spectral learning of mixtures of distributions. In *COLT*, pages 458–469, 2005.
- [2] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, and R. Tandon. Learning sparsely used overcomplete dictionaries via alternating minimization. *arXiv:1310.7991*, 2013.
- [3] A. Agarwal, A. Anandkumar, and P. Netrapalli. Exact recovery of sparsely used overcomplete dictionaries. *arXiv:1309.1952*, 2013.
- [4] M. Aharon. Overcomplete dictionaries for sparse representation of signals. PhD thesis, 2006.
- [5] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54(11):4311–4322, 2006.
- [6] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory*, 48(3):569–579, 2002.
- [7] N. Alon. Tools from higher algebra. In *Handbook of Combinatorics*, editors: R. L. Graham, M. Grötschel, and L. Lovász. Cambridge, MA: MIT Press, 1996, pages 1749–1783.
- [8] A. Anandkumar, D. Foster, D. Hsu, S. Kakade, and Y. Liu. A spectral algorithm for latent Dirichlet allocation. In *NIPS*, pages 926–934, 2012.
- [9] A. Anandkumar, R. Ge, D. Hsu, and S. Kakade. A tensor spectral approach to learning mixed membership community models. In *COLT*, pages 867–881, 2013.
- [10] A. Anandkumar, D. Hsu, and S. Kakade. A method of moments for hidden Markov models and multi-view mixture models. In *COLT*, pages 33.1–33.34, 2012.
- [11] J. Anderson, M. Belkin, N. Goyal, L. Rademacher, and J. Voss. The more the merrier: The blessing of dimensionality for learning large Gaussian mixtures. *arXiv:1311.2891*, 2013.
- [12] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, Y. Wu, and M. Zhu. A practical algorithm for topic modeling with provable guarantees. In *ICML*, pages 280–288, 2013.
- [13] S. Arora, R. Ge, R. Kannan, and A. Moitra. Computing a nonnegative matrix factorization – provably. In *STOC*, pages 145–162, 2012.

- [14] S. Arora, R. Ge, and A. Moitra. Learning topic models – going beyond SVD. In *FOCS*, pages 1–10, 2012.
- [15] S. Arora, R. Ge, and A. Moitra. New algorithms for learning incoherent and overcomplete dictionaries. *arXiv:1308.6273*, 2013.
- [16] S. Arora, R. Ge, T. Ma, and A. Moitra. Simple, efficient, and neural algorithms for sparse coding. In *COLT*, pages 113–149, 2015.
- [17] S. Arora, R. Ge, A. Moitra, and S. Sachdeva. Provable ICA with unknown Gaussian noise, and implications for Gaussian mixtures and autoencoders. In *NIPS*, pages 2384–2392, 2012.
- [18] S. Arora, R. Ge, S. Sachdeva, and G. Schoenebeck. Finding overlapping communities in social networks: Towards a rigorous approach. In *EC*, 2012.
- [19] S. Arora and R. Kannan. Learning mixtures of separated nonspherical Gaussians. *Ann. Appl. Probab.*, 15(1A):69–92, 2005.
- [20] M. Balcan, A. Blum, and A. Gupta. Clustering under approximation stability. *J. ACM*, 60(2): 1–34, 2013.
- [21] M. Balcan, A. Blum, and N. Srebro. On a theory of learning with similarity functions. *Mach. Learn.*, 72(1–2):89–112, 2008.
- [22] M. Balcan, C. Borgs, M. Braverman, J. Chayes, and S.-H. Teng. Finding endogenously formed communities. In *SODA*, 2013.
- [23] A. Bandeira, P. Rigollet, and J. Weed. Optimal rates of estimation for multi-reference alignment. *arXiv:1702.08546*, 2017.
- [24] B. Barak, S. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *FOCS*, pages 428–437, 2016.
- [25] B. Barak, J. Kelner, and D. Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *STOC*, pages 143–151, 2015.
- [26] B. Barak and A. Moitra. Noisy tensor completion via the sum-of-squares hierarchy. In *COLT*, pages 417–445, 2016.
- [27] M. Belkin and K. Sinha. Toward learning Gaussian mixtures with arbitrary separation. In *COLT*, pages 407–419, 2010.
- [28] M. Belkin and K. Sinha. Polynomial learning of distribution families. In *FOCS*, pages 103–112, 2010.
- [29] Q. Berthet and P. Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *COLT*, pages 1046–1066, 2013.
- [30] A. Bhaskara, M. Charikar, and A. Vijayaraghavan. Uniqueness of tensor decompositions with applications to polynomial identifiability. In *COLT*, pages 742–778, 2014.
- [31] A. Bhaskara, M. Charikar, A. Moitra, and A. Vijayaraghavan. Smoothed analysis of tensor decompositions. In *STOC*, pages 594–603, 2014.
- [32] Y. Bilu and N. Linial. Are stable instances easy? In *Combinatorics, Probability and Computing*, 21(5):643–660, 2012.
- [33] V. Bittorf, B. Recht, C. Re, and J. Tropp. Factoring nonnegative matrices with linear programs. In *NIPS*, 2012.
- [34] D. Blei. Introduction to probabilistic topic models. *Commun. ACM*, 55(4):77–84, 2012.

- [35] D. Blei and J. Lafferty. A correlated topic model of science. *Ann. Appl. Stat.*, 1(1):17–35, 2007.
- [36] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.
- [37] A. Blum, A. Kalai, and H. Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *J. ACM*, 50:506–519, 2003.
- [38] A. Blum and J. Spencer. Coloring random and semi-random k -colorable graphs. *Journal of Algorithms*, 19(2):204–234, 1995.
- [39] K. Borgwardt. *The Simplex Method: A Probabilistic Analysis*. New York: Springer, 2012.
- [40] S. C. Brubaker and S. Vempala. Isotropic PCA and affine-invariant clustering. In *FOCS*, pages 551–560, 2008.
- [41] E. Candes and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2008.
- [42] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.
- [43] E. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005.
- [44] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):1–37, 2011.
- [45] V. Chandrasekaran and M. Jordan. Computational and statistical tradeoffs via convex relaxation. *Proc. Natl. Acad. Sci. U.S.A.*, 110(13):E1181–E1190, 2013.
- [46] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.
- [47] J. Chang. Full reconstruction of Markov models on evolutionary trees: Identifiability and consistency. *Math. Biosci.*, 137(1):51–73, 1996.
- [48] K. Chaudhuri and S. Rao. Learning mixtures of product distributions using correlations and independence. In *COLT*, pages 9–20, 2008.
- [49] K. Chaudhuri and S. Rao. Beyond Gaussians: Spectral methods for learning mixtures of heavy-tailed distributions. In *COLT*, pages 21–32, 2008.
- [50] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.
- [51] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k -term approximation. *J. AMS*, 22(1):211–231, 2009.
- [52] J. Cohen and U. Rothblum. Nonnegative ranks, decompositions and factorizations of nonnegative matrices. *Linear Algebra Appl.*, 190:149–168, 1993.
- [53] P. Comon. Independent component analysis: A new concept? *Signal Processing*, 36(3):287–314, 1994.
- [54] A. Dasgupta. *Asymptotic Theory of Statistics and Probability*. New York: Springer, 2008.
- [55] A. Dasgupta, J. Hopcroft, J. Kleinberg, and M. Sandler. On learning mixtures of heavy-tailed distributions. In *FOCS*, pages 491–500, 2005.
- [56] S. Dasgupta. Learning mixtures of Gaussians. In *FOCS*, pages 634–644, 1999.
- [57] S. Dasgupta and L. J. Schulman. A two-round variant of EM for Gaussian mixtures. In *UAI*, pages 152–159, 2000.

- [58] G. Davis, S. Mallat, and M. Avellaneda. Greedy adaptive approximations. *Constr. Approx.*, 13:57–98, 1997.
- [59] L. De Lathauwer, J. Castaing, and J. Cardoso. Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Trans. Signal Process.*, 55(6):2965–2973, 2007.
- [60] S. Deerwester, S. Dumais, T. Landauer, G. Furnas, and R. Harshman. Indexing by latent semantic analysis. *J. Assoc. Inf. Sci. Technol.*, 41(6):391–407, 1990.
- [61] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Series B Stat. Methodol.*, 39(1):1–38, 1977.
- [62] D. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 -minimization. *Proc. Natl. Acad. Sci. U.S.A.*, 100(5):2197–2202, 2003.
- [63] D. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory*, 47(7):2845–2862, 1999.
- [64] D. Donoho and P. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, 1989.
- [65] D. Donoho and V. Stodden. When does nonnegative matrix factorization give the correct decomposition into parts? In *NIPS*, 2003.
- [66] R. Downey and M. Fellows. *Parameterized Complexity*. New York: Springer, 2012.
- [67] M. Elad. *Sparse and Redundant Representations*. New York: Springer, 2010.
- [68] K. Engan, S. Aase, and J. Hakon-Husoy. Method of optimal directions for frame design. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 5:2443–2446, 1999.
- [69] P. Erdos, M. Steel, L. Szekely, and T. Warnow. A few logs suffice to build (almost) all trees. I. *Random Struct. Algorithms*, 14:153–184, 1997.
- [70] M. Fazel. Matrix rank minimization with applications. PhD thesis, Stanford University, 2002.
- [71] U. Feige and J. Kilian. Heuristics for semirandom graph problems. *J. Comput. Syst. Sci.*, 63(4):639–671, 2001.
- [72] U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Struct. Algorithms*, 16(2):195–208, 2009.
- [73] J. Feldman, R. A. Servedio, and R. O'Donnell. PAC learning axis-aligned mixtures of Gaussians with no separation assumption. In *COLT*, pages 20–34, 2006.
- [74] A. Frieze, M. Jerrum, and R. Kannan. Learning linear transformations. In *FOCS*, pages 359–368, 1996.
- [75] A. Garnaev and E. Gluskin. The widths of a Euclidean ball. *Sov. Math. Dokl.*, 277(5):200–204, 1984.
- [76] R. Ge and T. Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. In *RANDOM*, pages 829–849, 2015.
- [77] A. Gilbert, S. Muthukrishnan, and M. Strauss. Approximation of functions over redundant dictionaries using coherence. In *SODA*, pages 243–252, 2003.
- [78] N. Gillis. Robustness analysis of hotttopixx, a linear programming model for factoring nonnegative matrices. *arXiv:1211.6687*, 2012.

- [79] N. Goyal, S. Vempala, and Y. Xiao. Fourier PCA. In *STOC*, pages 584–593, 2014.
- [80] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *arXiv:0910.1879*, 2009.
- [81] D. Gross, Y.-K. Liu, S. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(15):150401, 2010.
- [82] V. Guruswami, J. Lee, and A. Razborov. Almost Euclidean subspaces of ℓ_1^n via expander codes. *Combinatorica*, 30(1):47–68, 2010.
- [83] M. Hardt. Understanding alternating minimization for matrix completion. In *FOCS*, pages 651–660, 2014.
- [84] R. Harshman. Foundations of the PARAFAC procedure: model and conditions for an “explanatory” multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.
- [85] J. Håstad. Tensor rank is *NP*-complete. *J. Algorithms*, 11(4):644–654, 1990.
- [86] C. Hillar and L.-H. Lim. Most tensor problems are *NP*-hard. *arXiv:0911.1393v4*, 2013.
- [87] T. Hofmann. Probabilistic latent semantic analysis. In *UAI*, pages 289–296, 1999.
- [88] R. Horn and C. Johnson. *Matrix Analysis*. New York: Cambridge University Press, 1990.
- [89] D. Hsu and S. Kakade. Learning mixtures of spherical Gaussians: Moment methods and spectral decompositions. In *ITCS*, pages 11–20, 2013.
- [90] P. J. Huber. Projection pursuit. *Ann. Stat.*, 13:435–475, 1985.
- [91] R. A. Hummel and B. C. Gidas. Zero crossings and the heat equation. Courant Institute of Mathematical Sciences, TR-111, 1984.
- [92] R. Impagliazzo and R. Paturi. On the complexity of *k*-SAT. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001.
- [93] P. Jain, P. Netrapalli, and S. Sanghavi. Low rank matrix completion using alternating minimization. In *STOC*, pages 665–674, 2013.
- [94] A. T. Kalai, A. Moitra, and G. Valiant. Efficiently learning mixtures of two Gaussians. In *STOC*, pages 553–562, 2010.
- [95] R. Karp. Probabilistic analysis of some combinatorial search problems. In *Algorithms and Complexity: New Directions and Recent Results*. New York: Academic Press, 1976, pages 1–19.
- [96] B. Kashin and V. Temlyakov. A remark on compressed sensing. Manuscript, 2007.
- [97] L. Khachiyan. On the complexity of approximating extremal determinants in matrices. *J. Complexity*, 11(1):138–153, 1995.
- [98] D. Koller and N. Friedman. *Probabilistic Graphical Models*. Cambridge, MA: MIT Press, 2009.
- [99] J. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions with applications to arithmetic complexity and statistics. *Linear Algebra Appl.*, 18(2):95–138, 1977.
- [100] A. Kumar, V. Sindhwani, and P. Kambadur. Fast conical hull algorithms for near-separable non-negative matrix factorization. In *ICML*, pages 231–239, 2013.

- [101] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [102] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000.
- [103] S. Leurgans, R. Ross, and R. Abel. A decomposition for three-way arrays. *SIAM J. Matrix Anal. Appl.*, 14(4):1064–1083, 1993.
- [104] M. Lewicki and T. Sejnowski. Learning overcomplete representations. *Comput.*, 12:337–365, 2000.
- [105] W. Li and A. McCallum. Pachinko allocation: DAG-structured mixture models of topic correlations. In *ICML*, pp. 633–640, 2007.
- [106] B. Lindsay. *Mixture Models: Theory, Geometry and Applications*. Hayward, CA: Institute for Mathematical Statistics, 1995.
- [107] B. F. Logan. Properties of high-pass signals. PhD thesis, Columbia University, 1965.
- [108] L. Lovász and M. Saks. Communication complexity and combinatorial lattice theory. *J. Comput. Syst. Sci.*, 47(2):322–349, 1993.
- [109] F. McSherry. Spectral partitioning of random graphs. In *FOCS*, pages 529–537, 2001.
- [110] S. Mallat. *A Wavelet Tour of Signal Processing*. New York: Academic Press, 1998.
- [111] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.
- [112] A. Moitra. An almost optimal algorithm for computing nonnegative rank. In *SODA*, pages 1454–1464, 2013.
- [113] A. Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *STOC*, pages 821–830, 2015.
- [114] A. Moitra and G. Valiant. Setting the polynomial learnability of mixtures of Gaussians. In *FOCS*, pages 93–102, 2010.
- [115] E. Mossel and S. Roch. Learning nonsingular phylogenies and hidden Markov models. In *STOC*, pages 366–375, 2005.
- [116] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. New York: Springer, 2004.
- [117] B. Olshausen and B. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37(23):3311–3325, 1997.
- [118] C. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala. Latent semantic indexing: A probabilistic analysis. *J. Comput. Syst. Sci.*, 61(2):217–235, 2000.
- [119] Y. Pati, R. Rezaifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. *Asilomar Conference on Signals, Systems, and Computers*, pages 40–44, 1993.
- [120] K. Pearson. Contributions to the mathematical theory of evolution. *Philos. Trans. Royal Soc. A*, 185: 71–110, 1894.
- [121] Y. Rabani, L. Schulman, and C. Swamy. Learning mixtures of arbitrary distributions over large discrete domains. In *ITCS*, pages 207–224, 2014.
- [122] R. Raz. Tensor-rank and lower bounds for arithmetic formulas. In *STOC*, pages 659–666, 2010.

- [123] B. Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12:3413–3430, 2011.
- [124] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.
- [125] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, 26(2):195–239, 1984.
- [126] J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. *J. Symb. Comput.*, 13(1):255–352, 1991.
- [127] T. Rockafellar. *Convex Analysis*. Princeton, NJ: Princeton University Press, 1996.
- [128] A. Seidenberg. A new decision method for elementary algebra. *Ann. Math.*, 60(2):365–374, 1954.
- [129] V. de Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low rank approximation problem. *SIAM J. Matrix Anal. Appl.*, 30(3):1084–1127, 2008.
- [130] D. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. In *Journal of the ACM*, 51(3):385–463, 2004.
- [131] D. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. *J. Mach. Learn. Res.*, 23:1–18, 2012.
- [132] N. Srebro and A. Shraibman. Rank, trace-norm and max-norm. In *COLT*, pages 545–560, 2005.
- [133] M. Steel. Recovering a tree from the leaf colourations it generates under a Markov model. *Appl. Math. Lett.*, 7:19–24, 1994.
- [134] A. Tarski. A decision method for elementary algebra and geometry. Berkeley and Los Angeles: University of California Press, 1951.
- [135] H. Teicher. Identifiability of mixtures. *Ann. Math. Stat.*, 31(1):244–248, 1961.
- [136] J. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.
- [137] J. Tropp, A. Gilbert, S. Muthukrishnan, and M. Strauss. Improved sparse approximation over quasi-incoherent dictionaries. *IEEE International Conference on Image Processing*, 1:37–40, 2003.
- [138] L. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.
- [139] S. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM J. Optim.*, 20(3):1364–1377, 2009.
- [140] S. Vempala and Y. Xiao. Structure from local optima: Learning subspace juntas via higher order PCA. *arXiv:abs/1108.3329*, 2011.
- [141] S. Vempala and G. Wang. A spectral algorithm for learning mixture models. *J. Comput. Syst. Sci.*, 68(4):841–860, 2004.
- [142] M. Wainwright and M. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2): 1–305, 2008.
- [143] P. Wedin. Perturbation bounds in connection with singular value decompositions. *BIT Numer. Math.*, 12:99–111, 1972.
- [144] M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. Syst. Sci.*, 43(3):441–466, 1991.