

# References for the “Algorithmic High Dimensional Geometry” lectures at the Big Data Boot Camp, Simons Institute, Berkeley

Alexandr Andoni  
Microsoft Research SVC

September 4, 2013

Lecture 1: [AM93, Cla94, AMN<sup>+</sup>98, Kle97, HP01, Cha02a, IM98, HIM12, KOR00, GIM99, Cha02b, DIIM04, CR04, Pan06, AC09, AI06b, AI08, JL84, AMS99, Ach03, TZ04, AL08, AL11, DKS10, KN12, CS02, BC03, LN04, NR10, ANN10, Ind01, MNP06, OWZ11, PTW08, PTW10, AIP06, AI06a, Bro97, BGMZ97, AINR13].

Lecture 2: [AMN<sup>+</sup>98, KR02, Ass83, Cla99, GKL03, KL04, BW06, Cla08, CNBYM01, FK97, IN07, Cla06, HPM06, BKL06, Cha02b, IT03, SA12, AIK08, Ind07, OR07, CK06, MS00, CM07, NS07, KN06, KR06, AK10, Cor03, Enf69, Mat96, Ind98, FCI99, JL01, AIK09, Ind02, Ind06, EKK<sup>+</sup>00, ACCL07, GJKK07, GG07, EJ08, Yao79, KN97, AMS99, KOR00, IW03, Woo04, JKS08, BJKS04, AHPV04, AJP10, PRR95, Dub10, Val12, LLR94, Bou85, Rao99, ARV04, ALN05, KMS98].

## References

- [AC09] Nir Ailon and Bernard Chazelle. The fast johnson–lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009.
- [ACCL07] Nir Ailon, Bernard Chazelle, Seshadhri Comandur, and Ding Liu. Estimating the distance to a monotone function. *Random Structures and Algorithms*, 31:371–383, 2007. Previously appeared in RANDOM’04.
- [Ach03] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. System Sci.*, 66(4):671–687, 2003. Special issue on PODS 2001 (Santa Barbara, CA).
- [AHPV04] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Approximating extent measure of points. *Journal of the ACM*, pages 606–635, 2004.
- [AI06a] Alexandr Andoni and Piotr Indyk. Efficient algorithms for substring near neighbor problem. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1203–1212, 2006.
- [AI06b] Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 459–468, 2006.

- [AI08] Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of the ACM*, 51(1):117–122, 2008.
- [AIK08] Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Earth mover distance over high-dimensional spaces. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 343–352, 2008.
- [AIK09] Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Overcoming the  $\ell_1$  non-embeddability barrier: Algorithms for product metrics. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 865–874, 2009.
- [AINR13] Alexandr Andoni, Piotr Indyk, Huy L. Nguyen, and Ilya Razenshteyn. Beyond locality-sensitive hashing. *arXiv:1306.1547*, 2013. <http://arxiv.org/abs/1306.1547>.
- [AIP06] Alexandr Andoni, Piotr Indyk, and Mihai Pătraşcu. On the optimality of the dimensionality reduction method. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 449–458, 2006.
- [AJP10] Alexandr Andoni, T.S. Jayram, and Mihai Pătraşcu. Lower bounds for edit distance and product metrics via Poincaré-type inequalities. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010.
- [AK10] Alexandr Andoni and Robert Krauthgamer. The computational hardness of estimating edit distance. *SIAM Journal on Computing*, 39(6):2398–2429, 2010. Previously appeared in FOCS’07.
- [AL08] Nir Ailon and Edo Liberty. Fast dimension reduction using Rademacher series on dual BCH codes. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1–9, 2008.
- [AL11] Nir Ailon and Edo Liberty. An almost optimal unrestricted fast Johnson-Lindenstrauss transform. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 185–191, 2011.
- [ALN05] Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 553–562, 2005.
- [AM93] Sunil Arya and David M. Mount. Approximate nearest neighbor searching. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 271–280, 1993.
- [AMN<sup>+</sup>98] Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, and Angela Y. Wu. An optimal algorithm for approximate nearest neighbor searching. *J. ACM*, 6(45):891–923, 1998. Previously appeared in SODA’94.
- [AMS99] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. Comp. Sys. Sci.*, 58:137–147, 1999. Previously appeared in STOC’96.

- [ANN10] Alexandr Andoni, Assaf Naor, and Ofer Neiman. Improved upper bounds for  $\ell_1$  dimension reduction. 2010. Manuscript.
- [ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 222–231, 2004.
- [Ass83] Patrice Assouad. Plongements lipschitziens dans  $\mathbf{R}^n$ . *Bull. Soc. Math. France*, 111(4):429–448, 1983.
- [BC03] Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in  $\ell_1$ . In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 2003.
- [BGMZ97] Andrei Broder, Steve Glassman, Mark Manasse, and Geoffrey Zweig. Syntactic clustering of the web. *Proceedings of the Sixth International World Wide Web Conference*, pages 391–404, 1997.
- [BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. Previously in FOCS’02.
- [BKL06] Alina Beygelzimer, Sham Kakade, and John Langford. Cover trees for nearest neighbor. In *Proceedings of the Twenty-Third International Conference on Machine Learning*, pages 97–104, 2006.
- [Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.
- [Bro97] Andrei Broder. On the resemblance and containment of documents. *Proceedings of Compression and Complexity of Sequences*, pages 21–29, 1997.
- [BW06] Richard G. Baraniuk and Michael B. Wakin. Random projections of smooth manifolds. In *Foundations of Computational Mathematics*, pages 941–944, 2006.
- [Cha02a] Timothy M. Chan. Closest-point problems simplified on the RAM. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 472–473, 2002.
- [Cha02b] Moses Charikar. Similarity estimation techniques from rounding. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 380–388, 2002.
- [CK06] Moses Charikar and Robert Krauthgamer. Embedding the Ulam metric into  $\ell_1$ . *Theory of Computing*, 2(11):207–224, 2006.
- [Cla94] Ken Clarkson. An algorithm for approximate closest-point queries. *Proceedings of the Tenth Annual ACM Symposium on Computational Geometry*, pages 160–164, 1994.
- [Cla99] Ken Clarkson. Nearest neighbor queries in metric spaces. *Discrete Comput. Geom.*, 22(1):63–93, 1999. Previously in SoCG97.

- [Cla06] Kenneth L. Clarkson. Nearest-neighbor searching and metric space dimensions. In Gregory Shakhnarovich, Trevor Darrell, and Piotr Indyk, editors, *Nearest-Neighbor Methods for Learning and Vision: Theory and Practice*, pages 15–59. MIT Press, 2006.
- [Cla08] Kenneth L. Clarkson. Tighter bounds for random projections of manifolds. In *Proceedings of the ACM Symposium on Computational Geometry (SoCG)*, 2008.
- [CM07] Graham Cormode and S. Muthukrishnan. The string edit distance matching problem with moves. *ACM Transactions on Algorithms*, 3(1), 2007. Previously in SODA’02.
- [CNBYM01] Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, and José L. Marroquin. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
- [Cor03] Graham Cormode. *Sequence Distance Embeddings*. Ph.D. Thesis, University of Warwick. 2003.
- [CR04] Amit Chakrabarti and Oded Regev. An optimal randomised cell probe lower bounds for approximate nearest neighbor searching. *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 2004.
- [CS02] Moses Charikar and Amit Sahai. Dimension reduction in the  $\ell_1$  norm. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 551–560, 2002.
- [DIIM04] Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. *Proceedings of the ACM Symposium on Computational Geometry (SoCG)*, 2004.
- [DKS10] Anirban Dasgupta, Ravi Kumar, and Tamás Sarlós. A sparse Johnson Lindenstrauss transform. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 341–350, 2010.
- [Dub10] Moshe Dubiner. Bucketing coding and information theory for the statistical high dimensional nearest neighbor problem. *IEEE Transactions on Information Theory*, 56(8):4166–4179, 2010.
- [EJ08] Funda Ergün and Hossein Jowhari. On distance to monotonicity and longest increasing subsequence of a data stream. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 730–736, 2008.
- [EKK<sup>+</sup>00] Funda Ergün, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, and Manesh Viswanathan. Spot-checkers. *J. Comput. Syst. Sci.*, 60(3):717–751, 2000.
- [Enf69] P. Enflo. On the nonexistence of uniform homeomorphisms between  $L_p$ -spaces. *Ark. Mat.*, 8:103–105, 1969.
- [FCI99] Martin Farach-Colton and Piotr Indyk. Approximate nearest neighbor algorithms for hausdorff metrics via embeddings. *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 1999.

- [FK97] Christos Faloutsos and Ibrahim Kamel. Relaxing the uniformity and independence assumptions using the concept of fractal dimension. *Journal of Computer and System Sciences*, 55(2):229–240, 1997.
- [GG07] Anna Gál and Parikshit Gopalan. Lower bounds on streaming algorithms for approximating the length of the longest increasing subsequence. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 294–304, 2007.
- [GIM99] Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Similarity search in high dimensions via hashing. *Proceedings of the 25th International Conference on Very Large Data Bases (VLDB)*, 1999.
- [GJKK07] Parikshit Gopalan, T. S. Jayram, Robert Krauthgamer, and Ravi Kumar. Estimating the sortedness of a data stream. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 318–327, 2007.
- [GKL03] A. Gupta, R. Krauthgamer, and J. Lee. Bounded geometries, fractals, and low-distortion embeddings. *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 2003.
- [HIM12] Sariel Har-Peled, Piotr Indyk, and Rajeev Motwani. Approximate nearest neighbor: Towards removing the curse of dimensionality. *Theory of Computing*, 1(8):321–350, 2012.
- [HP01] Sariel Har-Peled. A replacement for voronoi diagrams of near linear size. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 94–103, 2001.
- [HPM06] Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM J. Comput.*, 35(5), 2006.
- [IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbor: towards removing the curse of dimensionality. *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 604–613, 1998.
- [IN07] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. *ACM Transactions on Algorithms*, 2007.
- [Ind98] P. Indyk. On approximate nearest neighbors in non-euclidean spaces. *Proceedings of the Symposium on Foundations of Computer Science*, pages 148–155, 1998.
- [Ind01] Piotr Indyk. *High-dimensional computational geometry*. Ph.D. Thesis. Department of Computer Science, Stanford University, 2001.
- [Ind02] Piotr Indyk. Approximate nearest neighbor algorithms for Frechet metric via product metrics. *Proceedings of the ACM Symposium on Computational Geometry (SoCG)*, pages 102–106, 2002.
- [Ind06] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. *J. ACM*, 53(3):307–323, 2006. Previously appeared in FOCS’00.

- [Ind07] Piotr Indyk. A near linear time constant factor approximation for euclidean bichromatic matching (cost). In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007.
- [IT03] Piotr Indyk and Nitin Thaper. Fast color image retrieval via embeddings. *Workshop on Statistical and Computational Theories of Vision (at ICCV)*, 2003.
- [IW03] Piotr Indyk and David Woodruff. Tight lower bounds for the distinct elements problem. *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 283–290, 2003.
- [JKS08] TS Jayram, Ravi Kumar, and D Sivakumar. The one-way communication complexity of hamming distance. *Theory of Computing*, 4(1):129–135, 2008.
- [JL84] William B. Johnson and Joram Lindenstrauss. Extensions of lipshitz mapping into hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
- [JL01] William B. Johnson and Joram Lindenstrauss, editors. *Handbook of the geometry of Banach spaces. Vol. I*. North-Holland Publishing Co., Amsterdam, 2001.
- [KL04] Robert Krauthgamer and James R. Lee. Navigating nets: Simple algorithms for proximity search. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2004.
- [Kle97] Jon Kleinberg. Two algorithms for nearest-neighbor search in high dimensions. *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, 1997.
- [KMS98] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *Journal of the ACM*, 1998.
- [KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.
- [KN06] Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. *Math. Ann.*, 334(4):821–852, 2006. Preliminary version appeared in FOCS’05.
- [KN12] Daniel M. Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss transforms. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1195–1206, 2012.
- [KOR00] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 30(2):457–474, 2000. Preliminary version appeared in STOC’98.
- [KR02] D. Karger and M. Ruhl. Finding nearest neighbors in growth-restricted metrics. *Proceedings of the Symposium on Theory of Computing (STOC)*, 2002.
- [KR06] Robert Krauthgamer and Yuval Rabani. Improved lower bounds for embeddings into  $L_1$ . In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1010–1017, 2006.

- [LLR94] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 577–591, 1994.
- [LN04] J. R. Lee and A. Naor. Embedding the diamond graph in  $L_p$  and dimension reduction in  $L_1$ . *Geom. Funct. Anal.*, 14(4):745–747, 2004.
- [Mat96] Jiří Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics*, 93:333–344, 1996.
- [MNP06] Rajeev Motwani, Assaf Naor, and Rina Panigrahy. Lower bounds on locality sensitive hashing. In *Proceedings of the ACM Symposium on Computational Geometry (SoCG)*, pages 253–262, 2006.
- [MS00] S. Muthukrishnan and S. C. Sahinalp. Approximate nearest neighbors and sequence comparisons with block operations. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 416–424, 2000.
- [NR10] Ilan Newman and Yuri Rabinovich. Finite volume spaces and sparsification. *arXiv:1002.3541v3*, 2010. <http://arxiv.org/abs/1002.3541>.
- [NS07] Assaf Naor and Gideon Schechtman. Planar earthmover is not in  $L_1$ . *SIAM Journal on Computing*, 37(3):804–826, 2007. An extended abstract appeared in FOCS’06.
- [OR07] Rafail Ostrovsky and Yuval Rabani. Low distortion embedding for edit distance. *J. ACM*, 54(5), 2007. Preliminary version appeared in STOC’05.
- [OWZ11] Ryan O’Donnell, Yi Wu, and Yuan Zhou. Optimal lower bounds for locality sensitive hashing (except when  $q$  is tiny). In *Innovations in Computer Science*, pages 275–283, 2011.
- [Pan06] Rina Panigrahy. Entropy-based nearest neighbor algorithm in high dimensions. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2006.
- [PRR95] Ramamohan Paturi, Sanguthevar Rajasekaran, and John Reif. The light bulb problem. *Information and Computation*, 117(2):187–192, 1995.
- [PTW08] Rina Panigrahy, Kunal Talwar, and Udi Wieder. A geometric approach to lower bounds for approximate near-neighbor search and partial match. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 414–423, 2008.
- [PTW10] Rina Panigrahy, Kunal Talwar, and Udi Wieder. Lower bounds on near neighbor search via metric expansion. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 805–814, 2010.
- [Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.

- [SA12] R. Sharathkumar and Pankaj K. Agarwal. A near-linear time  $\epsilon$ -approximation algorithm for geometric bipartite matching. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 385–394, 2012.
- [TZ04] M. Thorup and Y. Zhang. Tabulation based 4-universal hashing with applications to second moment estimation. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2004.
- [Val12] Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and juntas with noise. In *FOCS. Preliminary version as ECCV TR12-006*, 2012.
- [Woo04] David Woodruff. Optimal space lower bounds for all frequency moments. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 167–175, 2004.
- [Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.