The Fast Johnson-Lindenstrauss Transform and Applications

Nir Ailon

School of Mathematics, Institute for Advanced Study, Princeton NJ, USA
nailon@math.ias.edu
www.math.ias.edu/~nailon

Abstract

Dimension reduction is a highly useful tool in algorithm design, with applications in nearest neighbor searching, clustering, streaming, sketching, learning, approximation algorithms, vision and others. It removes redundancy from data and can be plugged into algorithms suffering from a "curse of dimensionality".

In my talk, I will describe a novel technique for reducing the dimension of points in Euclidean space, improving a now classic algorithm by Johnson and Lindenstrauss from the mid 80's. Our technique is the first to offer an asymptotic improvement, and has already been used in design of efficient algorithms for nearest neighbor searching and high dimensional linear algebraic numerical computations.

I will present our algorithm, its proof, applications, and interesting open questions. Joint work with Bernard Chazelle.

1 Introduction

Most of the talk will be based on a paper by the speaker with Bernard Chazelle [1]. Other related work [2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13].

References

[1] N. Ailon and B. Chazelle Approximate Nearest Neighbors and the Fast Johnson-Lindenstrauss Transform *In Proceedings ofSTOC'06; To appear in SICOMP*

[2] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003.

[3] E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Knowledge Discovery and Data Mining*, pages 245–250, 2001.

[4] S. DasGupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. *Technical Report, UC Berkeley*, 99-006, 1999.

[5] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory Series A*, 44:355–362, 1987.

[6] P. Indyk. Dimensionality reduction techniques for proximity problems. In *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 371–378, 2000.

[7] P. Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 189–197, 2000.

[8] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of L2 into L1. Technical Report TR06-126, ECCC, 2006.

[9] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC)*, pages 604–613, 1998.

[10] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[11] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM Journal on Computing*, 30(2):457–474, 2000.

[12] S. Muthukrishnan and S. C. Sahinalp. Simple and practical sequence nearest neighbors with block operations. In *Proceedings of the 13th Annual Symposium on Combinatorial Pattern Matching (CPM)*, pages 262–278, 2002.

[13] T. Sarlós. Improved approximation algorithms for large matrices via random projections. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, Berkeley, CA, 2006.