The Johnson Lindenstrauss Lemma

History and context, focusing on Jelani Nelson and major Stanford contributors: there are many others, this lecture focuses on Stanford so you can see how it ties to research and people here

Moses Charikar

Stanford CS114

Spring 2022

58th Annual IEEE Symposium on Foundations of Computer Science

Optimality of the Johnson-Lindenstrauss lemma

Kasper Green Larsen
Computer Science Department
Aarhus University
Aarhus, Denmark
larsen@cs.au.dk

Jelani Nelson
SEAS
Harvard University
Cambridge, MA, USA
minilek@seas.harvard.edu

***Abstract*—For any $d, n \geq 2$ and $1/(\min\{n, d\})^{0.4999} < \varepsilon < 1$, we show the existence of a set of n vectors $X \subset \mathbb{R}^d$ such that any embedding $f : X \to \mathbb{R}^m$ satisfying**

$$\forall x, y \in X,\ (1-\varepsilon)\|x-y\|_2^2 \leq \|f(x)-f(y)\|_2^2 \leq (1+\varepsilon)\|x-y\|_2^2$$

must have

$$m = \Omega(\varepsilon^{-2} \lg n).$$

This lower bound matches the upper bound given by the Johnson-Lindenstrauss lemma [JL84]. Furthermore, our lower bound holds for nearly the full range of ε of interest, since there is always an isometric embedding into dimension $\min\{d, n\}$ (either the identity map, or projection onto $span(X)$).

Previously such a lower bound was only known to hold against *linear* maps f, and not for such a wide range of parameters ε, n, d [LN16]. The best previously known lower bound for general f was $m = \Omega(\varepsilon^{-2} \lg n / \lg(1/\varepsilon))$ [Wel74], [Alo03], which is suboptimal for any $\varepsilon = o(1)$.

Johnson and Lindenstrauss [JL84], it was proven that for any $\varepsilon < 1/2$, there exists n point sets $X \subset \mathbb{R}^n$ for which any embedding $f : X \to \mathbb{R}^m$ providing (1) must have $m = \Omega(\lg n)$. This was later improved in [Alo03], which showed the existence of an n point set $X \subset \mathbb{R}^n$, such that any f providing (1) must have $m = \Omega(\min\{n, \varepsilon^{-2} \lg n / \lg(1/\varepsilon)\})$, which falls short of the JL lemma for any $\varepsilon = o(1)$. This lower bound can also be obtained from the Welch bound [Wel74], which states $\varepsilon^{2k} \geq (1/(n-1))(n/\binom{m+k-1}{k} - 1)$ for any positive integer k, by choosing $2k = \lceil \lg n / \lg(1/\varepsilon) \rceil$. The lower bound can also be extended to hold for any $n \leq e^{c\varepsilon^2 d}$ for some constant $c > 0$.

Our Contribution: In this paper, we finally settle the optimality of the JL lemma. Furthermore, we do so for almost the full range of ε.

Q & A

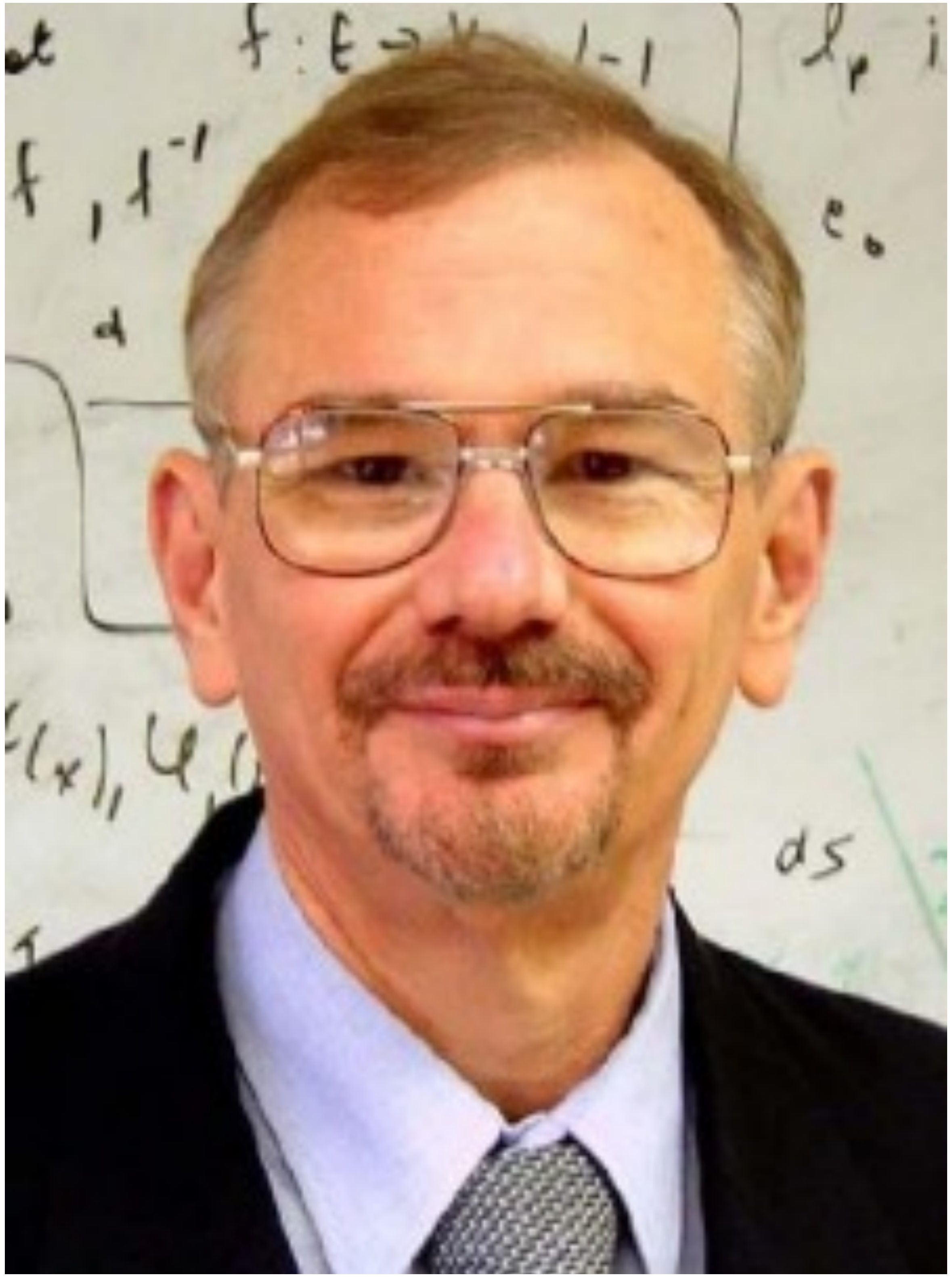
The Computer Scientist Who Shrinks Big Data

Jelani Nelson designs clever algorithms that only have to remember slivers of massive data sets. He also teaches kids in Ethiopia how to code.

Computer scientist Jelani Nelson with his daughter at their home in Berkeley, California.

Constanza Hevia for Quanta Magazine

The Johnson Lindenstrauss Lemma



http://dx.doi.org/10.1090/conm/026/737400

Contemporary Mathematics
Volume 26, 1984

EXTENSIONS OF LIPSCHITZ MAPPINGS INTO A HILBERT SPACE

William B. Johnson[1] and Joram Lindenstrauss[2]

INTRODUCTION

In this note we consider the following extension problem for Lipschitz functions: Given a metric space X and $n = 2, 3, 4, \ldots$, estimate the smallest constant $L = L(X, n)$ so that every mapping f from every n-element subset of X into ℓ_2 extends to a mapping $\tilde{f}$ from X into ℓ_2 with

$$\|\tilde{f}\|_{\ell ip} \leq L \, \|f\|_{\ell ip} \; .$$

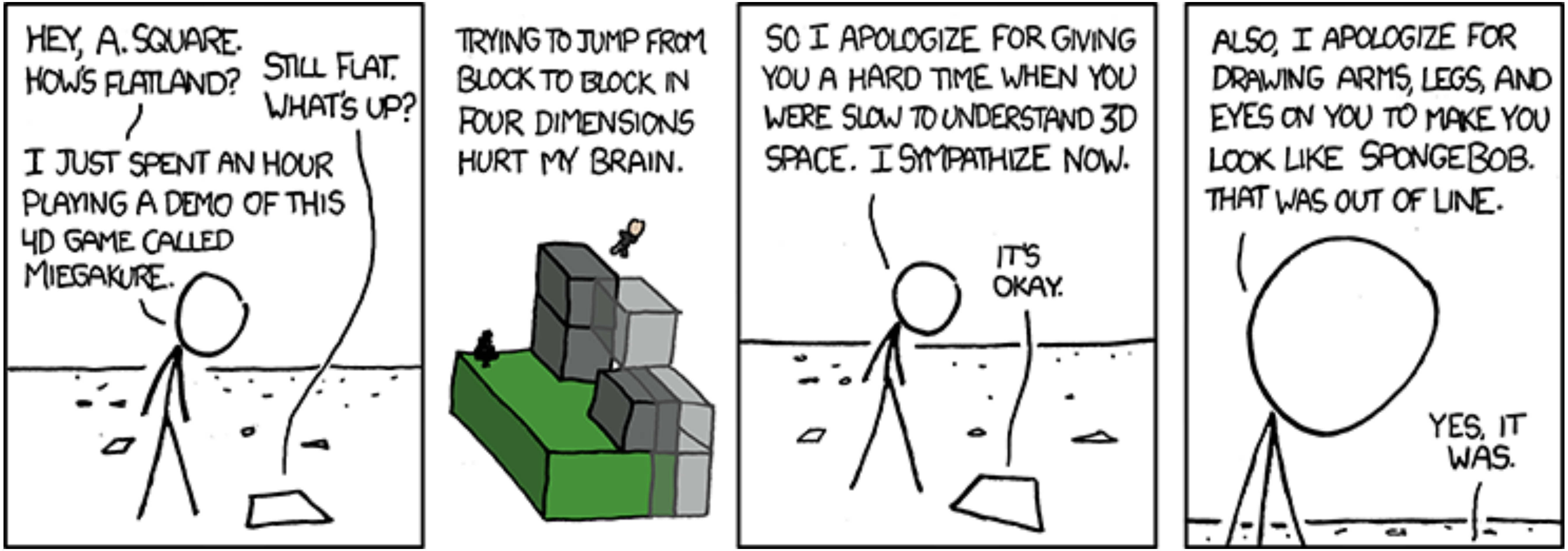
(Here $\|g\|_{\ell ip}$ is the Lipschitz constant of the function g.) A classical result of Kirszbraun's [14, p. 48] states that $L(\ell_2, n) = 1$ for all n, but it is easy to see that $L(X, n) \to \infty$ as $n \to \infty$ for many metric spaces X.

We begin with the geometrical lemma mentioned in the introduction.

LEMMA 1. *For each* $1 > \tau > 0$ *there is a constant* $K = K(\tau) > 0$ *so that if* $A \subset \ell_2^n$, $\overline{\overline{A}} = n$ *for some* $n = 2, 3, \ldots$, *then there is a mapping* f *from* A *onto a subset of* ℓ_2^k $(k \equiv [K \log n])$ *which satisfies*

$$\|\tilde{f}\|_{\ell ip} \, \|\tilde{f}^{-1}\|_{\ell ip} \le \frac{1 + \tau}{1 - \tau} .$$

Curse of Dimensionality


HEY, A. SQUARE. HOW'S FLATLAND?
STILL FLAT. WHAT'S UP?
I JUST SPENT AN HOUR PLAYING A DEMO OF THIS 4D GAME CALLED MIEGAKURE.
TRYING TO JUMP FROM BLOCK TO BLOCK IN FOUR DIMENSIONS HURT MY BRAIN.
SO I APOLOGIZE FOR GIVING YOU A HARD TIME WHEN YOU WERE SLOW TO UNDERSTAND 3D SPACE. I SYMPATHIZE NOW.
IT'S OKAY.
ALSO, I APOLOGIZE FOR DRAWING ARMS, LEGS, AND EYES ON YOU TO MAKE YOU LOOK LIKE SPONGEBOB. THAT WAS OUT OF LINE.
YES, IT WAS.

Near Neighbor Search in Large Metric Spaces

Sergey Brin*
Department of Computer Science
Stanford University
sergey@cs.stanford.edu

Abstract

Given user data, one often wants to find approximate matches in a large database. A good example of such a task is finding images similar to a given image in a large collection of images. We focus on the important and technically difficult case where each data element is high dimensional, or more generally, is represented by a point in a large metric space and distance calculations are computationally expensive.

the triangle inequality (see Section 3). Hence metric spaces are a very general concept and can be applied to vectors (for example, under Euclidean distance) as well as objects like strings and graphs which cannot be easily represented as vectors (if at all). Finding near neighbors in a metric space refers to selecting the elements of a data set (a finite subset of the space) which are within a certain distance of a given point.

The problem of finding the near neighbors in a large data set has been studied well and has a number of good solutions, *if* the data is in a simple (e.g. Euclidean), low-dimensional vector space. However, if the data lies in a large metric space the problem becomes much more difficult. By a *large* metric space


Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality

Piotr Indyk* Rajeev Motwani†

Department of Computer Science

Stanford University

Stanford, CA 94305

{indyk,rajeev}@cs.stanford.edu

Abstract

The *nearest neighbor* problem is the following: Given a set of n points $P = \{p_1, \ldots, p_n\}$ in some metric space X, preprocess P so as to efficiently answer queries which require finding the point in P closest to a query point $q \in X$. We focus on the particularly interesting case of the d-dimensional Euclidean space where $X = \Re^d$ under some l_p norm. Despite decades of effort, the current solutions are far from satisfactory; in fact, for large d, in theory or in practice, they provide little improvement over the brute-force algorithm which compares the query point to each data point. Of late, there has been some interest in the *approximate nearest neighbors* problem, which is: Find a point $p \in P$ that is an ϵ-approximate nearest neighbor of the query q in that for all $p' \in P$, $d(p, q) \leq (1 + \epsilon)d(p', q)$.

1 Introduction

The nearest neighbor search (NNS) problem is: Given a set of n points $P = \{p_1, \ldots, p_n\}$ in a metric space X with distance function d, preprocess P so as to efficiently answer queries for finding the point in P closest to a query point $q \in X$. We focus on the particularly interesting case of the d-dimensional Euclidean space where $X = \Re^d$ under some l_p norm. The low-dimensional case is well-solved [26], so the main issue is that of dealing with the "curse of dimensionality" [16]. The problem was originally posed in the 1960s by Minsky and Papert [53, pp. 222–225], and despite decades of effort the current solutions are far from satisfactory. In fact, for large d, in theory or in practice, they provide little improvement over a brute-force algorithm which compares a query q to each $p \in P$. The known algorithms are of two

A The Dimension Reduction Technique

We first outline our proof for the random projections technique for dimension reduction. Combining this with Proposition 2, we obtain the result given in Proposition 3.

Definition 8 *Let $\mathcal{M} = (X,d)$ and $\mathcal{M}' = (X',d')$ be two metric spaces. The space $\mathcal{M}$ is said to have a c-*isometric embedding*, or simply a c-*embedding*, in $\mathcal{M}'$ if there exists a map $f : \mathcal{M} \to \mathcal{M}'$ such that*

$$(1-\epsilon)d(p,q) < d'(f(p),f(q)) < (1+\epsilon)d(p,q)$$

for all $p,q \in X$. We call c the distortion *of the embedding; if $c = 1$, we call the embedding* isometric.

Frankl and Maehara [32] gave the following improvement to the Johnson-Lindenstrauss Lemma [41] on $(1+\epsilon)$-embedding of any $S \subset l_2^d$ in $l_2^{O(\log|S|)}$.

Lemma 6 (Frankl-Maehara [32]) *For any $0 < \epsilon < \frac{1}{2}$, any (sufficiently large) set S of points in $\Re^d$, and $k = \lceil 9(\epsilon^2 - 2\epsilon^3/3)^{-1} \ln|S| \rceil + 1$, there exists a map $f : S \to \Re^k$ such that for all $u,v \in S$,*

$$(1-\epsilon)||u-v||^2 < ||f(u)-f(v)||^2 < (1+\epsilon)||u-v||^2.$$

The proof proceeds by showing that the square of the length of a projection of any unit vector v on a random k

where P_t^α is a random variable following the Poisson distribution with parameter αt. Bounding the latter quantity is a matter of simple calculation. ■

An interesting question is if the Johnson-Lindenstrauss Lemma holds for other l_p norms. A partial answer is provided by the following two results.

Theorem 5 *For any $p \in [1,2]$, any n-point set $S \subset l_p^d$, and any $\epsilon > 0$, there exist a map $f : S \to l_2^k$ with $k = O(\log n)$ such that for all $u,v \in S$.*

$$(1-\epsilon)||u-v||_p < ||f(u)-f(v)||^2 < (1+\epsilon)||u-v||_p.$$

Theorem 6 *The Johnson-Lindenstrauss Lemma does not hold for l_∞. More specifically, there is a set S of n points in $\Re^d$ for some d such that any embedding of S in $\Re^f$ has distortion $\Omega(\frac{\log n}{\sqrt{f}})$.*

Proof Sketch: We give a sketch of the proof of Theorem 6 based on the following two known facts.

Fact 3 (Linial, London, and Rabinovich [48]) *Every n-point metric $\mathcal{M}$ can be isometrically embedded in l_∞^n.*

Fact 4 (Linial, London, and Rabinovich [48]) *There are graphs with n vertices which for any d cannot be embedded in l_2^d with distortion $o(\log n)$.*

An Elementary Proof of a Theorem of Johnson and Lindenstrauss

Sanjoy Dasgupta,[1] Anupam Gupta[2]

[1]AT&T Labs Research, Room A277, Florham Park, New Jersey 07932; e-mail: dasgupta@research.att.com

[2]Lucent Bell Labs, Room 2C-355, 600 Mountain Avenue, Murray Hill, New Jersey 07974; e-mail: anupamg@research.bell-labs.com

Lower bounds

Combinatorics, Probability and Computing (2009) 18, 3–15.
doi:10.1017/S0963548307008917 Printed in the United Kingdom

Perturbed Identity Matrices Have High Rank: Proof and Applications

NOGA ALON†

Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences,
Tel Aviv University, Tel Aviv 69978, Israel
(e-mail: nogaa@tau.ac.il)

Received 15 November 2006; revised 17 November 2007; first published online 16 January 2008

We describe a lower bound for the rank of any real matrix in which all diagonal entries are significantly larger in absolute value than all other entries, and discuss several applications of this result to the study of problems in Geometry, Coding Theory, Extremal Finite Set Theory and Probability. This is partly a survey, containing a unified approach for proving various known results, but it contains several new results as well.

3. Distortion in low-dimension embeddings

A well-known lemma of Johnson and Lindenstrauss, proved in [11] (see also [15]), asserts that for any $\epsilon > 0$, any set A of n points in an Euclidean space can be embedded in an Euclidean space of dimension $k = c(\epsilon) \log n$ with distortion at most ϵ. That is, there is a mapping $f : A \mapsto R^k$ such that for any $a, b \in A$, the distance between $f(a)$ and $f(b)$ is at least the distance between a and b, and at most that distance multiplied by $1 + \epsilon$. The proof gives that $c(\epsilon) \leqslant O(\frac{1}{\epsilon^2})$. Theorem 2.1 can be used to show that this is nearly tight: $c(\epsilon)$ must be at least $\Omega\left(\frac{1}{\epsilon^2 \log(1/\epsilon)}\right)$, even for embedding the set of points of a simplex. This is stated in the following proposition, proved in [1].

Proposition 3.1. *Let $P_0, P_1, \ldots, P_n$ be a set of $n+1$ points in R^k, and suppose that the distance between any two of them is at least 1 and at most $1+\epsilon$, where $\frac{1}{\sqrt{n}} \leqslant \epsilon \leqslant \frac{1}{10}$. Then $k \geqslant \frac{c'}{\epsilon^2 \log(1/\epsilon)} \log n$, where c' is an absolute positive constant.* □

Many Applications

To appear as a part of an upcoming textbook on dimensionality reduction and manifold learning.

Johnson-Lindenstrauss Lemma, Linear and Nonlinear Random Projections, Random Fourier Features, and Random Kitchen Sinks: Tutorial and Survey

Benyamin Ghojogh BGHOJOGH@UWATERLOO.CA
Department of Electrical and Computer Engineering,
Machine Learning Laboratory, University of Waterloo, Waterloo, ON, Canada

Ali Ghodsi ALI.GHODSI@UWATERLOO.CA
Department of Statistics and Actuarial Science & David R. Cheriton School of Computer Science,
Data Analytics Laboratory, University of Waterloo, Waterloo, ON, Canada

Fakhri Karray KARRAY@UWATERLOO.CA
Department of Electrical and Computer Engineering,
Centre for Pattern Analysis and Machine Intelligence, University of Waterloo, Waterloo, ON, Canada

Mark Crowley MCROWLEY@UWATERLOO.CA
Department of Electrical and Computer Engineering,
Machine Learning Laboratory, University of Waterloo, Waterloo, ON, Canada