On Fast Johnson-Lindenstrauss Embeddings

TopMath
Mathematik mit Promotion

Stefan Bamberger, Advisor: Prof. Dr. Felix Krahmer

Unit M15: Applied Numerical Analysis and Optimization and Data Analysis
Technische Universität München

Technische Universität München
Elitenetzwerk Bayern
Universität Augsburg University

Abstract

Johnson-Lindenstrauss embeddings are random maps from $\mathbb{R}^N$ to an $\mathbb{R}^m$ of lower dimension that preserve the norms of p (finitely many) given points. This approach can improve the computational complexity of problems by reducing the dimension of the containing space. Thus, we aim to minimize the dimension of the reduced space as well as the computation time for the transformation itself. Currently existing fast constructions either have an upper bound for the number of points p to preserve or an embedding dimension that is not optimal. We present a new construction for a Johnson-Lindenstrauss embedding that has an optimal embedding dimension and also a significantly higher upper bound for p. However, it accomplishes a fast transformation for all p points together rather than for each single point.

Introduction – Johnson-Lindenstrauss Embeddings

Johnson-Lindenstrauss embeddings are random transformations that reduce the dimension of data. We want these transformations to preserve the ℓ_2-norm of all points within a given finite set. For this norm preservation we tolerate a small deviation and aim to achieve it with a high probability. This leads to the following definition.

Definition (Johnson-Lindenstrauss Embedding). *Let $A \in \mathbb{R}^{m\times N}$ be a random matrix where $m < N$, $\epsilon, \eta \in (0,1)$ and $p \in \mathbb{Z}_{\geq 1}$. We say that A is a (p,ϵ,η)-JLE (Johnson-Lindenstrauss embedding) if for any subset $E \subseteq \mathbb{R}^N$ with $|E| = p$*

$$(1-\epsilon)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\epsilon)\|x\|_2^2 \qquad (1)$$

holds simultaneously for all $x \in E$ with a probability of at least $1-\eta$.

This notation of m, N, p, ϵ, η and E will be used throughout this work.
Note that randomness is necessary for this property to hold. For a deterministic matrix A, we can take an $x \in \ker A \backslash \{0\}$ (since $m < N$) and then $\|Ax\| = 0$ and (1) cannot hold for any $\epsilon < 1$.
Johnson-Lindenstrauss Embeddings have various applications in problems that can benefit from reduced data dimension. Examples include the Approximate Nearest Neighbor Search [4], the problem of finding the point in a finite set that is closest to a given position. Also, the matrices from Johnson-Lindenstrauss embeddings have interesting properties in data reconstruction [7].

Goals for the Construction of JLEs

As many applications benefit from a lower computational time due to the reduced data dimension, we pursue reaching a low embedding dimension along with a fast computation of the dimension reduction. Our goals are thus:

1. Optimal embedding dimension: This has been proven to be $\Theta(\epsilon^{-2}\log p)$, [6], this is independent of N.
2. Fast transformation of a **single point**: We aim to transform each single point with the Johnson-Lindenstrauss embedding in time $\mathcal{O}(N(\log N)^k)$ with fixed k.
3. Fast transformation of **all p points in E together**: We aim to transform all points from the set E in an overall time of $\mathcal{O}(pN(\log N)^k)$.

The third goal is a weaker version of the second one. However, it will turn out to be easier to achieve in some cases.

Existing Constructions for JLEs

All of the following existing Johnson-Lindenstrauss embeddings can achieve a transformation of a single point in the desired time of $\mathcal{O}(N\log N)$. However, there is an upper bound for the number p of points to be preserved and not all of them have the optimal embedding dimension.

Construction	Optimal Dimension?	Upper Bound for p
Bernoulli ± 1[1]	yes	$\mathcal{O}(N^k)$
Ailon, Liberty 2009[5]	yes	$\exp(\mathcal{O}(N^{\frac{1}{2}-\delta}))$
Krahmer, Ward 2011[2]	no	none (only a general bound)

The first construction simply uses independent entries, each taking the value ± 1 with probability $\frac{1}{2}$ each. The other two approaches are based on the discrete Fourier transform.
The given bound for Ailon, Liberty 2009 works for any fixed $\delta > 0$.
Based on the minimal embedding dimension, one can show that all Johnson-Lindenstrauss embeddings have the upper bound of $p \leq \exp(\mathcal{O}(N))$ for the number of points.

This means, for JLEs of optimal embedding dimension, there is still a gap for the values $p = \exp(\Theta(N^{\frac{1}{2}}))$ up to $p = \exp(\theta(N^1))$ between the construction of Ailon, Liberty and the general upper bound. We intend to close this gap with a new construction.

A New Construction for a Fast JLE I

Our new construction is based on the following lemma, stating that the composition of two JLEs is a JLE again with slightly larger parameter values.

Lemma (Composition of JLEs, S. B., Krahmer, 2017). *Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times N}$ be independent random matrices that are both $(p, \frac{\epsilon}{3}, \frac{\eta}{4})$-JLEs. Then $AB \in \mathbb{R}^{m\times N}$ is a (p,ϵ,η)-JLE.*

A New Construction for a Fast JLE II

The previous lemma can be used to show the Johnson-Lindenstrauss property for the following matrix

$$GA \in \mathbb{R}^{m\times N}$$

where

- $A \in \mathbb{R}^{n\times N}$ is the construction by Krahmer and Ward [2]. It can map points from the original $\mathbb{R}^N$ into a space of a reduced, but not yet optimal dimension n.
- $G \in \mathbb{R}^{m\times n}$ ist the construction using Bernoulli ± 1 matrices [1]. This consecutively maps points from the space $\mathbb{R}^n$ into a space of optimal dimension m.

For the choices $m = \Theta(\epsilon^{-2}\log\frac{p}{\eta})$ and $n = \Theta(\epsilon^{-2}\log(\frac{p}{\eta})(\log N)^4)$ one can prove G and A to be JLEs, so also the composition GA is a JLE. This means, we get the optimal embedding dimension m for fixed η and thus this achieves our first goal.

As pointed out before, the computation of A is always fast. The matrix G can cause difficulties for the fast transformation. However, due to the precedent dimension reduction by A, the matrix G here is smaller than a single complete Bernoulli matrix. Counting all the operations necessary for the transformation of one point, we get

- $\mathcal{O}(N\log N)$ for $p \leq \exp(\mathcal{O}(N^{\frac{1}{2}-\delta}))$ for every fixed δ.
- $\mathcal{O}(N(\log N)^4)$ for $p \leq \exp(\mathcal{O}(N^{\frac{1}{2}}))$.

The first case provides precisely the same bound as the construction by Ailon and Liberty [5]. However, the second case is an improvement as their construction does not work any more if the exponent of N is precisely $\frac{1}{2}$.

Speeding Up JLEs Using Fast Matrix Multiplication

The problem causing the upper bound for p in the previous construction GA is that we use the standard matrix-vector multiplication algorithm for G that requires too many operations if p and hence also the number of rows in G is too large.
Using faster algorithms for a general matrix-matrix multiplication, we can improve the complexity.
These algorithms benefit from matrices with many columns. So instead of computing the transformation GAx for a single point x, we compute the matrix-matrix product GAM_E where $M_E \in \mathbb{R}^{p\times N}$ consists of all vectors in E as its columns. This corresponds to the simultaneous transformation of all the points in the set E.
Using tools from the theory of fast matrix multiplication algorithms by Lotti and Romani [3], the following result can be shown.

Result (S. B., Krahmer, 2017). *For every fixed $\delta > 0$, there is an algorithm that can compute the transformation by GA of all p points in E together in time $\mathcal{O}(pN\log N)$ if $p \leq \exp(\mathcal{O}(N^{1-\delta}))$.*

Considering the general upper bound of $p \leq \exp(\mathcal{O}(N))$, this result can essentially exhaust the entire range of possible values for p.

Discussion

As demonstrated, the new construction can slightly improve the upper bound on p required for a fast transformation of a single point. It improves the simultaneous transformation of all p points to a larger extent. However, this new approach also has the follwing restrictions.

- The constants hidden in the $\mathcal{O}$-notation are very large. This can cause problems in practical computations.
- Depending on the application, the simultaneous transformation of all p points might not be useful.

Thus, it is still an important open research problem to construct a JLE that has an optimal embedding dimension and allows a fast transformation of a single point for the remaining values $p = \exp(\Theta(N^r))$ for $\frac{1}{2} < r \leq 1$.

References

[1] Dimitris Achlioptas. "Database-Friendly Random Projections: Johnson-Lindenstrauss with Binary Coins". In: *Journal of Computer and System Sciences* (2003).
[2] Felix Krahmer, Rachel Ward. "New and Improved Johnson-Lindenstrauss Embeddings via the Restricted Isometry Property". In: *SIAM J. Math. Anal.* (2011).
[3] Grazia Lotti, Francesco Romani. "On the Asymptotic Complexity of Rectangular Matrix Multiplication". In: *Theoretical Computer Science* (1983).
[4] Nir Ailon, Bernard Chazelle. "Approximate Nearest Neighbors and the Fast Johnson-Lindenstrauss Transform". In: *STOC '06 Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing* (2006).
[5] Nir Ailon, Edo Liberty. "Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes". In: *Discrete & Computational Geometry* (2009).
[6] Noga Alon. "Problems and Results in Extremal Combinatorics-I". In: *Discrete Mathematics* (2003).
[7] Richard Baraniuk, Mark Davenport, Ronald DeVore, Michael Wakin. "A Simple Proof of the Restricted Isometry Property for Random Matrices". In: *Constructive Approximation* (2008).