Spring 2015

Johnson-Lindenstrauss projection of high dimensional data

Shuhong Gao
Clemson University

Fiona Knoll
Clemson University

Yue Mao
Clemson University

Tighter Bounds on Johnson Lindenstrauss Transforms

FIONA KNOLL (JOINT WORK WITH DR. SHUHONG GAO AND DR. YUE MAO)
DEPT. MATH. SCI., CLEMSON UNIVERSITY, CLEMSON, SC 29631

ABSTRACT: Johnson and Lindenstrauss (1984) proved that any finite set of data in a high dimensional space can be projected into a low dimensional space with the Euclidean metric information of the set being preserved within any desired accuracy. Such dimension reduction plays a critical role in many applications with massive data. There has been extensive effort in the literature on how to find explicit constructions of Johnson-Lindenstrauss projections. In this poster, we show how algebraic codes over finite fields can be used for fast Johnson-Lindenstrauss projections of data in high dimensional Euclidean spaces.

Johnson Lindenstrauss Transform

Problem

Given data in a high dimensional space, we want to project the data to a low dimensional space so that the pairwise distances are preserved with high probability.

Johnson Lindenstrauss Lemma

- Let n be any positive integer, $0 < \epsilon, \delta < \frac{1}{2}$ and $m = \mathcal{O}(\epsilon^{-2}\log\frac{1}{\delta})$. Then there exists a probabilistic distribution on $A \in \mathbb{R}^{m\times n}$ such that for any vector $x \in \mathbb{R}^n$,
$$\Pr[(1-\epsilon)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\epsilon)\|x\|_2^2] \ge 1-\delta.$$
  - If $\|x\| = 1$, we have
$$\Pr[|\ \|Ax\|_2^2 - 1\ | > \epsilon] < \delta.$$
  - A transformation matrix A with this property is called a Johnson Lindenstrauss Transformation.

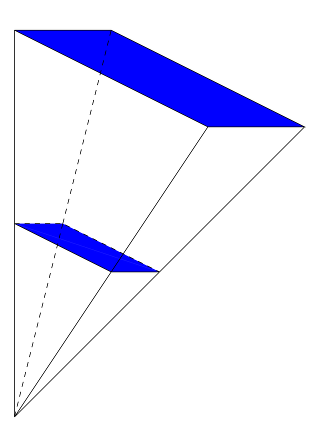
Motivation

Main Motivation

- A dimension reduction technique that preserves pairwise distances and norms.

Applications

- Speeds up algorithm processes:
  - Closest pair
  - Approximate nearest neighbor
  - Finding diameter and minimum spanning tree
- Reduces amount of storage required
  - One-pass streaming algorithms
  - Similarity measures (comparing text documents)



Comments on Parameters

Parameters δ and ϵ

- $\epsilon \in [0, 1/2)$: The desired accuracy
- $1-\delta \le 1$: The desired probability of success
  - Want $\delta \le \frac{1}{poly(n)}$ in order to compress $poly(n)$ points with a high probability of success.

Parameters n and m

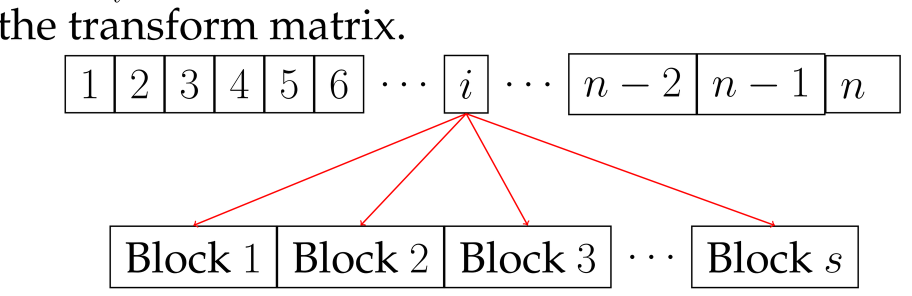
- n: Original dimension
- m: Desired dimension
  - Normally, $n >> m$ where $m \ge \mathcal{O}(\epsilon^{-2}\log\frac{1}{\delta})$

Previous Work: Initial Progress in Construction

- Johnson and Lindenstrauss (1984): Introduced and proved the JL lemma using a complicated transform matrix A with 3 conditions.
- Indyk and Motwani (1998): Dropped two of the conditions: orthogonality and normality of A
  - Construction: Choose independently and randomly $A_{i,j} \sim N(0, 1/n)$.
- Achlioptas (2003): Dropped the spherical symmetry condition of A
  - Relatively Sparse Construction: Choose independently and randomly
$$A_{i,j} = \begin{cases} (n/3)^{-1/2} & p = 1/6 \\ -(n/3)^{-1/2} & p = 1/6 \\ 0 & p = 2/3 \end{cases} \quad \text{with } p \text{ being the probability.}$$

Previous Work: Progress in Faster Computations

- Ailon and Chazelle (2009): Used Fourier in their Fast Johnson Lindenstrauss Transform
- Ailon and Liberty (2009): Used Rademacher and BCH Codes
- Kane and Nelson (2014): Used hash functions
  - Block Construction: Each coordinate of x is hashed to s coordinates where each of these s coordinates lies in one of the s blocks containing m/s entries.
    - Hashing determined by a code C:
      - $C = \{C_1, \ldots, C_n\} \subset [k/s]^s$ with relative distance $1 - O(s/k)$
      - $h : [d] \times [s] \to [k/s]$ can be determined by $(C_i)_j = h(i, j)$
      - The codeword C_i determines the location of the s nonzero elements in the i^{th} column of the transform matrix.



Previous Work: Progress Towards a Tighter Bound

Asymptotic Bounds

Kane's and Nelson's work was one of the first to give attention to the tightness of the success probability of the construction.

- Graph and Block Construction: For sparsity $s = \Omega\left(\epsilon^{-1}\log(1/\delta)\right)$ of matrix A, one may achieve distortion of $1 \pm \epsilon$ with success probability $1-\delta$.

Specific Bounds

- Kane and Nelson: Provided an upper bound for the block construction:
$$\Pr[|\ \|Ax\|_2^2 - 1\ | > \epsilon] \le \epsilon^{-\ell}\left(64\max\left\{\frac{\sqrt{l(s-d)}}{s}, \frac{l}{s}\right\}\right)^{l},$$
where d is the minimum distance of the code, s is the code length, and l a positive even integer.

Our Results: Tighter Bound

Let s be the length of the code used in the block construction and d the minimum distance.

- Tighter Constraint on Kane's and Nelson's Block Construction:
$$P[|\|Ax\|_2^2 - 1| > \epsilon] < \epsilon^{-\ell}\cdot\left(C_\ell\frac{\sqrt{\ell(s-d)}}{s}\right)^{\ell},$$
where $C_\ell \le 64$ for $\ell \le 7564$; more specifically $C_{32} = 4.21$, $C_{64} = 5.92$, and $C_{128} = 8.35$.

Explicit Construction Using AG Codes

AG Code:

(s, k, d)-AG Code over $\mathbb{F}_{q^2}$ from Garcia Stichtenoth Tower (GS tower):

- Code Length: $s = q^u(q^2 - q)$ for some integer $u \ge 1$ and prime power $q \ge 2$
- Dimension: $k < s$, an integer
- Minimum Distance: $d = s - k - g$ where $g = (q^{\lfloor\frac{u+1}{2}\rfloor} - 1)(q^{\lceil\frac{u+1}{2}\rceil} - 1)$ is the genus of the curve.

Bound in Terms of AG Code:

An (s, k, d)-AG Code over $\mathbb{F}_{q^2}$ from GS tower gives
$$\frac{s^2}{s-d} = \frac{(q^{u+2} - q^u)^2}{k + (q^{\lfloor\frac{u+1}{2}\rfloor} - 1)(q^{\lceil\frac{u+1}{2}\rceil} - 1)} \ge 4\epsilon^{-2}\cdot[(2\ell - 1)!!]^{\frac{1}{\ell}}.$$
Hence, $P[|\|Ax\|_2^2 - 1| > \epsilon] < 2^{-\ell}\left((2\ell - 1)!!\right)^{1/2}$.

Parameters

q	u	k	d	g	$s = q^u(q^2-q)$	$m = s\cdot q^2$	$n = (q^2)^k$	ϵ	$\delta = 0.5^\ell$
2	7	15	16	225	256	1024	1.07×10^{09}	0.42	0.5^{16}
2	10	17	78	1953	2048	8192	1.72×10^{10}	0.21	0.5^{32}
2	10	17	78	1953	2048	8192	1.72×10^{10}	0.30	0.5^{64}
4	2	8	139	45	192	3072	4.29×10^{09}	0.37	0.5^{32}

Consider the third line. If $\delta = .5^{64}$, then we can preserve pairwise distance for $p = 2^{20}$ points with 14% accuracy and by a success probability of $1 - \delta' = 1 - \left(\frac{1}{2}\right)^{25}$.

[1] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. Contemporary Mathematics, 26:189-206, 1984.

[2] Daniel M. Kane and Jelani Nelson. Sparser johnson-lindenstrauss transforms. J. ACM, 61(1):4:1-4:23, January 2014.

[3] Dimitris Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. Journal of Computer and System Sciences , 66(4):671-687, 2003. Special Issue on {PODS} 2001.

[4] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98, pages 604-613, New York, NY, USA, 1998. ACM.

[5] Nir Ailon and Bernard Chazelle. The fast johnson-lindenstrauss transform and approximate nearest neighbors. SIAM J. Comput, pages 302-322, 2009.

[6] Nir Ailon and Edo Liberty. Fast dimension reduction using rademacher series on dual bch codes. Technical report, 2007.

email: fknoll@clemson.edu