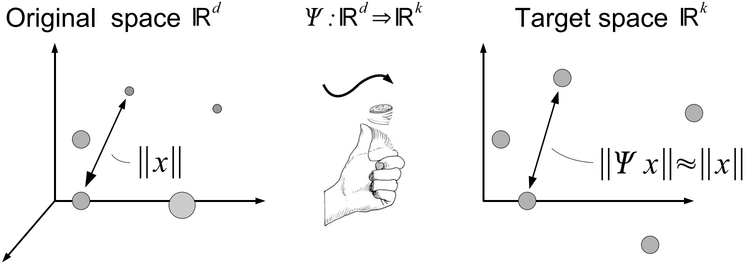
# Fast Dimension Reduction

Nir Ailon[1] Edo Liberty[2]

[1] Google Research
[2] Yale University

# Introduction




## Lemma (Johnson, Lindenstrauss (1984))

*A random projection $\Psi$ preserves all $\binom{n}{2}$ distances up to distortion $\varepsilon$ with constant probability if:*

$$k = \Omega\left(\frac{\log(n)}{\varepsilon^2}\right)$$

# Applications

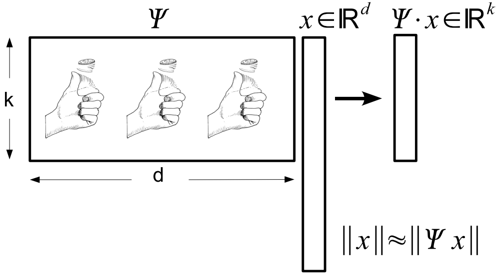
This idea is extremely useful in

- Approximate nearest neighbors searches
- Linear Embedding / Dimensionality reduction
- Matrix rank-k approximation
- $\ell_p$ regression
- Compressed sensing

and the list continues...

# JL Property definition

$\Pr[\textit{distortion}] \leq 1/n^2 \rightarrow$ embedding an $n$ point metric.




## Definition

A distribution $\mathcal{D}$ over $k \times d$ matrices exhibits the JL Property if for $\|x\|_2 = 1$ and $0 < \varepsilon < 1/2$

$$\Pr_{\Psi \sim \mathcal{D}}[|\|\Psi x\|_2 - 1| > \varepsilon] \leq c_1 e^{-c_2 k \varepsilon^2}$$

$k = c \log(n)/\varepsilon^2 \longrightarrow \Pr[\textit{distortion}] \leq 1/n^2$.

# Unstructured constructions

Other JL distributions for $\Psi(i,j)$ being i.i.d random variables:

| | | |
|---|---|---|
| Frankl and Maehara<br>Indyk and Motwani<br>DasGupta and Gupta | 1987<br>1998<br>1999 | $\Psi(i,j) \sim N(0,1)$ |
| Achlioptas | 2003 | $\Psi(i,j) \in \{0,-1,1\}$ |
| Matousek | 2006 | $\Psi(i,j)$ Symmetrically sub-gaussian distributed. |

However, these matrices are

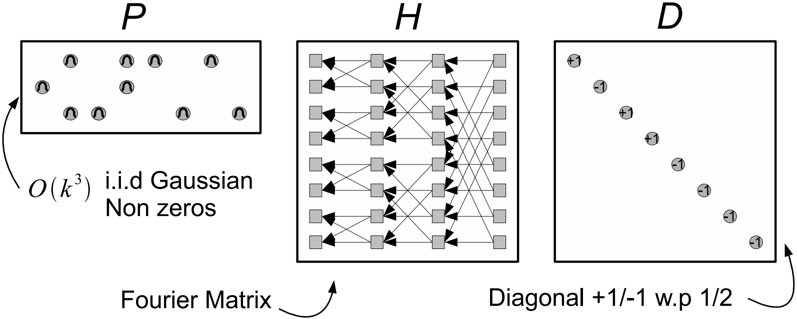
- **Slow to apply:** $O(kd)$
- **Heavy to store:** $O(kd)$

# Fast Johnson Lindenstrauss Transform

The first JL distribution for efficiently applicable matrices.

## Lemma (Ailon, Chazelle (2006))

*The distribution* $\Psi = PHD$ *exhibits the JL property*




- $O(d\log(d) + k^3)$ **to apply.**
- $O(d + k^3\log(d))$ **to store.**

# Our motivation

FJLT algorithm running time is $O(d\log(d) + k^3)$.

1. When the target dimension is very **small**

$$k = O(\log(d))$$

   The running time is $\Omega(kd)$.

2. When the target dimension is **large**

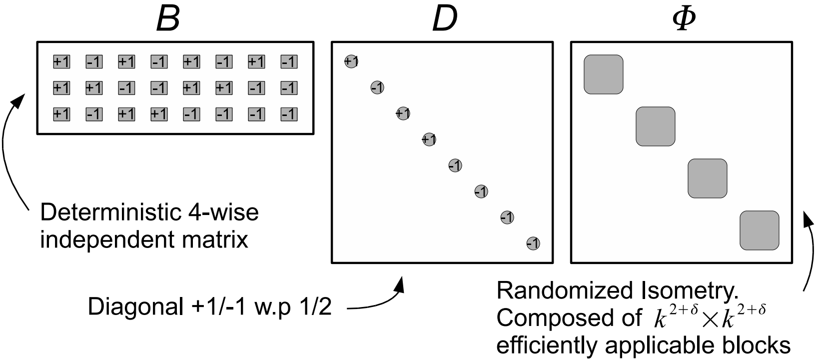
$$k = \Omega((d\log(d))^{1/3})$$

   The running time is dominated by $k^3$.

# This work, Faster Johnson Lindenstrauss Transform

## Lemma (Ailon, Liberty (2008))

*The distribution* $\Psi = BD\Phi$ *exhibits the JL property*



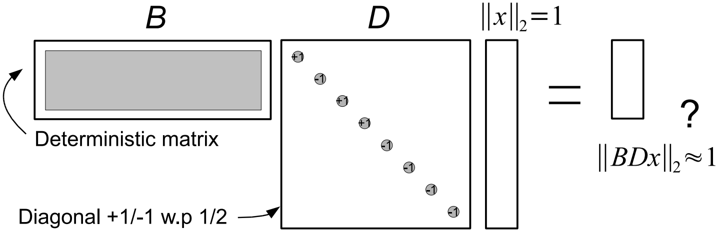

*for* $k = O(d^{1/2-\delta})$ *and* $\delta > 0$

- $O(d\log(k))$ **to apply**
- $O(d)$ **to store.**

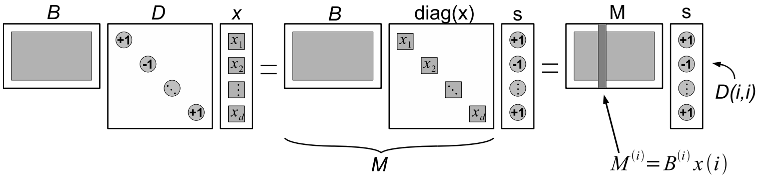
# Performance summary

| | Slow $- - - - -- \rightarrow$ Fast | | |
|---|---|---|---|
| $k$ in $o(\log d)$ | FJLT | JL | This work |
| $k$ in $\omega(\log d)$ and $o(\text{poly}(d))$ | JL | FJLT | This work |
| $k$ in $\Omega(\text{poly}(d))$ and $o((d \log(d)^{1/3})$ | JL | | This work, FJLT |
| $k$ in $\omega((d \log d)^{1/3})$ and $O(d^{1/2-\delta})$ | JL | FJLT | This work |

# Measure concentration




The randomness is only in *D*.




**How does $\|Ms\|_2$ concentrate around 1 ? ($\|x\|_2 = 1$)**

# Sufficient conditions

## Lemma (Talagrand)

*Denote:*

- $\mu$ *the median of* $\|Ms\|_2$,
- $\sigma = \|M\|_{2\to 2}$ *the spectral norm of M.*

$$\Pr\left[\,|\|Ms\|_2 - \mu| > t\right] \le 4e^{-t^2/8\sigma^2}$$

By substitution we get the JL property for *BD*:

$$\Pr(|\|BDx\|_2 - 1| > \varepsilon) \le c_1 e^{-c_2 k \varepsilon^2}$$

If:

- The columns of *B* are normalized to 1.
- $\sigma = \|M\|_{2\to 2} = O(k^{-1/2})$

# Breaking down $\|M\|_{2\to 2}$

Using the definition of the spectral norm and Cauchy-Schwartz:

$$\|M\|_{2\to 2} \leq \|B^T\|_{2\to 4}\|x\|_4$$

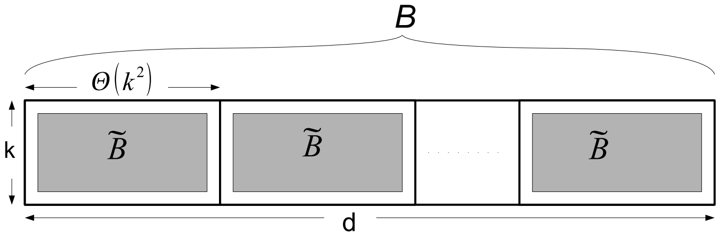
The $\ell_2 \to \ell_4$ operator norm is defined as such:

$$\|B^T\|_{2\to 4} \equiv \max_{\|z\|_2=1} \|B^T z\|_4$$

We seek a matrix $B$

- with a low $\ell_2 \to \ell_4$ operator norm, $\|B^T\|_{2\to 4}$.
- which is fast to apply.

# Our choice for $B$




**Lemma**

*$\widetilde{B}$ is obtained from $k$ rows of a $\Theta(k^2) \times \Theta(k^2)$ Hadamard matrix.*
*$\longrightarrow \quad B$ can be applied in $O(d \log(k))$ operations.*

**Lemma**

*Let $\widetilde{B}$ be a 4-wise independent matrix.*
*(The columns of $\widetilde{B}$ are the sample space)*
*$\longrightarrow \quad \|B^T\|_{2\to4} = O(d^{1/4}k^{-1/2})$.*

# Controlling $\|x\|_4$

Since $\|B^T\|_{2\to 4} = O(d^{1/4}k^{-1/2})$ our goal:

$$\|B^T\|_{2\to 4}\|x\|_4 = O(k^{-1/2})$$

is achieved if

$$\|x\|_4 = O(d^{-1/4})$$

**Problem:** $\|x\|_4$ might be as high as 1!
**Solution:** Preprocess $x$ using a random Isometry $\Phi$ such that:

- $\|\Phi x\|_4 = O(d^{-1/4})$ w.h.p
- $\Phi$ can be applied in $O(d\log(k))$ operations
- $\Phi$ requires $O(d)$ bits to store

## Lemma

*Such matrix distributions exist for* $k = O(d^{1/2-\delta})$.

# Recap and summary

Any matrix distribution $BD\Phi$ exhibits the JL property if

- The columns of $B$ are normalized to 1.
- $\Phi$ is an isometry.
- $\|B^T\|_{2\to 4}\|\Phi x\|_4 \leq O(k^{-1/2})$ w.h.p.

**Our result:** For $k = O(d^{1/2-\delta})$ there exist $B$ and $\Phi$ such that

- $BD\Phi$ requires $O(d\log(k))$ operations to apply
- $BD\Phi$ requires $O(d)$ bits to store

# Future work and open questions

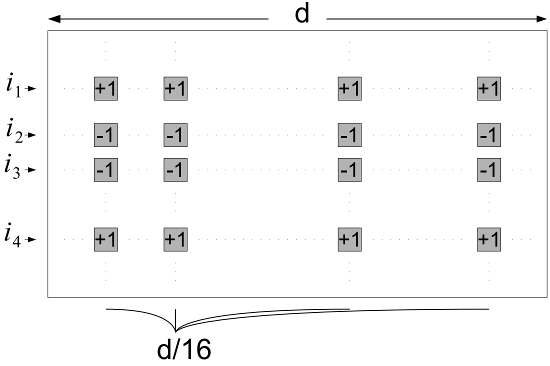
| | Slow $- - - - - - - \rightarrow$ Fast | | | $O(d)$ |
|---|---|---|---|---|
| $k$ in $o(\log d)$ | FJLT | JL | This work | |
| $k$ in $\omega(\log d)$ and $o(\text{poly}(d))$ | JL | FJLT | This work | |
| $k$ in $\Omega(\text{poly}(d))$ and $o((d\log(d)^{1/3})$ | JL | | This work, FJLT | |
| $k$ in $\omega((d\log d)^{1/3})$ and $O(d^{1/2-\delta})$ | JL | FJLT | This work | |
| $k$ in $\omega(d^{1/2-\delta})$ and $\leq d$ | JL, FJLT | | | |

# Thank you

# 4-wise independence

## Definition

$\widetilde{B}$ is 4-*wise independent* if:
There are $d/16$ columns $j$ for which $(\widetilde{B}^{(j)}_{i_1}, \widetilde{B}^{(j)}_{i_2}, \widetilde{B}^{(j)}_{i_3}, \widetilde{B}^{(j)}_{i_4}) = (b_1, b_2, b_3, b_4)$ for any choice of $1 \le i_1 < i_2 < i_3 < i_4 \le k$ and $(b_1, b_2, b_3, b_4) \in \{+1, -1\}^4$.



# Bounding $\|B^T\|_{2\to4}$

## Lemma

*Given a $k \times d$ 4-wise independent code matrix $B$,*

$$\|B^T\|_{2\to4} = O(d^{1/4}k^{-1/2}) \tag{1}$$

## Proof.

For $y \in \ell_2^k$, $\|y\| = 1$,

$$\begin{aligned}
\|y^T B\|_4^4 &= dE_{j\in[d]}[(y^T B(j))^4] \\
&= dk^{-2} \sum_{i_1,i_2,i_3,i_4=1}^{k} E_{b_1,b_2,b_3,b_4}[y_{i_1} y_{i_2} y_{i_3} y_{i_4} b_1 b_2 b_3 b_4] \qquad (2) \\
&= dk^{-2}(3\|y\|_2^4 - 2\|y\|_4^4) \leq 3dk^{-2}
\end{aligned}$$

$\square$

# Trimming the Hadamard transform

Applying $k$ rows from a Hadamard matrix, $H$, is like applying $PH$ where $P$ contains only $k$ entrees which are 1 and the rest are zero.

$$\begin{aligned} PH_d z &= \begin{pmatrix} P_1 & P_2 \end{pmatrix} \begin{pmatrix} H_{d/2} & H_{d/2} \\ H_{d/2} & -H_{d/2} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= P_1 H_{d/2}(z_1 + z_2) + P_2 H_{d/2}(z_1 - z_2) \end{aligned}$$

Which gives the relation $T(d,k) = T(d/2, k_1) + T(d/2, k_2) + d$.

Since $T(d,1) = d$, by induction

$$T(d,k) \leq 2d\log(k+1) = O(d\log(k)) \quad (3)$$