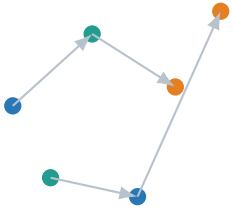
Lecture 6

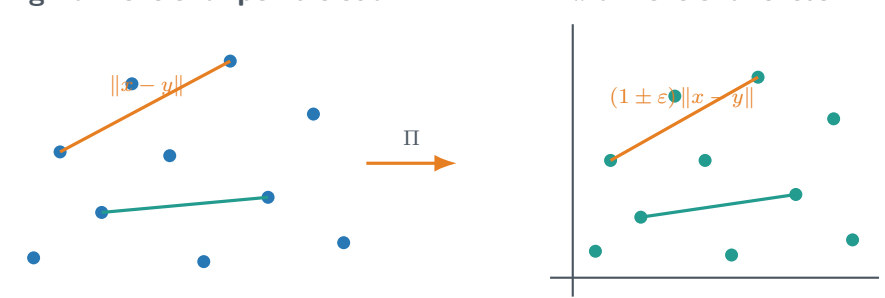
Johnson–Lindenstrauss Dimensionality Reduction

Concentration, dense random projections, and Fast JL



Geometry in fewer coordinates — without losing the distances that matter

The geometric promise



The surprise

For a finite set, the target dimension depends logarithmically on the number of points — and not at all on the ambient dimension n.

Johnson–Lindenstrauss lemma

Theorem

Let $S \subset \mathbb{R}^n$ be finite and let $0 < \varepsilon < 1$. There exists a linear map

$$\Pi : \mathbb{R}^n \longrightarrow \mathbb{R}^d, \qquad d = O\left(\frac{\log |S|}{\varepsilon^2}\right),$$

such that for every $x, y \in S$,

$$(1-\varepsilon)\,\|x-y\|_2 \le \|\Pi x - \Pi y\|_2 \le (1+\varepsilon)\,\|x-y\|_2 .$$

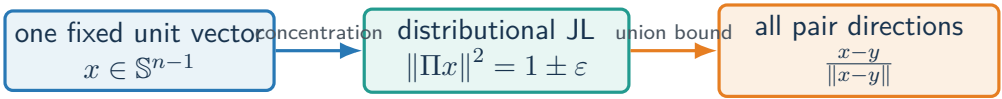
Dimension

$m = |S|$ points need only $O(\varepsilon^{-2} \log m)$ coordinates.

Linearity

One matrix preserves all $\binom{m}{2}$ pairwise distances simultaneously.

The proof has two reductions



1. Build a random matrix that preserves the norm of *one fixed* vector with failure probability $e^{-\Omega(\varepsilon^2 d)}$.
2. Apply the guarantee to at most $|S|^2$ normalized difference vectors.

What does the work?

The logarithm in JL is simply the price of a union bound over all pairs. The ε^{-2} comes from Gaussian-scale concentration.

Distributional JL

Definition

A distribution $\mathcal{P}$ over matrices $\Pi \in \mathbb{R}^{d \times n}$ is an (ε, δ)-JL distribution if, for every *fixed* $x \in \mathbb{S}^{n-1}$,

$$\mathbb{P}_{\Pi \sim \mathcal{P}} \left[\left| \|\Pi x\|_2^2 - 1 \right| > \varepsilon \right] \leq \delta.$$

Quantifiers

For every fixed x, most matrices work for x.

Not claimed

One short matrix cannot preserve every vector in $\mathbb{R}^n$ when $d < n$.

We need the failure probability $\delta = e^{-\Omega(\varepsilon^2 d)}$.

From one vector to all pairwise distances

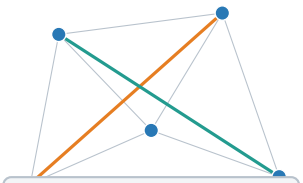
Let $m = |S|$ and define the normalized pair directions

$$S' = \left\{ \frac{x - y}{\|x - y\|_2} : x, y \in S,\ x \neq y \right\} \subset \mathbb{S}^{n-1}.$$

Since $|S'| \leq m^2$, an (ε, δ)-JL distribution gives

$$\begin{aligned} \mathbb{P}[\exists s \in S' : \left| \|\Pi s\|^2 - 1 \right| > \varepsilon] &\leq \sum_{s \in S'} \delta \\ &\leq m^2 \delta. \end{aligned}$$

Choose $\delta \leq 1/(2m^2)$: with probability at least $1/2$, every pair direction is preserved.



The union-bound proof, completed

On the good event, for each distinct $x, y \in S$,

$$\begin{aligned}\|\Pi x - \Pi y\|_2^2 &= \|\Pi(x-y)\|_2^2 \\ &= \|x-y\|_2^2 \left\| \Pi\left(\frac{x-y}{\|x-y\|_2}\right)\right\|_2^2 \\ &= (1 \pm \varepsilon)\,\|x-y\|_2^2 .\end{aligned}$$

For $0 < \varepsilon < 1$,

$$1 - \varepsilon \le \sqrt{1-\varepsilon} \quad \text{and} \quad \sqrt{1+\varepsilon} \le 1 + \varepsilon,$$

so the same event gives norm distortion at most $1 \pm \varepsilon$.

Dimension from the failure exponent

If $\delta \le 2e^{-c\varepsilon^2 d}$, then $m^2\delta \le \frac{1}{2}$ once

$$d \ge C\varepsilon^{-2} \log m.$$

Thus it remains only to construct a distributional JL map.

A Gaussian candidate

Let $G \in \mathbb{R}^{d \times n}$ have independent $N(0,1)$ entries and set

$$\boxed{\Pi = \frac{1}{\sqrt{d}} G.}$$

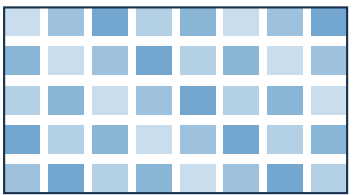
Fix a unit vector x. If g_i^{T} is row i of G, then

$$\|\Pi x\|_2^2 = \frac{1}{d} \sum_{i=1}^{d} \langle g_i, x \rangle^2 .$$

Rotational invariance gives

$$\langle g_i, x \rangle \sim N(0, \|x\|^2) = N(0,1),$$

independently across rows.

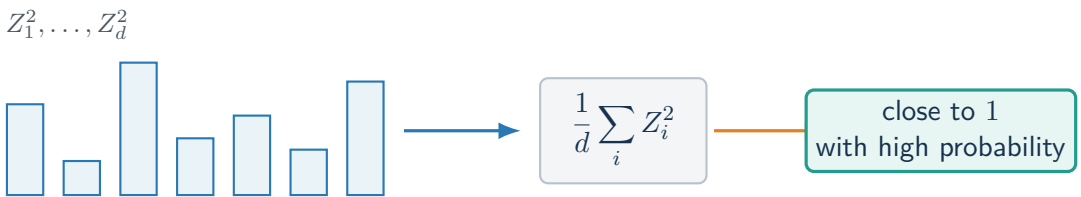


d independent random probes of x

The problem becomes a χ^2 concentration question

For every fixed $x \in \mathbb{S}^{n-1}$,

$$\|\Pi x\|_2^2 \overset{d}{=} \frac{1}{d}\sum_{i=1}^{d} Z_i^2, \qquad Z_1, \ldots, Z_d \overset{\text{i.i.d.}}{\sim} N(0,1).$$



Why last lecture's theorem is not enough

The variables Z_i are subgaussian, but $Z_i^2 - 1$ are not: $\mathbb{P}[Z_i^2 > t] = e^{-\Theta(t)}$ rather than $e^{-\Omega(t^2)}$. We need a slightly broader concentration class.

Concentration beyond Gaussian tails

Moment bounds, MGF bounds, and the subexponential class

Recall: the subgaussian MGF method

Definition

A centered random variable X is σ-subgaussian if

$$\mathbb{E}e^{\lambda X} \le \exp\left(\frac{\sigma^2\lambda^2}{2}\right) \qquad \text{for every } \lambda \in \mathbb{R}.$$

For $t > 0$ and $\lambda > 0$, Markov's inequality gives

$$\mathbb{P}[X \ge t] = \mathbb{P}[e^{\lambda X} \ge e^{\lambda t}] \le \exp\left(\frac{\sigma^2\lambda^2}{2} - \lambda t\right).$$

Optimizing at $\lambda = t/\sigma^2$, and applying the same argument to $-X$,

$$\boxed{\mathbb{P}[|X| \ge t] \le 2e^{-t^2/(2\sigma^2)}.}$$

Independent subgaussians add in quadrature

Suppose $X_1, \dots, X_N$ are independent and X_i is σ_i-subgaussian. Then

$$\begin{aligned}\mathbb{E}\exp\left(\lambda \sum_i X_i\right) &= \prod_i \mathbb{E} e^{\lambda X_i} \\ &\le \prod_i e^{\sigma_i^2 \lambda^2/2} = e^{(\sum_i \sigma_i^2)\lambda^2/2}.\end{aligned}$$

Therefore

$$\sum_i X_i \quad \text{is} \quad \sqrt{\sum_i \sigma_i^2}\text{-subgaussian.}$$

Hoeffding as a corollary

If $\mathbb{E}X_i = 0$, the variables are independent, and $|X_i| \le b_i$, then with $\sigma^2 = \sum_i b_i^2$,

$$\mathbb{P}\left[\left|\sum_i X_i\right| > t\sigma\right] \le 2e^{-ct^2}.$$

L_p norms: measuring every moment at once

Definition

For $p \geq 1$ and a random variable X,

$$\|X\|_p := \left(\mathbb{E}|X|^p\right)^{1/p}.$$

These are norms: in particular, $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$ and $\|X\|_p \leq \|X\|_q$ for $p \leq q$.

Apply Markov to $|X|^p$:

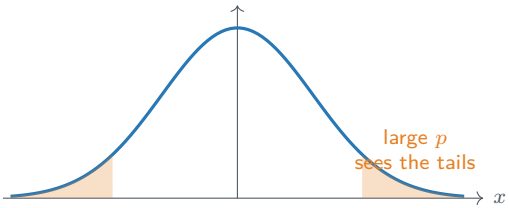
$$\boxed{\mathbb{P}[|X| > t] = \mathbb{P}[|X|^p > t^p] \leq \frac{\mathbb{E}|X|^p}{t^p} = \left(\frac{\|X\|_p}{t}\right)^p.}$$

The new freedom is that we may choose the moment p to match the deviation t.

Gaussian moments grow like $\sqrt{p}$

For $g \sim N(0,1)$,

$$\mathbb{E}|g|^p = \frac{2^{p/2}\Gamma((p+1)/2)}{\sqrt{\pi}}, \qquad c\sqrt{p} \le \|g\|_p \le C\sqrt{p}.$$



Interpretation

A variable has Gaussian-scale tails precisely when its moments grow no faster than $\sqrt{p}$ (up to universal constants).

Moment growth implies subgaussian tails

Assume that, for every $p \geq 1$,

$$\|X\|_p \leq C\sigma\sqrt{p}.$$

Moment Markov gives, for every choice of p,

$$\mathbb{P}[|X| > t] \leq \left(\frac{C\sigma\sqrt{p}}{t}\right)^p.$$

When $t \geq 2C\sigma$, choose

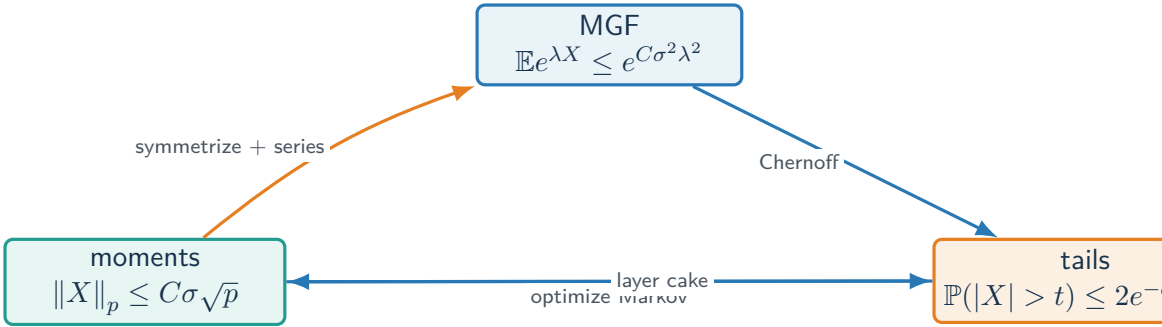
$$p = \left\lfloor \frac{t^2}{4C^2\sigma^2} \right\rfloor, \qquad \frac{C\sigma\sqrt{p}}{t} \leq \frac{1}{2}.$$

Thus

$$\mathbb{P}[|X| > t] \leq 2^{-p} \leq 2\exp\left(-c\frac{t^2}{\sigma^2}\right).$$

For smaller t the bound is made valid by adjusting the constant.

Three equivalent faces of subgaussianity



Theorem

For a centered random variable, these conditions are equivalent up to universal constant changes in the parameter σ.

Gaussian tails imply Gaussian moment growth

Suppose

$$\mathbb{P}[|X| > t] \le 2e^{-t^2/(2\sigma^2)}.$$

The layer-cake identity says

$$\mathbb{E}|X|^p = p \int_0^\infty t^{p-1} \mathbb{P}[|X| > t]\, dt.$$

Therefore, after substituting $u = t^2/(2\sigma^2)$,

$$\begin{aligned} \mathbb{E}|X|^p &\le 2p \int_0^\infty t^{p-1} e^{-t^2/(2\sigma^2)}\, dt \\ &= p(2\sigma^2)^{p/2}\Gamma(p/2) \le (C\sigma\sqrt{p})^p. \end{aligned}$$

Hence $\|X\|_p \le C\sigma\sqrt{p}$ for every $p \ge 1$.

Moments imply an MGF bound: symmetrize first

Assume $\mathbb{E}X = 0$ and $\|X\|_p \le C\sigma\sqrt{p}$ for every p. Let X' be an independent copy of X. Jensen's inequality gives

$$e^{\lambda X} = e^{\lambda(X - \mathbb{E}X')} \le \mathbb{E}_{X'} e^{\lambda(X - X')},$$
$$\therefore \qquad \mathbb{E}e^{\lambda X} \le \mathbb{E}e^{\lambda(X - X')}.$$

Set $Y = X - X'$. Then

$$Y \overset{d}{=} -Y, \qquad \|Y\|_p \le \|X\|_p + \|X'\|_p \le 2C\sigma\sqrt{p}.$$

Why symmetrization helps

All odd moments of Y vanish, so the exponential series has only nonnegative, even-moment terms.

Moments imply an MGF bound: compare with a Gaussian

Choose $g \sim N(0, C_0^2\sigma^2)$ with C_0 large enough that

$$\mathbb{E}|Y|^{2k} \le \mathbb{E}|g|^{2k} \qquad \text{for every } k \ge 0.$$

Because Y is symmetric,

$$\begin{aligned}\mathbb{E}e^{\lambda Y} &= \sum_{k=0}^{\infty} \frac{\lambda^{2k}\mathbb{E}Y^{2k}}{(2k)!} \\ &\le \sum_{k=0}^{\infty} \frac{\lambda^{2k}\mathbb{E}g^{2k}}{(2k)!} = \mathbb{E}e^{\lambda g} = e^{C_0^2\sigma^2\lambda^2/2}.\end{aligned}$$

Together with symmetrization,

$$\mathbb{E}e^{\lambda X} \le e^{C_0^2\sigma^2\lambda^2/2}.$$

This closes the equivalence between MGF, moment, and tail formulations.

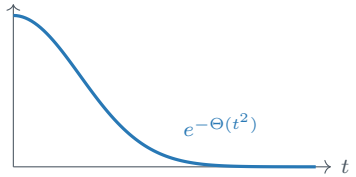
Squaring changes the tail scale

Gaussian variable

If $g \sim N(0,1)$, then

$$\mathbb{P}[|g| > t] = e^{-\Theta(t^2)}.$$

The exponent is quadratic in t.

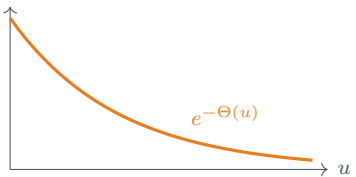


Squared Gaussian

For $u > 0$,

$$\mathbb{P}[g^2 > u] = \mathbb{P}[|g| > \sqrt{u}] = e^{-\Theta(u)}.$$

The exponent is only linear in u.



Subexponential random variables

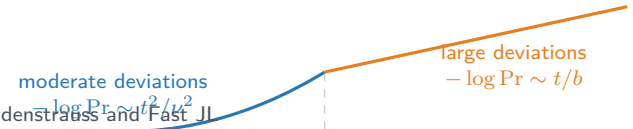
Definition

A centered random variable X is (ν, b)-subexponential if

$$\mathbb{E}e^{\lambda X} \le \exp\left(\frac{\nu^2\lambda^2}{2}\right) \qquad \text{for every } |\lambda| < \frac{1}{b}.$$

The same quadratic MGF bound holds, but only in a neighborhood of the origin. This produces two tail regimes:

$$\boxed{\mathbb{P}[|X| > t] \le 2\exp\left[-c\min\left\{\frac{t^2}{\nu^2}, \frac{t}{b}\right\}\right].}$$



The two-regime tail bound from Chernoff

For any $0 < \lambda < 1/b$,

$$\mathbb{P}[X \geq t] \leq \exp\left(\frac{\nu^2\lambda^2}{2} - \lambda t\right).$$

There are two possible optimizers.

Moderate t

If $t \lesssim \nu^2/b$, take $\lambda = t/\nu^2$:

$$\mathbb{P}[X \geq t] \leq e^{-t^2/(2\nu^2)}.$$

Large t

The unconstrained optimizer is illegal.
Take $\lambda = 1/(2b)$:

$$\mathbb{P}[X \geq t] \leq e^{-ct/b}.$$

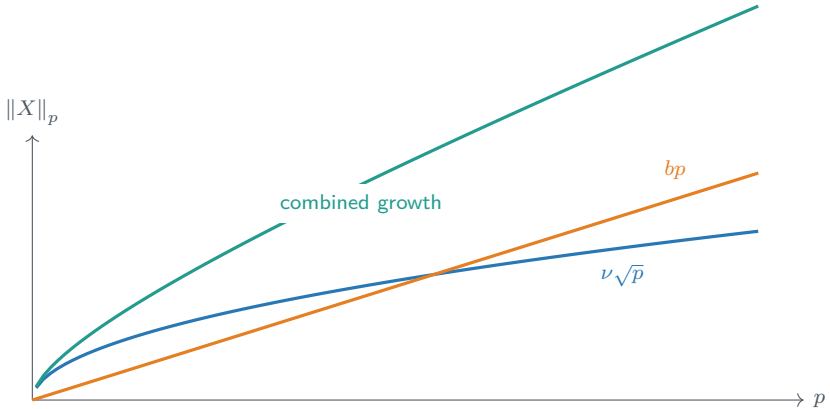
Applying the same bound to $-X$ gives

$$\mathbb{P}[|X| > t] \leq 2e^{-c\min\{t^2/\nu^2, t/b\}}.$$

Equivalent moment growth

Up to universal constants, the subexponential tail bound is equivalent to

$$\boxed{\|X\|_p \lesssim \nu\sqrt{p} + bp \qquad (p \geq 1).}$$



Independent subexponentials and Bernstein's inequality

If X_i are independent and (ν_i, b_i)-subexponential, then

$$\mathbb{E}e^{\lambda \sum_i X_i} = \prod_i \mathbb{E}e^{\lambda X_i} \le e^{(\sum_i \nu_i^2)\lambda^2/2} \quad \text{for } |\lambda| < \frac{1}{\max_i b_i}.$$

Thus

$$\sum_i X_i \text{ is } \left(\sqrt{\sum_i \nu_i^2}, \max_i b_i\right) \text{-subexponential.}$$

Bernstein tail

Writing $\nu^2 = \sum_i \nu_i^2$ and $b = \max_i b_i$,

$$\mathbb{P}\left[\left|\sum_i X_i\right| > t\right] \le 2\exp\left[-c\min\left\{\frac{t^2}{\nu^2}, \frac{t}{b}\right\}\right].$$

A familiar form: Bernstein for bounded variables

Let $X_1, \dots, X_N$ be independent and centered, with

$$|X_i| \le B, \qquad V := \sum_{i=1}^{N} \mathbb{E}X_i^2.$$

The boundedness controls the MGF for $|\lambda| \lesssim 1/B$, while the quadratic term is governed by V. Consequently,

$$\boxed{\mathbb{P}\left[\left|\sum_{i=1}^{N} X_i\right| > t\right] \le 2\exp\left(-\frac{t^2}{2(V + Bt/3)}\right).}$$

In the two natural ranges,

$$\mathbb{P}\left[\left|\sum_{i} X_i\right| > t\right] \le 2\exp\left[-c\min\left\{\frac{t^2}{V}, \frac{t}{B}\right\}\right].$$

Compared with Hoeffding

Bernstein uses the actual variance V; this refinement will be decisive in Fast JL.

A square of a subgaussian is subexponential

Lemma

If X is K-subgaussian, then $X^2 - \mathbb{E}X^2$ is subexponential with scale $O(K^2)$.

Indeed, subgaussian moment growth gives $\|X\|_q \lesssim K\sqrt{q}$. For every $p \geq 1$,

$$\begin{aligned}\left\|X^2 - \mathbb{E}X^2\right\|_p &\leq \left\|X^2\right\|_p + \mathbb{E}X^2 \\ &= \|X\|_{2p}^2 + \|X\|_2^2 \\ &\lesssim K^2 p + K^2 \lesssim K^2 p.\end{aligned}$$

This is exactly subexponential moment growth.

The variable we needed

For $Z \sim N(0,1)$, the centered square $Z^2 - 1$ is $(O(1), O(1))$-subexponential.

Dense random projections

Subgaussian rows give the optimal Johnson–Lindenstrauss dimension

Dense subgaussian JL theorem

Theorem

Let $G \in \mathbb{R}^{d \times n}$ have independent entries satisfying

$$\mathbb{E}G_{ij} = 0, \qquad \mathbb{E}G_{ij}^2 = 1, \qquad G_{ij} \text{ is } K\text{-subgaussian},$$

where $K = O(1)$. Then

$$\Pi := \frac{1}{\sqrt{d}} G$$

is an $\left(\varepsilon, 2e^{-c\varepsilon^2 d}\right)$-JL distribution for $0 < \varepsilon < 1$. The constant $c > 0$ depends only on K.

Gaussian map

$G_{ij} \sim N(0,1)$ independently.

Random-sign map

$G_{ij} \sim \mathrm{Unif}\{-1, +1\}$ independently.

One row: the expectation is exactly right

Fix $x \in \mathbb{S}^{n-1}$ and write g_i^{T} for row i of G. Let

$$Z_i := \langle g_i, x \rangle = \sum_{j=1}^{n} G_{ij} x_j.$$

Then

$$\begin{aligned} \mathbb{E} Z_i^2 &= \sum_{j,k} x_j x_k \mathbb{E}[G_{ij} G_{ik}] \\ &= \sum_{j} x_j^2 \mathbb{E} G_{ij}^2 + \sum_{j \neq k} x_j x_k \underbrace{\mathbb{E} G_{ij}}_{0} \underbrace{\mathbb{E} G_{ik}}_{0} \\ &= \sum_{j} x_j^2 = 1. \end{aligned}$$

Consequently,

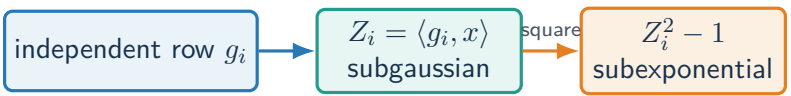
$$\mathbb{E} \, \|\Pi x\|_2^2 = \frac{1}{d} \sum_{i=1}^{d} \mathbb{E} Z_i^2 = 1.$$

One row: the measurement is subgaussian

Independence of the entries in row i gives

$$\begin{aligned}\mathbb{E}e^{\lambda Z_i} &= \prod_{j=1}^{n} \mathbb{E}e^{\lambda x_j G_{ij}} \\ &\le \prod_{j=1}^{n} \exp\left(\frac{K^2\lambda^2 x_j^2}{2}\right) \\ &= \exp\left(\frac{K^2\lambda^2 \|x\|_2^2}{2}\right) = e^{K^2\lambda^2/2}.\end{aligned}$$

So each Z_i is K-subgaussian, has variance 1, and the Z_i are independent across rows.



Average the independent squared measurements

We have the exact decomposition

$$\|\Pi x\|_2^2 - 1 = \frac{1}{d}\sum_{i=1}^{d}(Z_i^2 - 1).$$

Each $Y_i := Z_i^2 - 1$ is centered and $(O(1), O(1))$-subexponential. Hence $\sum_i Y_i$ is $(O(\sqrt{d}), O(1))$-subexponential, and Bernstein gives

$$\begin{aligned}\mathbb{P}\Big[\Big|\|\Pi x\|_2^2 - 1\Big| > \varepsilon\Big] &= \mathbb{P}\left[\left|\sum_{i=1}^{d} Y_i\right| > \varepsilon d\right] \\ &\le 2\exp\left[-c\min\left\{\frac{\varepsilon^2 d^2}{d}, \varepsilon d\right\}\right] \\ &\le 2e^{-c\varepsilon^2 d}, \qquad 0 < \varepsilon < 1.\end{aligned}$$

Conclusion

$G/\sqrt{d}$ is an $(\varepsilon, 2e^{-c\varepsilon^2 d})$-JL distribution.

The JL lemma is now immediate



Choose d so that

$$|S|^2 \cdot 2e^{-c\varepsilon^2 d} \leq \frac{1}{2}.$$

Equivalently,

$$\boxed{d \geq C\frac{\log |S|}{\varepsilon^2}.}$$

With probability at least $1/2$, the sampled matrix preserves every pairwise distance in S. In particular, a suitable matrix exists.

Application: oblivious subspace embedding

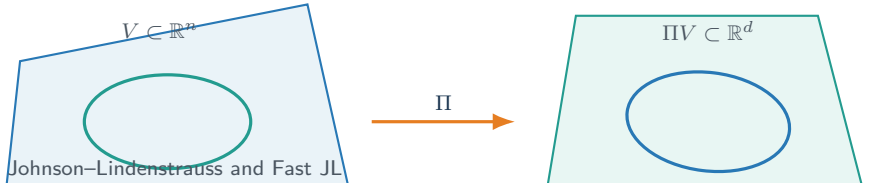
Theorem

Let $V \subset \mathbb{R}^n$ be a fixed k-dimensional subspace. A normalized dense subgaussian matrix $\Pi = G/\sqrt{d} \in \mathbb{R}^{d \times n}$ as above, with

$$d \geq C \frac{k + \log(1/\delta)}{\varepsilon^2}$$

satisfies, with probability at least $1 - \delta$,

$$(1 - \varepsilon) \|x\|_2^2 \leq \|\Pi x\|_2^2 \leq (1 + \varepsilon) \|x\|_2^2 \qquad \text{for every } x \in V.$$



A net reduces a whole subspace to finitely many vectors

Let $U \in \mathbb{R}^{n \times k}$ have orthonormal columns spanning V, and define

$$M := U^{\mathsf{T}} \Pi^{\mathsf{T}} \Pi U - I_k.$$

For $z \in \mathbb{S}^{k-1}$,

$$z^{\mathsf{T}} M z = \|\Pi U z\|_2^2 - 1.$$

Take a $1/4$-net $\mathcal{N} \subset \mathbb{S}^{k-1}$ with $|\mathcal{N}| \le 9^k$. Pointwise JL and a union bound imply

$$|z^{\mathsf{T}} M z| \le \varepsilon/2 \quad \text{for all } z \in \mathcal{N}$$

except with probability

$$9^k \cdot 2e^{-c\varepsilon^2 d} \le \delta$$

when $d \ge C(k + \log(1/\delta))/\varepsilon^2$.

Lift the net guarantee to every vector

For a symmetric matrix M and an η-net $\mathcal{N} \subset \mathbb{S}^{k-1}$,

$$\|M\| \le \frac{1}{1-2\eta} \max_{z \in \mathcal{N}} |z^{\mathsf{T}} M z|.$$

Proof. Choose a unit u with $|u^{\mathsf{T}} M u| = \|M\|$ and $z \in \mathcal{N}$ with $\|u - z\| \le \eta$. Then

$$\begin{aligned} \|M\| &= |u^{\mathsf{T}} M u| \\ &\le |z^{\mathsf{T}} M z| + |(u-z)^{\mathsf{T}} M u| + |z^{\mathsf{T}} M (u-z)| \\ &\le \max_{w \in \mathcal{N}} |w^{\mathsf{T}} M w| + 2\eta \, \|M\| . \end{aligned}$$

For $\eta = 1/4$, the factor is 2. Therefore the net bound $\varepsilon/2$ gives $\|M\| \le \varepsilon$, which is exactly the subspace embedding guarantee.

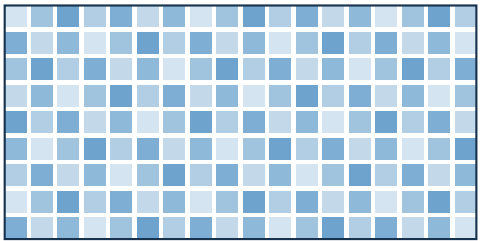
Fast Johnson–Lindenstrauss

Random signs flatten; a fast transform spreads; coordinate sampling compresses

Why look for a faster transform?

A dense JL matrix has dn nonzero entries, so computing Πx costs

$$\Theta(nd) \quad \text{arithmetic operations.}$$



dense $d \times n$ projection

Goal

Keep a distributional JL guarantee while replacing dense multiplication by structured, implicit operations.

$$O(n \log n + d) \quad \text{instead of} \quad O(nd).$$

The price in this elementary construction is a

The Fast JL transform

For clarity assume n is a power of two. Define

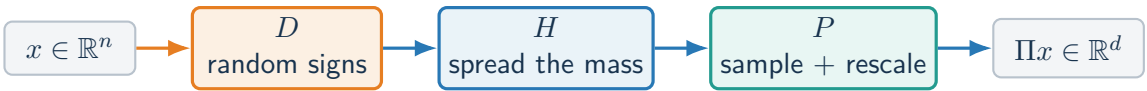
$$\boxed{\Pi = PHD \in \mathbb{R}^{d \times n}.}$$

- $D = \mathrm{diag}(\xi_1, \ldots, \xi_n)$, where $\xi_j \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\{\pm 1\}$.
- $H \in \mathbb{R}^{n \times n}$ is the normalized Walsh–Hadamard matrix: $H^\mathsf{T} H = I$ and $|H_{ij}| = 1/\sqrt{n}$.
- Draw $I_1, \ldots, I_d \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[n]$ and set

$$R = \begin{bmatrix} e_{I_1}^\mathsf{T} \\ \vdots \\ e_{I_d}^\mathsf{T} \end{bmatrix}, \qquad P = \sqrt{\frac{n}{d}}\, R.$$

The samples are independent of D. Sampling is with replacement, so repeated rows are allowed.

Fast JL in one pipeline



1. Destroy alignment

Random signs make cancellation available to every Hadamard coordinate.

2. Flatten

The orthogonal transform preserves ℓ_2 norm and spreads it across coordinates.

3. Subsample

Once no coordinate is heavy, a few random coordinates estimate the full norm.

Random signs and H flatten a spiky vector



Flattening lemma

For every fixed $x \in \mathbb{S}^{n-1}$ and every $\gamma > 0$,

$$\mathbb{P}_D\left[\|HDx\|_\infty > \frac{\gamma}{\sqrt{n}}\right] \leq 2ne^{-\gamma^2/2}.$$

Proof of the flattening lemma

Let $y = HDx$. For a fixed coordinate i,

$$y_i = \sum_{j=1}^{n} H_{ij}\xi_j x_j.$$

The summands are independent and centered. Their subgaussian variance proxies add:

$$\sum_{j=1}^{n} H_{ij}^2 x_j^2 = \frac{1}{n}\sum_{j=1}^{n} x_j^2 = \frac{1}{n}.$$

Hence y_i is $1/\sqrt{n}$-subgaussian and

$$\mathbb{P}\left[|y_i| > \frac{\gamma}{\sqrt{n}}\right] \leq 2e^{-\gamma^2/2}.$$

A union bound over $i = 1, \ldots, n$ gives

$$\mathbb{P}\left[\|y\|_\infty > \frac{\gamma}{\sqrt{n}}\right] \leq 2ne^{-\gamma^2/2}.$$

Also $\|y\|_2 = \|HDx\|_2 = \|x\|_2 = 1$ because H and D are orthogonal.

Sampling a flat vector

Condition on a vector $y = HDx$ satisfying

$$\|y\|_2 = 1, \qquad \|y\|_\infty \le \frac{\gamma}{\sqrt{n}}.$$



$I_1, \dots, I_d$ are independent uniform coordinates; repeats are allowed

Define $W_s := n y_{I_s}^2$. Then

$$\|\Pi x\|_2^2 = \frac{n}{d} \sum_{s=1}^{d} y_{I_s}^2 = \frac{1}{d} \sum_{s=1}^{d} W_s.$$

Moreover,

$$\mathbb{E} W_s = n \cdot \frac{1}{n} \sum_{j=1}^{n} y_j^2 = 1.$$

So the sampled norm is an unbiased estimator of the full norm.

Flatness gives a sharp variance bound

On the flattening event,

$$0 \le W_s = n y_{I_s}^2 \le \gamma^2.$$

More importantly,

$$\begin{aligned}
\mathbb{E} W_s^2 &= \frac{1}{n} \sum_{j=1}^{n} n^2 y_j^4 \\
&= n \sum_{j=1}^{n} y_j^4 \\
&\le n \, \|y\|_\infty^2 \sum_{j=1}^{n} y_j^2 \\
&\le \gamma^2.
\end{aligned}$$

Thus $\mathrm{Var}(W_s) \le \mathbb{E} W_s^2 \le \gamma^2$. For the chosen $\gamma \ge 1$, also $|W_s - 1| \le \gamma^2$. Applying Bernstein to $W_s - 1$, for $0 < \varepsilon \le 1$,

$$\mathbb{P}\left[\left| \frac{1}{d} \sum_{k=1}^{d} W_{s_k} - 1 \right| \ge \varepsilon \right] \le 2 \exp\left(-\frac{d\varepsilon^2}{2\gamma^2 + \frac{2}{3}\gamma^2 \varepsilon} \right)$$

Why Bernstein, rather than Hoeffding?

Both inequalities see $0 \le W_s \le \gamma^2$, but only Bernstein uses $\mathrm{Var}(W_s) \le \gamma^2$.

	information used	exponent for the average
Hoeffding	range γ^2	$-\Theta(d\varepsilon^2/\gamma^4)$
Bernstein	range γ^2 + variance γ^2	$-\Theta(d\varepsilon^2/\gamma^2)$

One logarithm matters

Flattening will give $\gamma^2 = \Theta(\log(n/\delta))$. Using the variance saves a full factor of $\log(n/\delta)$ in the sample count.

Choose the parameters

There are two bad events.

Flattening fails

Choose

$$\gamma^2 = 2 \log \frac{4n}{\delta}.$$

Then

$$2ne^{-\gamma^2/2} \leq \delta/2.$$

Sampling fails

It is enough that

$$2e^{-cd\varepsilon^2/\gamma^2} \leq \delta/2,$$

or

$$d \geq C\varepsilon^{-2}\gamma^2 \log \frac{4}{\delta}.$$

Therefore it suffices to take

$$d \geq C\varepsilon^{-2} \log \frac{4n}{\delta} \log \frac{4}{\delta} \leq C\varepsilon^{-2} \log^2 \frac{4n}{\delta}.$$

The two failure probabilities add to at most δ.

Fast JL theorem

Theorem

Let D, H, P be as above. For every fixed unit vector x and $0 < \varepsilon, \delta < 1/2$, the map $\Pi = PHD$ satisfies

$$\mathbb{P}\Big[\Big|\|\Pi x\|_2^2 - 1\Big| > \varepsilon\Big] \le \delta$$

provided

$$d \ge C\varepsilon^{-2}\log^2\left(\frac{4n}{\delta}\right).$$

Randomize

D prevents a fixed input from aligning with a row of H.

Flatten

H turns ℓ_2 mass into many small coordinates.

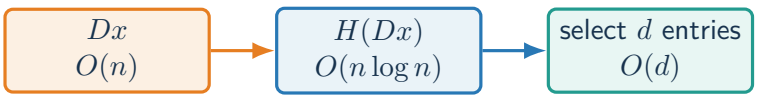
Sample

P estimates the norm from only d coordinates.

The “fast” in Fast JL

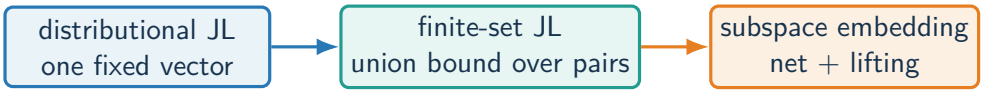
The matrix is never formed explicitly.

transform	target dimension	time for $x \mapsto \Pi x$
dense subgaussian JL	$O(\varepsilon^{-2}\log(1/\delta))$	$O(nd)$
elementary Fast JL	$O(\varepsilon^{-2}\log^2(n/\delta))$	$O(n\log n + d)$



For a finite set of m points, take $\delta \asymp m^{-2}$ and union-bound as before. This elementary Fast JL analysis trades an extra logarithm in dimension for much faster multiplication.

Takeaways



- MGF, moments, and tails are interchangeable languages for concentration.
- Squaring a subgaussian produces a subexponential variable; Bernstein then controls averages of squared measurements.
- Dense subgaussian projections achieve $d = O(\varepsilon^{-2} \log |S|)$.
- Fast JL uses the structural recipe

$$\boxed{\text{random signs} \longrightarrow \text{fast flattening} \longrightarrow \text{coordinate sampling.}}$$