Computer Science 2: Advanced Algorithms Fall 2026

Lecture 7 — September 14, 2026

Prof. Jarosław Błasiok *Scribe: Yang Liu, Fangqing Yuan*

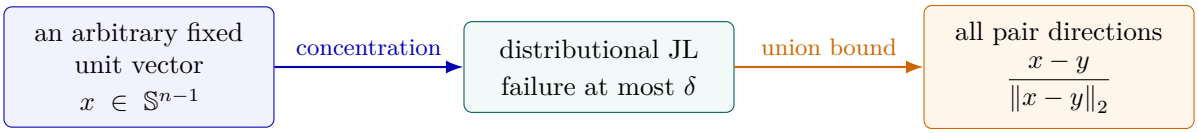
Overview

This lecture studies randomized dimension reduction. The Johnson–Lindenstrauss (JL) lemma states that a finite set of points in Euclidean space admits a linear embedding into a lower-dimensional space that approximately preserves every pairwise distance. The target dimension depends logarithmically on the number of points, rather than on the ambient dimension. The lemma originates in the work of Johnson and Lindenstrauss [1].

Theorem 1 (Johnson–Lindenstrauss lemma)**.** *Let $S \subset \mathbb{R}^n$ have $m \geq 2$ points and let $0 < \varepsilon < 1$. There exists a linear map $\Pi : \mathbb{R}^n \to \mathbb{R}^d$, with $d = O(\varepsilon^{-2} \log m)$, such that, for all $x, y \in S$,*

$$(1-\varepsilon)\left\|x-y\right\|_2 \leq \left\|\Pi x - \Pi y\right\|_2 \leq (1+\varepsilon)\left\|x-y\right\|_2 . \tag{1}$$

The same matrix must preserve all pairwise distances simultaneously. Sets with at most one point impose no distance constraints. We first show how a guarantee for an arbitrary fixed vector implies the finite-set result. We then prove that guarantee for dense subgaussian matrices using concentration inequalities. Finally, we extend the argument to fixed subspaces and analyze an elementary fast transform based on random signs, a Hadamard transform, and coordinate sampling.



Notation. Throughout, n is the ambient dimension, d the target dimension, m the number of points, and k the subspace dimension. All logarithms are natural unless a base is specified. The constants $c, C > 0$ may change from line to line; a subscript K allows dependence on the fixed subgaussian parameter K. We write $a \lesssim b$ for $a \leq Cb$ with a universal constant. We use δ for the failure probability of a fixed vector or fixed subspace, and ρ for total failure probability over a finite point set. For vectors, $\|x\|_2$ is the Euclidean norm and $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. For random variables, $\|X\|_{L^p} = (\mathbb{E}\,|X|^p)^{1/p}$, $p \geq 1$; $\|M\|_{\mathrm{op}} = \sup_{\|u\|_2=1} \|Mu\|_2$ is the Euclidean operator norm of a matrix.

1 From distributional JL to finite-set JL

1.1 Distributional JL and its quantifiers

Definition 2 (Distributional JL)**.** For $0 < \varepsilon, \delta < 1$, a distribution $\mathcal{P}$ over matrices in $\mathbb{R}^{d\times n}$ is an (ε, δ)-JL distribution if, for every fixed $x \in \mathbb{S}^{n-1}$,

$$\mathop{\mathbb{P}}_{\Pi\sim\mathcal{P}}\Big(\Big|\|\Pi x\|_2^2 - 1\Big| > \varepsilon\Big) \le \delta. \tag{2}$$

The vector x is fixed independently of the sampled matrix Π. The guarantee holds for every such fixed vector with the same parameters; it does not assert that one sampled matrix preserves every vector simultaneously. Indeed, if $d < n$, every linear map $\mathbb{R}^n \to \mathbb{R}^d$ has a nonzero kernel vector, whose norm it cannot preserve.

By linearity, the unit-vector guarantee also controls any fixed nonzero v: apply it to $v/\|v\|_2$ to obtain $|\,\|\Pi v\|_2^2 - \|v\|_2^2\,| \le \varepsilon\,\|v\|_2^2$ with probability at least $1-\delta$.

1.2 The union bound over pair directions

For a fixed set S of $m \ge 2$ points, define

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

There are at most m^2 directions. Applying (2) to each one and taking a union bound gives

$$\mathbb{P}\Big(\exists s \in S' : \Big|\|\Pi s\|_2^2 - 1\Big| > \varepsilon\Big) \le |S'|\delta \le m^2\delta. \tag{3}$$

The bad events for different directions need not be independent. On the complementary event, linearity gives, for every distinct $x, y \in S$,

$$\|\Pi x - \Pi y\|_2^2 = \|x-y\|_2^2 \left\|\Pi \frac{x-y}{\|x-y\|_2}\right\|_2^2 \in [(1-\varepsilon)\,\|x-y\|_2^2\,, (1+\varepsilon)\,\|x-y\|_2^2].$$

Taking square roots and using $1-\varepsilon \le \sqrt{1-\varepsilon}$ and $\sqrt{1+\varepsilon} \le 1+\varepsilon$ proves (1); the case $x = y$ is immediate. Thus a squared-norm guarantee with error ε implies the stated norm guarantee.

If the fixed-vector failure probability is at most $2e^{-c\varepsilon^2 d}$, the failure probability for the whole set is at most $2m^2e^{-c\varepsilon^2 d}$. To make it at most $\rho \in (0,1)$, it suffices that

$$d \ge \frac{1}{c\varepsilon^2}\log\frac{2m^2}{\rho}. \tag{4}$$

In particular, $\rho = 1/2$ gives $d = O(\varepsilon^{-2}\log m)$ and proves existence. The logarithmic dependence on m is the cost of this union bound.

2 A Gaussian candidate

Let $G \in \mathbb{R}^{d\times n}$ have independent $N(0,1)$ entries and set $\Pi = G/\sqrt{d}$. Fix $x \in \mathbb{S}^{n-1}$, write g_i^T for row i of G, and define $Z_i = \langle g_i, x\rangle$. A linear combination of independent Gaussian variables is Gaussian, so

$$Z_i = \sum_{j=1}^{n} G_{ij}x_j \sim N\left(0, \sum_{j=1}^{n} x_j^2\right) = N(0,1).$$

The variables Z_i are independent across rows, and

$$\|\Pi x\|_2^2 = \frac{1}{d}\sum_{i=1}^{d} Z_i^2, \qquad \mathbb{E}\,\|\Pi x\|_2^2 = 1. \tag{5}$$

Thus $d\,\|\Pi x\|_2^2$ has the chi-square distribution with d degrees of freedom. We need to concentrate an average of squared measurements. Although Z_i is subgaussian, $Z_i^2 - 1$ is not. The following concentration tools explain the appropriate larger class and will also cover non-Gaussian entries.

3 Concentration beyond Gaussian tails

3.1 Subgaussian MGFs and Chernoff bounds

A random variable X is *centered* if $\mathbb{E}\,X = 0$; concentration around the mean can be expressed in terms of $Y - \mathbb{E}\,Y$.

Definition 3 (Subgaussian variable)**.** A centered random variable X is σ-subgaussian, with $\sigma > 0$, if its moment-generating function (MGF) satisfies

$$\mathbb{E}\,e^{\lambda X} \le \exp(\sigma^2\lambda^2/2) \qquad \text{for every } \lambda \in \mathbb{R}. \tag{6}$$

The quantity σ^2 is a *variance proxy*; it need not equal $\mathrm{Var}(X)$. For $g \sim N(0,\sigma^2)$, completing the square in the Gaussian density gives $\mathbb{E}\,e^{\lambda g} = e^{\sigma^2\lambda^2/2}$. A Rademacher random variable ξ, uniform on $\{-1,1\}$, is 1-subgaussian since $\mathbb{E}\,e^{\lambda\xi} = \cosh\lambda \le e^{\lambda^2/2}$. The last inequality follows by comparing the series, using $(2r)! \ge 2^r r!$.

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

$$\mathbb{P}(X \ge t) \le e^{-\lambda t}\,\mathbb{E}\,e^{\lambda X} \le \exp(\sigma^2\lambda^2/2 - \lambda t).$$

Choosing $\lambda = t/\sigma^2$ and applying the same argument to $-X$ yields

$$\mathbb{P}(|X| \ge t) \le 2\exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{7}$$

3.2 Independent sums and Hoeffding's inequality

Lemma 4 (Weighted subgaussian sums)**.** *If X_i are independent and σ_i-subgaussian, then $\sum_i a_i X_i$ is $(\sum_i a_i^2\sigma_i^2)^{1/2}$-subgaussian for any real coefficients a_i, unless this scale is zero, in which case the sum is zero almost surely.*

Proof. Independence and (6) give, for every $\lambda \in \mathbb{R}$,

$$\mathbb{E}\exp\left(\lambda \sum_i a_i X_i\right) = \prod_i \mathbb{E}\, e^{\lambda a_i X_i} \le \exp\left(\frac{\lambda^2}{2}\sum_i a_i^2 \sigma_i^2\right). \qquad \square$$

For a centered variable X with $|X| \le b$ and $b > 0$, convexity gives

$$e^{\lambda X} \le \frac{b+X}{2b} e^{\lambda b} + \frac{b-X}{2b} e^{-\lambda b}.$$

Taking expectations yields $\mathbb{E}\, e^{\lambda X} \le \cosh(\lambda b) \le e^{\lambda^2 b^2/2}$. Hence bounded centered variables are subgaussian. Combining this fact with Lemma 4 gives the following form of *Hoeffding's inequality*: independent centered variables with $|X_i| \le b_i$ satisfy

$$\mathbb{P}\left(\left|\sum_i X_i\right| > t\right) \le 2\exp\left(-\frac{t^2}{2\sum_i b_i^2}\right), \qquad t > 0. \tag{8}$$

If $\sum_i b_i^2 = 0$, every summand is zero almost surely.

3.3 Moments and Gaussian moment growth

For $p \ge 1$, the L^p norms satisfy the triangle inequality $\|X+Y\|_{L^p} \le \|X\|_{L^p} + \|Y\|_{L^p}$ and monotonicity $\|X\|_{L^p} \le \|X\|_{L^q}$ for $p \le q$. The former is Minkowski's inequality; the latter follows from Jensen's inequality applied to $|X|^p$ and the convex power q/p. Applying Markov to $|X|^p$ gives

$$\mathbb{P}(|X| > t) \le \frac{\mathbb{E}\,|X|^p}{t^p} = \left(\frac{\|X\|_{L^p}}{t}\right)^p. \tag{9}$$

We may choose the real number $p \ge 1$ to match the deviation t.

For a standard Gaussian g, direct integration with $u = t^2/2$ gives

$$\mathbb{E}\,|g|^p = \frac{2}{\sqrt{2\pi}} \int_0^\infty t^p e^{-t^2/2}\, dt = \frac{2^{p/2}\Gamma((p+1)/2)}{\sqrt{\pi}},$$
$$c\sqrt{p} \le \|g\|_{L^p} \le C\sqrt{p}, \qquad p \ge 1. \tag{10}$$

Here $\Gamma(s) = \int_0^\infty u^{s-1}e^{-u}\, du$, and the bounds follow from Stirling's estimates for Γ (with constants adjusted on a bounded interval). Thus Gaussian L^p norms grow like $\sqrt{p}$; the pth moments themselves have order $(\sqrt{p})^p$ up to factors exponential in p.

Figure 1 illustrates the weighting of the tails. For $t > 0$, the logarithmic derivative of $t^p\phi(t)$ is $p/t - t$, so this moment integrand is maximized at $t = \sqrt{p}$.

3.4 Moment growth implies subgaussian tails

Suppose $\|X\|_{L^p} \le C\sigma\sqrt{p}$ for all $p \ge 1$. For $t \ge 2C\sigma$, choose $p = t^2/(4C^2\sigma^2) \ge 1$ in (9). Then

$$\mathbb{P}(|X| > t) \le \left(\frac{C\sigma\sqrt{p}}{t}\right)^p \le 2^{-p} = \exp\left(-\frac{(\log 2)t^2}{4C^2\sigma^2}\right).$$

For smaller t, decreasing the universal constant makes the bound $2e^{-ct^2/\sigma^2}$ at least 1. Hence the same form of tail bound holds for all $t > 0$.

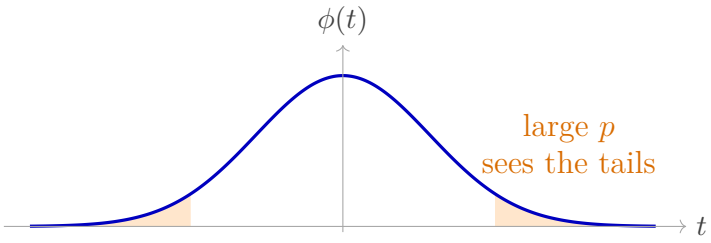


Figure 1: Schematic tail regions under the standard Gaussian density $\phi(t) = (2\pi)^{-1/2}e^{-t^2/2}$. In $\mathbb{E}\,|g|^p = \int_{\mathbb{R}} |t|^p \phi(t)\,dt$, increasing p gives larger magnitudes more weight; the Gaussian distribution and its tail probabilities remain unchanged.

3.5 Three equivalent faces of subgaussianity

For a centered variable, MGF, tail, and moment bounds are equivalent up to universal changes in the scale parameter. Figure 2 records the implications; the two remaining arguments follow below.

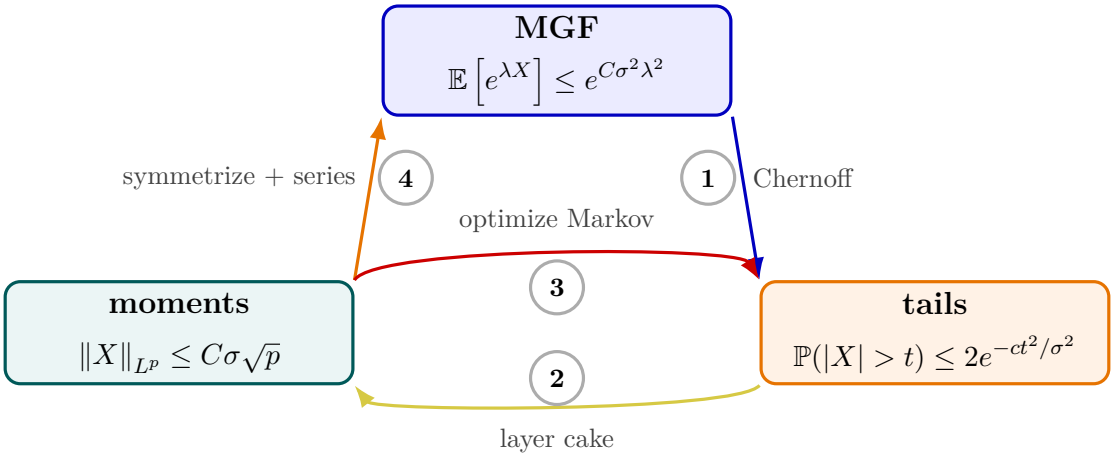


Figure 2: For a centered random variable, the MGF bound for all $\lambda \in \mathbb{R}$, the moment bound for all $p \ge 1$, and the tail bound for all $t > 0$ are equivalent up to universal changes in constants. The numbered arrows identify the four proof steps.

3.6 Layer cake: tails imply moments

The elementary identity $|X|^p = \int_0^\infty p t^{p-1} \mathbf{1}_{\{|X|>t\}}\,dt$ has a nonnegative integrand. Tonelli's theorem therefore gives the *layer-cake identity*

$$\mathbb{E}\,|X|^p = p \int_0^\infty t^{p-1}\,\mathbb{P}(|X| > t)\,dt. \tag{11}$$

Under the tail bound (7), substituting $u = t^2/(2\sigma^2)$ yields

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

Taking pth roots proves the moment bound. Centering is unnecessary for either direction between absolute tails and moments; it is needed when recovering an MGF bound with no linear term.

3.7 Moments imply an MGF bound: symmetrization

Suppose $\mathbb{E}\,X = 0$ and $\|X\|_{L^p} \le C\sigma\sqrt{p}$ for all $p \ge 1$. Let X' be an independent copy of X. For each fixed value of X, Jensen's inequality over X' gives

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

Set $Y = X - X'$. Exchanging X and X' does not change their joint law, so Y and $-Y$ have the same distribution. Also,

$$\|Y\|_{L^p} \le \|X\|_{L^p} + \|X'\|_{L^p} \le 2C\sigma\sqrt{p}.$$

This symmetrization removes the odd moments without assuming that X itself is symmetric.

Let $g_0 \sim N(0,1)$ and put $g = C_0\sigma g_0$. By the Gaussian lower bound in (10), one universal C_0 suffices to ensure

$$\|Y\|_{L^{2r}} \le 2C\sigma\sqrt{2r} \le C_0\sigma\,\|g_0\|_{L^{2r}} = \|g\|_{L^{2r}}\,, \qquad r \ge 1.$$

Thus $\mathbb{E}\,Y^{2r} \le \mathbb{E}\,g^{2r}$ for every integer $r \ge 1$; for $r = 0$ both moments are 1.

By symmetry, $\mathbb{E}\,e^{\lambda Y} = \mathbb{E}\cosh(\lambda Y)$, initially allowing extended expectations. The series of cosh has nonnegative terms, so Tonelli's theorem and the Gaussian comparison give

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

This finite upper bound also establishes existence of the MGF for every real λ. Combining it with Jensen's inequality proves the required MGF bound for X and completes the equivalence.

3.8 Squaring changes the tail scale

For a standard Gaussian g, the large-deviation tails satisfy

$$\mathbb{P}(|g| > t) = e^{-\Theta(t^2)} \quad (t \ge 1), \qquad \mathbb{P}(g^2 > u) = \mathbb{P}(|g| > \sqrt{u}) = e^{-\Theta(u)} \quad (u \ge 1).$$

For completeness, the upper bound follows from (7); a lower bound follows by integrating the Gaussian density over $[t, t + 1/t]$. For $t \ge 1$ this gives $\mathbb{P}(g > t) \ge (\sqrt{2\pi}\,t)^{-1}e^{-(t+1/t)^2/2}$, which is at least e^{-Ct^2} for a universal C.

In particular, a centered Gaussian square is not subgaussian. Its exact MGF makes the distinction explicit:

$$\mathbb{E}\,e^{\lambda(g^2-1)} = e^{-\lambda}(1-2\lambda)^{-1/2} \quad (\lambda < 1/2), \tag{12}$$

while the expectation is infinite for $\lambda \ge 1/2$. This identity follows by combining $e^{\lambda g^2}$ with the Gaussian density and evaluating the resulting Gaussian integral. More generally, centering the square of a subgaussian variable produces a subexponential variable, as proved below. Some such squares remain subgaussian; for a Rademacher sign, $\xi^2 - \mathbb{E}\,\xi^2 = 0$.

3.9 Subexponential variables and two tail regimes

Definition 5 (Subexponential variable)**.** A centered random variable X is (ν, b)-subexponential, with $\nu, b > 0$, if

$$\mathbb{E}\, e^{\lambda X} \le \exp(\nu^2\lambda^2/2) \qquad \text{whenever } |\lambda| < 1/b. \tag{13}$$

The quadratic MGF bound now holds only near the origin. The parameter ν^2 is a variance proxy, and b controls the size of this neighborhood.

Lemma 6 (Subexponential tail bound)**.** *If X is (ν, b)-subexponential, then, for every $t > 0$,*

$$\mathbb{P}(|X| > t) \le 2\exp\left[-\frac{1}{4}\min\left\{\frac{t^2}{\nu^2}, \frac{t}{b}\right\}\right]. \tag{14}$$

Proof. For $0 < \lambda < 1/b$, Chernoff's inequality gives $\mathbb{P}(X \ge t) \le \exp(\nu^2\lambda^2/2 - \lambda t)$. Choose $\lambda = \min\{t/\nu^2, 1/(2b)\}$. If $t \le \nu^2/(2b)$, the exponent is $-t^2/(2\nu^2)$. Otherwise,

$$\frac{\nu^2}{8b^2} - \frac{t}{2b} \le -\frac{t}{4b}.$$

In both cases the exponent is at most $-\frac{1}{4}\min\{t^2/\nu^2, t/b\}$. Apply the same argument to $-X$ and add the two bounds. □

The two expressions in the minimum agree at $t = \nu^2/b$: the upper bound has a quadratic exponent for moderate deviations and a linear exponent for large deviations. This is a concentration upper bound, not a claim that every subexponential variable has exactly that tail decay.

3.10 Subexponential tails and moments

Up to universal constants, the two-regime tail bound is equivalent to

$$\|X\|_{L^p} \lesssim \nu\sqrt{p} + bp, \qquad p \ge 1. \tag{15}$$

We prove both directions. Suppose first that $\mathbb{P}(|X| > t) \le 2e^{-c\min\{t^2/\nu^2, t/b\}}$ for a fixed universal $c > 0$. Since $e^{-\min\{a,b\}} \le e^{-a} + e^{-b}$, layer cake gives

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

Taking pth roots and using $(u^p + v^p)^{1/p} \le u + v$ for $u, v \ge 0$ proves (15).

Conversely, suppose $\|X\|_{L^p} \le A(\nu\sqrt{p} + bp)$ for all $p \ge 1$, where A is a fixed universal constant. Choose a universal $\alpha > 0$ such that $A(\sqrt{\alpha} + \alpha) \le 1/2$, and put

$$p = \alpha\min\left\{\frac{t^2}{\nu^2}, \frac{t}{b}\right\}.$$

If $p \geq 1$, then $\|X\|_{L^p} \leq t/2$, and Markov's inequality yields

$$\mathbb{P}(|X| > t) \leq 2^{-p} = \exp\left[-\alpha \log 2 \min\left\{\frac{t^2}{\nu^2}, \frac{t}{b}\right\}\right].$$

If $p < 1$, the same bound with a prefactor 2 exceeds 1. This proves the claimed equivalence between tails and moments. It does not by itself recover a local MGF with the same arbitrary pair (ν, b) up to constants. For centered squares, the following single-scale implication supplies the MGF bound we need.

Lemma 7 (Linear moment growth implies a local MGF bound)**.** *Let Y be centered and suppose $\|Y\|_{L^p} \leq Lp$ for all $p \geq 1$, where $L > 0$. Then Y is $(2eL, 2eL)$-subexponential.*

Proof. For $|\lambda| < 1/(2eL)$, the inequality $r! \geq (r/e)^r$ gives

$$\sum_{r=2}^{\infty} \frac{|\lambda|^r \, \mathbb{E}\,|Y|^r}{r!} \leq \sum_{r=2}^{\infty} (eL|\lambda|)^r \leq 2e^2 L^2 \lambda^2.$$

Together with the finite terms $r = 0, 1$, this justifies integrating the exponential series term by term. Since $\mathbb{E}\, Y = 0$, the linear term vanishes, and

$$\mathbb{E}\, e^{\lambda Y} \leq 1 + 2e^2 L^2 \lambda^2 \leq e^{2e^2 L^2 \lambda^2}. \qquad \square$$

3.11 Independent subexponential sums: Bernstein's inequality

Lemma 8 (Subexponential sums)**.** *If X_i are independent and (ν_i, b_i)-subexponential, then $\sum_i X_i$ is (ν, b)-subexponential for $\nu^2 = \sum_i \nu_i^2$ and $b = \max_i b_i$. Consequently,*

$$\mathbb{P}\left(\left|\sum_i X_i\right| > t\right) \leq 2 \exp\left[-\frac{1}{4} \min\left\{\frac{t^2}{\sum_i \nu_i^2}, \frac{t}{\max_i b_i}\right\}\right]. \tag{16}$$

Proof. For $|\lambda| < 1/b$, independence gives

$$\mathbb{E}\, e^{\lambda \sum_i X_i} = \prod_i \mathbb{E}\, e^{\lambda X_i} \leq \exp\left(\frac{\lambda^2}{2} \sum_i \nu_i^2\right).$$

The tail bound follows from Lemma 6. $\square$

Unlike subgaussian concentration at all scales, this Bernstein bound switches from a quadratic to a linear exponent for large deviations.

3.12 Bounded Bernstein and the actual variance

For bounded variables, a variance-sensitive form is especially useful. It will save one logarithmic factor in the Fast JL analysis.

Theorem 9 (Bounded Bernstein inequality)**.** *Let X_i be independent centered random variables with $|X_i| \le B$, where $B > 0$, and let $V = \sum_i \mathbb{E}\, X_i^2$. Then, for every $t > 0$,*

$$\mathbb{P}\left(\left|\sum_i X_i\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{2(V + Bt/3)}\right). \tag{17}$$

Proof. For every integer $r \ge 2$, boundedness gives $\mathbb{E}\,|X_i|^r \le B^{r-2}\,\mathbb{E}\, X_i^2$, and $r! \ge 2 \cdot 3^{r-2}$. Using $\mathbb{E}\, X_i = 0$, for $|\lambda| B < 3$ we obtain

$$\begin{aligned}\mathbb{E}\, e^{\lambda X_i} &\le 1 + \sum_{r=2}^{\infty} \frac{|\lambda|^r B^{r-2}\,\mathbb{E}\, X_i^2}{r!} \\ &\le 1 + \frac{\lambda^2\,\mathbb{E}\, X_i^2}{2(1 - |\lambda| B/3)} \le \exp\left(\frac{\lambda^2\,\mathbb{E}\, X_i^2}{2(1 - |\lambda| B/3)}\right).\end{aligned}$$

Independence bounds the log-MGF of the sum by $\lambda^2 V/[2(1 - |\lambda| B/3)]$. For $V > 0$, choose $\lambda = t/(V + Bt/3) < 3/B$ in Chernoff's inequality:

$$\mathbb{P}\left(\sum_i X_i \ge t\right) \le \exp\left(-\lambda t + \frac{\lambda^2 V}{2(1 - \lambda B/3)}\right) = \exp\left(-\frac{t^2}{2(V + Bt/3)}\right).$$

Apply the same argument to $-X_i$ and add the two bounds. If $V = 0$, every X_i is zero almost surely, so the result is immediate. □

For $V > 0$, this also yields the equivalent-order upper bound $2\exp[-c \min\{t^2/V, t/B\}]$. In contrast to a bound based only on the range B, Bernstein retains the potentially much smaller total variance V.

3.13 Centered squares of subgaussians

Lemma 10 (A centered square is subexponential)**.** *If X is K-subgaussian, then $Y = X^2 - \mathbb{E}\, X^2$ is (CK^2, CK^2)-subexponential for a universal constant C.*

Proof. The subgaussian moment bound and the triangle inequality imply, for $p \ge 1$,

$$\begin{aligned}\left\|X^2 - \mathbb{E}\, X^2\right\|_{L^p} &\le \left\|X^2\right\|_{L^p} + \mathbb{E}\, X^2 \\ &= \|X\|_{L^{2p}}^2 + \|X\|_{L^2}^2 \le C_1 K^2 p.\end{aligned}$$

The identity in the second line follows directly from the definition of the L^p norm. Since Y is centered, applying Lemma 7 with $L = C_1 K^2$ proves the claim. □

In particular, for $Z \sim N(0,1)$, the variable $Z^2 - 1$ has subexponential parameters $(O(1), O(1))$. This is the concentration class needed for the average in (5).

4 Dense random projections

Theorem 11 (Dense subgaussian JL)**.** *Let $G \in \mathbb{R}^{d\times n}$ have independent entries satisfying $\mathbb{E}\, G_{ij} = 0$, $\mathbb{E}\, G_{ij}^2 = 1$, and the K-subgaussian MGF bound, where $K = O(1)$. Set $\Pi = G/\sqrt{d}$. For every fixed $x \in \mathbb{S}^{n-1}$ and $0 < \varepsilon < 1$,*

$$\mathbb{P}\Big(\Big|\|\Pi x\|_2^2 - 1\Big| > \varepsilon\Big) \le 2e^{-c_K \varepsilon^2 d}. \tag{18}$$

In particular, $d \ge C_K \varepsilon^{-2} \log(2/\delta)$ makes this an (ε, δ)-JL distribution for any $0 < \delta < 1$.

Proof. Fix a unit vector x and write $Z_i = \sum_j G_{ij} x_j$. We verify the mean, the concentration class, and the concentration of the sum.

The mean is exactly right. By centering and independence within row i,

$$\mathbb{E}\, Z_i^2 = \sum_j x_j^2\, \mathbb{E}\, G_{ij}^2 + \sum_{j\neq\ell} x_j x_\ell\, \mathbb{E}\, G_{ij}\, \mathbb{E}\, G_{i\ell} = \sum_j x_j^2 = 1.$$

The mixed terms vanish, and therefore $\mathbb{E}\, \|\Pi x\|_2^2 = d^{-1} \sum_i \mathbb{E}\, Z_i^2 = 1$.

Each row measurement is subgaussian. For every $\lambda \in \mathbb{R}$, independence within the row gives

$$\mathbb{E}\, e^{\lambda Z_i} = \prod_j \mathbb{E}\, e^{\lambda x_j G_{ij}} \le \exp\left(\frac{K^2\lambda^2}{2} \sum_j x_j^2\right) = e^{K^2\lambda^2/2}.$$

Thus Z_i is K-subgaussian. The variables Z_i are independent across rows, and Lemma 10 makes $Y_i = Z_i^2 - 1$ centered and (C_K, C_K)-subexponential.

Average the squared measurements. The exact identity $\|\Pi x\|_2^2 - 1 = d^{-1} \sum_i Y_i$ and Lemma 8 give

$$\begin{aligned}\mathbb{P}\Big(\Big|\|\Pi x\|_2^2 - 1\Big| > \varepsilon\Big) &= \mathbb{P}\left(\left|\sum_i Y_i\right| > \varepsilon d\right) \\ &\le 2\exp[-c_K \min\{\varepsilon^2 d, \varepsilon d\}] = 2e^{-c_K\varepsilon^2 d},\end{aligned}$$

where the last equality uses $0 < \varepsilon < 1$. Choosing $d \ge C_K \varepsilon^{-2} \log(2/\delta)$ completes the proof. □

Independent standard Gaussian entries and independent Rademacher signs both satisfy the assumptions with $K = 1$. Combining Theorem 11 with the pair-direction union bound in Section 1 proves Theorem 1. More generally, total failure probability at most ρ is obtained with $d = O_K(\varepsilon^{-2} \log(m/\rho))$ for $m \ge 2$.

5 Application: oblivious subspace embeddings

A union bound handles a finite set directly. To preserve a whole subspace, which contains infinitely many vectors, we use a finite net and then lift the guarantee by an operator-norm argument.

Theorem 12 (Oblivious subspace embedding)**.** *Let $V \subset \mathbb{R}^n$ be a fixed k-dimensional subspace, with $1 \le k \le n$, and let Π be the normalized dense subgaussian matrix of Theorem 11. For $0 < \varepsilon, \delta < 1$, if*

$$d \ge C_K \varepsilon^{-2}(k + \log(2/\delta)), \tag{19}$$

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

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

The distribution of Π does not depend on V; this is the meaning of *oblivious*. The subspace is fixed independently of the sampled matrix. The theorem does not assert that one short matrix preserves all subspaces simultaneously.

5.1 A small net of the unit sphere

An η-net $\mathcal{N} \subset \mathbb{S}^{k-1}$ is a finite set such that every $u \in \mathbb{S}^{k-1}$ is within Euclidean distance η of some $z \in \mathcal{N}$. For $0 < \eta < 1$, there is such a net with

$$|\mathcal{N}| \le (1 + 2/\eta)^k. \tag{20}$$

To see this, take a maximal set of sphere points whose mutual distances exceed η. The open balls of radius $\eta/2$ around these points are disjoint and lie inside the ball of radius $1 + \eta/2$ about the origin. Write $B_2^k = \{z \in \mathbb{R}^k : \|z\|_2 \le 1\}$ for the Euclidean unit ball. Comparing their k-dimensional volumes gives

$$|\mathcal{N}|(\eta/2)^k \operatorname{vol}(B_2^k) \le (1+\eta/2)^k \operatorname{vol}(B_2^k),$$

which proves (20). The same packing bound ensures that a greedy construction terminates. Maximality ensures the covering property: a point farther than η from every chosen point could otherwise be added. In particular, a $1/4$-net has at most 9^k points.

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

$$M = U^T \Pi^T \Pi U - I_k. \tag{21}$$

For $z \in \mathbb{S}^{k-1}$, Uz is a unit vector and $z^T M z = \|\Pi U z\|_2^2 - 1$. Choose a fixed $1/4$-net $\mathcal{N}$. The dense JL tail bound with error $\varepsilon/2$ and a union bound imply

$$\mathbb{P}\Big(\exists z \in \mathcal{N} : |z^T M z| > \varepsilon/2\Big) \le 2 \cdot 9^k e^{-c_K \varepsilon^2 d} \le \delta, \tag{22}$$

where the factor $1/4$ in the exponent is absorbed into c_K. The last inequality follows from (19) for a sufficiently large C_K.

5.2 Lifting from the net to every vector

Lemma 13 (Quadratic-form lifting)**.** *For a real symmetric matrix M and an η-net $\mathcal{N}$ of $\mathbb{S}^{k-1}$ with $0 < \eta < 1/2$,*

$$\|M\|_{\mathrm{op}} \le \frac{1}{1-2\eta} \max_{z \in \mathcal{N}} |z^T M z|. \tag{23}$$

Proof. By the spectral theorem, choose a unit vector u such that $|u^T M u| = \|M\|_{\mathrm{op}}$. Take $z \in \mathcal{N}$ with $\|u - z\|_2 \le \eta$. The decomposition $u^T M u - z^T M z = (u-z)^T M u + z^T M (u-z)$ gives

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

Here we used $|a^T M b| \le \|a\|_2 \|M\|_{\mathrm{op}} \|b\|_2$ and $\|u\|_2 = \|z\|_2 = 1$. Rearranging proves the claim since $1 - 2\eta > 0$. □

On the good event in (22), the choice $\eta = 1/4$ gives $\|M\|_{\mathrm{op}} \le 2(\varepsilon/2) = \varepsilon$. Every $x \in V$ can be written as $x = Ua$, with $\|x\|_2 = \|a\|_2$. Therefore

$$\left| \|\Pi x\|_2^2 - \|x\|_2^2 \right| = |a^T M a| \le \|M\|_{\mathrm{op}} \|a\|_2^2 \le \varepsilon \|x\|_2^2.$$

This completes the proof of Theorem 12. A constant-radius net suffices because the lifting bound absorbs the approximation error into the unknown operator norm; an ε-net and its extra $\log(1/\varepsilon)$ in the dimension bound are unnecessary.

6 Fast Johnson–Lindenstrauss

Ailon and Chazelle introduced the fast Johnson–Lindenstrauss transform (FJLT) [2]. Here we analyze the elementary construction from the lecture: random signs, a Hadamard transform, and uniform coordinate sampling. Its dimension bound is weaker than the dense JL bound, but its structured form reduces the cost of applying the map.

6.1 The structured transform

Applying an explicitly stored dense $d \times n$ matrix by standard matrix–vector multiplication takes $\Theta(nd)$ arithmetic operations. The structured transform below can be applied in $O(n \log n + d)$ time. Assume for now that n is a power of two and define

$$\Pi = PHD \in \mathbb{R}^{d \times n}. \tag{24}$$

The three factors are as follows:

- $D = \mathrm{diag}(\xi_1, \dots, \xi_n)$, with independent Rademacher signs ξ_j.
- $H \in \mathbb{R}^{n \times n}$ is the normalized Walsh–Hadamard matrix, with $H^T H = I_n$ and $|H_{ij}| = 1/\sqrt{n}$.
- Independently of D, choose $I_1, \dots, I_d$ independently and uniformly from $[n] = \{1, \dots, n\}$. Set $R_{s,:} = e_{I_s}^T$ and $P = \sqrt{n/d}\, R$.

Sampling is with replacement, so repeated rows are allowed. The factor $\sqrt{n/d}$ makes the sampled squared norm an unbiased estimator of the full squared norm. The actual vector pipeline is

$$x \xrightarrow{D} Dx \xrightarrow{H} HDx \xrightarrow{P} \sqrt{n/d}\,\big((HDx)_{I_1}, \dots, (HDx)_{I_d}\big)^T.$$

The signs do not change coordinate magnitudes by themselves. Their role is to prevent a fixed input from aligning with the Hadamard basis: without D, the fixed unit vector $x = H^T e_j$ satisfies $Hx = e_j$. Uniform sampling misses that nonzero coordinate with probability $(1 - 1/n)^d$, which is large when $d \ll n$. The combination HD instead makes every coordinate small with high probability.

6.2 Flattening by random signs and a Hadamard transform

Lemma 14 (Flattening)**.** *For every fixed $x \in \mathbb{S}^{n-1}$ and $\gamma > 0$,*

$$\mathop{\mathbb{P}}_{D}\left(\|HDx\|_\infty > \frac{\gamma}{\sqrt{n}} \right) \le 2ne^{-\gamma^2/2}. \tag{25}$$

Proof. Write $y = HDx$. For a fixed coordinate i, $y_i = \sum_j H_{ij}\xi_j x_j$ is a sum of independent centered variables. Lemma 4 gives the subgaussian variance proxy

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

Thus $\mathbb{P}_D(|y_i| > \gamma/\sqrt{n}) \le 2e^{-\gamma^2/2}$. A union bound over all n coordinates proves (25). □

The coordinates of HDx need not be independent. Independence is used within each coordinate sum; the union bound needs no independence between coordinates. Also $\|HDx\|_2 = \|x\|_2 = 1$ deterministically, since both H and D are orthogonal.

6.3 Sampling a fixed flat vector

Fix a realization of D for which $y = HDx$ satisfies $\|y\|_2 = 1$ and $\|y\|_\infty \le \gamma/\sqrt{n}$, with $\gamma \ge 1$. Since the indices I_s are independent of D, they remain independent and uniform after conditioning on D. All expectations and probabilities in this sampling argument are over these indices, with y held fixed; we denote them by $\mathbb{E}_R$ and $\mathbb{P}_R$.

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

$$\|Py\|_2^2 = \frac{n}{d}\sum_{s=1}^{d} y_{I_s}^2 = \frac{1}{d}\sum_{s=1}^{d} W_s, \qquad \mathop{\mathbb{E}}_{R} W_s = \sum_{j=1}^{n} y_j^2 = 1. \tag{26}$$

Flatness gives both a range bound and a useful second-moment bound:

$$0 \le W_s \le \gamma^2, \qquad \mathop{\mathbb{E}}_{R} W_s^2 = n \sum_j y_j^4 \le n \, \|y\|_\infty^2 \, \|y\|_2^2 \le \gamma^2. \tag{27}$$

Therefore $\mathrm{Var}_R(W_s) \le \gamma^2$ and $|W_s - 1| \le \gamma^2$. Crucially, the variance has order at most γ^2, although squaring the range bound alone would give γ^4.

The variables $W_s - 1$ are independent and centered under this conditional distribution. Applying bounded Bernstein with $V \le d\gamma^2$, $B = \gamma^2$, and $t = d\varepsilon$ yields, for $0 < \varepsilon \le 1$,

$$\begin{aligned} \mathop{\mathbb{P}}_{R}\Big(\big| \|Py\|_2^2 - 1 \big| > \varepsilon \Big) &\le 2 \exp\left(- \frac{d\varepsilon^2}{2\gamma^2(1 + \varepsilon/3)} \right) \\ &\le 2 \exp\left(- \frac{3d\varepsilon^2}{8\gamma^2} \right). \end{aligned} \tag{28}$$

The last step uses $1 + \varepsilon/3 \le 4/3$.

6.4 Why the variance bound matters

Using only $0 \le W_s \le \gamma^2$, Hoeffding gives the upper bound $2\exp(-cd\varepsilon^2/\gamma^4)$. Bernstein also uses $\mathrm{Var}_R(W_s) \le \gamma^2$ and improves it to $2\exp(-cd\varepsilon^2/\gamma^2)$. Since our choice below has $\gamma^2 = \Theta(\log(n/\delta))$ for $0 < \delta < 1/2$, the variance estimate saves a factor of $\log(n/\delta)$ in the sufficient sample count. These statements compare concentration upper bounds, not matching upper and lower estimates of the actual failure probability.

6.5 Choosing parameters and removing the conditioning

Fix $0 < \varepsilon, \delta < 1/2$ and set $\gamma^2 = 2\log(4n/\delta)$. Let

$$F = \left\{ \|HDx\|_\infty \le \frac{\gamma}{\sqrt{n}} \right\}, \qquad \mathcal{B} = \left\{ \big| \|\Pi x\|_2^2 - 1 \big| > \varepsilon \right\}.$$

The flattening lemma gives $\mathbb{P}_D(F^c) \le \delta/2$. For every fixed realization of D in F, the conditional sampling bound (28) is at most $\delta/2$ provided

$$d \ge \frac{8\gamma^2}{3\varepsilon^2} \log \frac{4}{\delta} = \frac{16}{3\varepsilon^2} \log \frac{4n}{\delta} \log \frac{4}{\delta}. \tag{29}$$

The law of total probability now gives

$$\mathop{\mathbb{P}}_{D,R}(\mathcal{B}) \le \mathop{\mathbb{P}}_{D}(F^c) + \mathop{\mathbb{E}}_{D}\Big[\mathbf{1}_F \mathop{\mathbb{P}}_{R}(\mathcal{B} \mid D) \Big] \le \delta. \tag{30}$$

This step uses conditional independence of the sampled variables, not an unconditional independence claim for variables sharing the random vector HDx.

Theorem 15 (Elementary Fast JL)**.** *Let n be a power of two, and let D, H, P be defined as above. For every fixed $x \in \mathbb{S}^{n-1}$ 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

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

In particular, $d \ge C\varepsilon^{-2} \log^2(4n/\delta)$ is sufficient.

The product-of-logarithms bound is the direct result of the proof. The larger square-of-a-logarithm bound follows from $\log(4/\delta) \le \log(4n/\delta)$.

6.6 Running time, padding, and finite sets

The matrix Π need not be formed explicitly. Applying D takes $O(n)$ operations. The normalized Hadamard matrices satisfy

$$H_1 = [1], \qquad H_{2r} = \frac{1}{\sqrt{2}} \begin{pmatrix} H_r & H_r \\ H_r & -H_r \end{pmatrix}.$$

This recursion preserves $H^T H = I$ and the equal entry magnitudes. For an input (u, v), recursively compute $H_r u$ and $H_r v$, then form their sum and difference, each scaled by $1/\sqrt{2}$. Thus the multiplication time satisfies $T(n) = 2T(n/2) + O(n) = O(n \log n)$ (with constant cost for $n = 1$). Selecting and rescaling d coordinates costs $O(d)$, giving total time

$$O(n \log n + d). \tag{31}$$

The signs and sampled indices require $O(n + d)$ words of storage; the Hadamard matrix is implicit.

If n is not a power of two, pad every input with zeros to dimension $N = 2^{\lceil \log_2 n \rceil} < 2n$. Padding preserves all distances and changes these bounds only by constant factors. Sampling with replacement also allows $d > n$; in that regime the identity map already preserves norms exactly in n dimensions. The structured multiplication is useful in regimes where its cost is smaller; for very small d, dense multiplication can be cheaper.

For a fixed set of $m \ge 2$ points and total failure probability $\rho \in (0, 1/2)$, take $\delta = \rho/m^2$ and apply the union bound to the normalized pair differences. It suffices to use

$$d = O\left(\varepsilon^{-2} \log \frac{4nm^2}{\rho} \log \frac{4m^2}{\rho}\right). \tag{32}$$

For constant ρ, this is $O(\varepsilon^{-2} \log(nm) \log m)$. All points are mapped using the same sampled matrix. Compared with dense JL's sufficient dimension $O(\varepsilon^{-2} \log m)$, this elementary analysis pays an extra factor of order $\log(nm)$ in the dimension bound to obtain structured multiplication.

Dimension and computation at a glance

The sufficient dimensions below suppress universal constants and assume $K = O(1)$. In each row, d denotes that construction's output dimension.

Guarantee	Sufficient dimension d	Time per vector
Dense, one fixed vector	$O(\varepsilon^{-2}\log(2/\delta))$	$O(nd)$
Dense, m fixed points	$O(\varepsilon^{-2}\log(m/\rho))$	$O(nd)$
Dense, fixed k-subspace	$O(\varepsilon^{-2}(k+\log(2/\delta)))$	$O(nd)$
Fast, one fixed vector	$O(\varepsilon^{-2}\log(4n/\delta)\log(4/\delta))$	$O(n\log n + d)$
Fast, m fixed points	$O(\varepsilon^{-2}\log(4nm^2/\rho)\log(4m^2/\rho))$	$O(n\log n + d)$

Here δ is the failure probability for a fixed vector or fixed subspace; ρ is the total failure probability for a finite set. The proof mechanisms are concentration for one vector, a union bound for finitely many pair directions, and a net plus lifting for a whole subspace.

AI use disclosure

GPT-5.6 was used to assist with generating the TikZ figures, polishing the English, and checking the mathematical derivations. ChatGPT also assisted with revision, proofreading, and notation checks.

References

[1] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in Modern Analysis and Probability*, Contemporary Mathematics, volume 26, pages 189–206. American Mathematical Society, 1984. DOI: 10.1090/conm/026/737400.

[2] Nir Ailon and Bernard Chazelle. The Fast Johnson–Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM Journal on Computing*, 39(1):302–322, 2009. DOI: 10.1137/060673096.