

Deck 9: Random embeddings

Math 7870: Topics in Randomized Numerical Linear Algebra

Spring 2026

Akil Narayan

(Linear) Embeddings

We now consider the task of compressing high-dimensional vectors into lower-dimensional space

- We've already seen this through the Johnson-Lindenstrauss Lemma, though that operated on finite sets of points.
- Accomplishing this compression in more general scenarios will reveal new capabilities for us: randomized least squares and the randomized SVD.

Let $V \subset \mathbb{R}^n$ be an arbitrary set. A linear map (matrix) $\mathbf{S} : \mathbb{R}^n \rightarrow \mathbb{R}^d$ is called an ℓ^2 **embedding of distortion** $\epsilon \in (0, 1)$ on V if,

$$(1 - \epsilon)\|\mathbf{x}\|_2 \leq \|\mathbf{S}\mathbf{x}\|_2 \leq (1 + \epsilon)\|\mathbf{x}\|_2, \quad \mathbf{x} \in V.$$

Of particular interest:

- $d \ll n$
- How large must d be to achieve a specific ϵ ?
- Generate \mathbf{S} randomly.

Random embeddings

Randomness is a conceptually reasonable strategy to generate \mathbf{S} since if \mathbf{S} is random, then for every deterministic \mathbf{x} :

$$\mathbb{E}\|\mathbf{S}\mathbf{x}\|_2^2 = \mathbf{x}^* \mathbb{E}[\mathbf{S}^* \mathbf{S}] \mathbf{x}.$$

Therefore, if \mathbf{S} is isotropic: $\mathbb{E}\mathbf{S}^* \mathbf{S} = \mathbf{I}$, then this *on average* is an embedding with zero distortion.

By now, we know that we can make low-distortion a reality with high probability if we take enough rows of \mathbf{S} .

E.g., if $\mathbf{x} \in \mathbb{R}^n$ is an isotropic random vector, then filling the rows of \mathbf{S} with sufficiently many iid copies of \mathbf{x} should do the trick.

Restricted singular values

Before discussing randomness, let's probe some machinery for embeddings a bit more. Recall that \mathbf{S} is an ϵ -distortion of V if,

$$(1 - \epsilon)\|\mathbf{x}\|_2 \leq \|\mathbf{S}\mathbf{x}\|_2 \leq (1 + \epsilon)\|\mathbf{x}\|_2, \quad \mathbf{x} \in V.$$

The (extremal) V -restricted singular values of \mathbf{S} are,

$$\sigma_{\min}(\mathbf{S}; V) := \inf_{\mathbf{x} \in V \setminus \{0\}} \frac{\|\mathbf{S}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}, \quad \sigma_{\max}(\mathbf{S}; V) := \sup_{\mathbf{x} \in V \setminus \{0\}} \frac{\|\mathbf{S}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}.$$

Note $V = \mathbb{R}^n$ recovers the standard definition of singular values.

What's really important: $\sigma_{\min}(\mathbf{S}; V) \neq 0$.

It's ok if $\sigma_{\max}(\mathbf{S}; V)$ is large. (This is in general an injection, which works just fine.)

Note then that \mathbf{S} is an ϵ -distortion embedding of V iff $\sigma_{\min}(\mathbf{S}; V) \geq 1 - \epsilon$ and $\sigma_{\max}(\mathbf{S}; V) \leq 1 + \epsilon$.

Gaussian embeddings and width

To study randomness, let's keep things simple to start, and consider Gaussian embeddings.

E.g., if $\mathbf{S} \in \mathbb{R}^{d \times n}$ is a matrix with iid $\mathcal{N}(0, \frac{1}{d})$ (Gaussian) entries, then it's isotropic.

A key tool to understand random Gaussian embeddings is the following concept: Given a set $V \subset \mathbb{R}^n$, the quantity

$$w(V) := \mathbb{E} \sup_{\mathbf{x} \in V} \langle \mathbf{x}, \mathbf{y} \rangle, \quad \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

is called the **Gaussian width** of V .

Informally: the Gaussian width measures the size or complexity of V . The quantity $\sup_{\mathbf{x} \in V} \langle \mathbf{x}, \mathbf{y} \rangle$ measures the maximum reach of V in the (unnormalized) direction \mathbf{y} .

The Gaussian width averages these maximum reaches.

Gaussian width properties

Here are some important properties of Gaussian width:

- $w(V) < \infty$ iff V is a bounded set.
- $\frac{1}{\sqrt{2\pi}}\text{diam}(V) \leq w(V) \leq \frac{\sqrt{n}}{2}\text{diam}(V)$, where V is the diameter of V .
- The Gaussian width is invariant under taking convex hulls: $w(V) = w(\text{conv}(V))$.
- The Gaussian width is monotonic: if $V_1 \subset V_2$, then $w(V_1) \leq w(V_2)$.
- For any deterministic unitary matrix \mathbf{A} and any vector \mathbf{b} , we have $w(V) = w(\mathbf{U}V + \mathbf{b})$.
- For any $c \in \mathbb{R}$, $w(cV) = |c|w(V)$.
- For any sets V_1 and V_2 , $w(V_1 + V_2) = w(V_1) + w(V_2)$, where $V_1 + V_2$ is defined as Minkowski (set) addition.

Gaussian width examples

Let B_p^n denote the unit ball in \mathbb{R}^n in the $\ell^p(\mathbb{R}^n)$ norm. Then for absolute constants c_j :

$$w(B_2^n) \in [\sqrt{n} - c_1, \sqrt{n} + c_1], \quad w(B_1^n) \in [c_2\sqrt{\log n}, c_3\sqrt{\log n}], \quad w(B_\infty^n) = n \frac{\sqrt{2\pi}}{.}$$

In fact, in general $w(B_p^n) \leq c_4\sqrt{q}n^{1/q}$, where (p, q) are conjugate exponents: $\frac{1}{p} + \frac{1}{q} = 1$.

If V is a finite point set, $|V| < \infty$, then $w(V) \leq c_5 \text{diam}(V)\sqrt{\log |V|}$.

If V is a subspace of dimension k , restricted to the unit sphere, then

$$w(V) \in [\sqrt{k-1}, \sqrt{k}].$$

Gaussian widths to embeddings

We can tie many of these results together: The Gaussian width controls concentration of a Gaussian embedding.

Theorem 1. *Let V be a subset of the unit sphere in \mathbb{R}^n . Let $\mathbf{S} \in \mathbb{R}^{d \times n}$ be a Gaussian embedding. (It has iid $\mathcal{N}(0, \frac{1}{d})$ entries. Then the restricted singular values of \mathbf{S} satisfy,*

$$\Pr \left(\sigma_{\min}(\mathbf{S}; V) \geq 1 + \frac{w(V)}{\sqrt{d}} + t \right) \leq \exp(-dt^2/2)$$
$$\Pr \left(\sigma_{\max}(\mathbf{S}; V) \leq 1 - \frac{1 + w(V)}{\sqrt{d}} - t \right) \leq \exp(-dt^2/2).$$

The punch line: picking $t = \epsilon$ and $d \gtrsim w(V)^2/\epsilon^2$ achieves an embedding with high probability.

NB: While Gaussian widths care about sizes of vectors, *subspace* embeddings do not: Given $\sigma_{\min}(V; \mathbf{S})$ for a spherical set V , the set cV for any nontrivial constant c has the same restricted singular values.

Quick-and-dirty application: Johnson-Lindenstrauss

Let's revisit the Johnson-Lindenstrauss Lemma: Let $\{\mathbf{x}_j\}_{j \in [N]} \subset \mathbb{R}^n$ be given. The conclusion of the Johnson-Lindenstrauss Lemma is that there's a linear embedding $\mathbf{S} \in \mathbb{R}^{d \times n}$ such that,

$$1 - \epsilon \leq \frac{\|\mathbf{S}(\mathbf{x}_i - \mathbf{x}_j)\|_2}{\|\mathbf{x}_i - \mathbf{x}_j\|_2} \leq 1 + \epsilon.$$

Define the set of points,

$$V := \left\{ \frac{\mathbf{x}_i - \mathbf{x}_j}{\|\mathbf{x}_i - \mathbf{x}_j\|_2} \right\}_{i < j}, \quad |V| \sim N^2.$$

This is a finite point set, so $w(V) \sim \sqrt{\log |V|} \sim \sqrt{N}$.

Then from the previous slide, we know choosing

$$d \sim \frac{w(V)^2}{\epsilon^2} \sim (\log N)/\epsilon^2$$

achieves an ϵ -embedding with high probability.

Application 2: Subspace embeddings

We can now immediately craft subspace embeddings: Let V be a k -dimensional subspace of \mathbb{R}^n . We wish to identify $\mathbf{S} \in \mathbb{R}^{d \times n}$ such that,

$$1 - \epsilon \leq \frac{\|\mathbf{S}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq 1 + \epsilon$$



$$\sigma_{\min}(\mathbf{S}; V) \leq 1 - \epsilon, \sigma_{\max}(\mathbf{S}; V) \geq 1 + \epsilon.$$

We know that the Gaussian width $w(V \cap S^{n-1}) \leq \sqrt{k}$, so taking

$$d \gtrsim \frac{w(V \cap S^{n-1})^2}{\epsilon^2} = k\epsilon^2,$$

and choosing \mathbf{S} as a Gaussian embedding does the trick.

Structured Embeddings

The Gaussian strategy for constructing embeddings is useful as a benchmark. It has some weaknesses:

- it's a dense matrix
- naive matrix-vector multiplications cost $\mathcal{O}(dn)$ complexity per vector.

Of course, Gaussian embeddings are a great barometer as they provide essentially optimal complexity (in terms of d vs ϵ), and they are *oblivious*, i.e., they don't depend on the set V .

There are two classes of structured embeddings that directly address some of the weaknesses mentioned above:

- randomized sign matrices
- subsampled randomized Fourier transforms

Analyzing subspace embeddings, I

We will be interested in subspace embeddings, so we specialize to that situation now. With V is a k -dimensional subspace of \mathbb{R}^n , let $\mathbf{U} \in \mathbb{R}^{n \times k}$ be any semi-unitary matrix spanning V :

$$\mathbf{U}^* \mathbf{U} = \mathbf{I}_k, \quad \text{range}(\mathbf{U}) = V.$$

In order to accomplish an ϵ -embedding, we want,

$$1 - \epsilon \leq \frac{\|\mathbf{S}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq 1 + \epsilon \quad \forall \mathbf{x} \in V \setminus \{\mathbf{0}\}.$$

This is equivalent to requiring,

$$1 - \epsilon \leq \frac{\|\mathbf{S}\mathbf{U}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq 1 + \epsilon, \quad \forall \mathbf{y} \in \mathbb{R}^k.$$

Hence, we seek to study the extremal singular values of $\mathbf{S}\mathbf{U}$.

Analyzing subspace embeddings, II

The above is equivalent to studying the extremal eigenvalues of,

$$\mathbf{Y} := (\mathbf{S}\mathbf{U})^* \mathbf{S}\mathbf{U} = \mathbf{U}^* \mathbf{S}^* \mathbf{S}\mathbf{U}.$$

We wish to design \mathbf{S} so that,

$$\lambda_{\min}(\mathbf{Y}) \geq (1 - \epsilon)^2, \quad \lambda_{\max}(\mathbf{Y}) \leq (1 + \epsilon)^2.$$

Suppose that we can write $\mathbf{S}^* \mathbf{S}$ as a sum of iid matrices:

$$\mathbf{S}^* \mathbf{S} = \sum_{j \in [d]} \mathbf{X}_j, \quad \mathbf{X}_j \geq \mathbf{0}.$$

If this were the case, we'd be able to use our matrix concentration inequalities, in particular Matrix Chernoff....

Sparse sign matrices

Consider the following *sparse sign matrix* model: $\mathbf{S} \in \mathbb{R}^{n \times d}$ is a matrix with iid entries, having distribution:

$$(\mathbf{S})_{i,j} = \begin{cases} \alpha, & \text{with probability } p/2 \\ -\alpha, & \text{with probability } p/2 \\ 0, & \text{otherwise} \end{cases}$$

where $p \in (0, 1)$ is a free parameter, and α is chosen to maintain isotropy: $p d \alpha^2 = 1$.

The value of p controls sparsity: larger p yields denser matrices.

Since each row of \mathbf{S} is independent, then letting \mathbf{s}_j denote the j th row of \mathbf{S} , we have,

$$\mathbf{Y} = \mathbf{U}^* \mathbf{S}^* \mathbf{S} \mathbf{U} = \sum_{j \in [d]} \mathbf{U}^* \mathbf{s}_j \mathbf{s}_j^* \mathbf{U} =: \sum_{j \in [d]} \mathbf{X}_j,$$

where each \mathbf{X}_j is independent.

It's possible to compute somewhat sophisticated bounds on $\|\mathbf{X}_j\|_2$, yielding to an application of Matrix Chernoff:

If $d \gtrsim (k + \log n) \log k$, and $p \gtrsim (\log k)/d$, then the sparse sign matrix model is a $\frac{1}{2}$ -distortion subspace embedding with high probability.

Subsampled Randomized Fourier transforms

Another “structured” random matrix yields similar guarantees. Let $\mathbf{F} \in \mathbb{C}^{n \times n}$ denote the discrete Fourier transform matrix, which is unitary.

We construct \mathbf{S} as,

$$\mathbf{S} = \frac{1}{\sqrt{d}} \mathbf{R} \mathbf{F} \mathbf{\Sigma},$$

where:

- $\mathbf{\Sigma}$ is a square, diagonal matrix with iid Bernoulli(1/2) entries.
- $\mathbf{R} \in \mathbb{R}^{d \times n}$ is a random restriction matrix (its rows are \mathbf{e}_{q_j} , where $q_j, j \in [d]$ are iid uniform random variables on $[n]$).

The advantage/motivation: applying \mathbf{S} is quite cheap, requiring $\mathcal{O}(n \log d)$ complexity.

This random transform $\mathbf{F} \mathbf{\Sigma}$ attempts to construct output vectors whose entries have similar magnitudes.

Subsampled Randomized Fourier transforms

Like the sparse sign model, the subsampled randomized Fourier transform can be written as,

$$\mathbf{Y} = \mathbf{U}^* \mathbf{S}^* \mathbf{S} \mathbf{U} = \frac{1}{d} \sum_{j \in [d]} \mathbf{w}_j \mathbf{w}_j^*,$$

The following bound also holds:

$$\|\mathbf{X}_j\| \leq \frac{1}{d} \|\mathbf{w}_j\|_2^2 = \frac{1}{d} \|\mathbf{e}_{q_j}^* \mathbf{F} \mathbf{\Sigma} \mathbf{U}\|_2^2 \leq \frac{k + c \log n}{d}.$$

This last step requires proving a bound on the last vector norm, which requires Hoeffding's inequality.

With all of this, we have that if $d \gtrsim (k + \log n) \log k$, then the subsampled randomized Fourier transform is a subspace embedding with distortion $1/2$ with high probability.