1 Overview

In this lecture we will construct fast J-L embeddings via BOS RIP matrices, and then use them to quickly solve overdetermined least-squares problems.

2 A Fast J-L Embedding Matrix

- We choose a BOS with $K = 1$, $D = [N]$, and $\phi_\omega(t) = e^{-2\pi i (t-1)(\omega-1)/N}$, for all $t, \omega \in [N]$. Then, 
  \[ \Phi = \{ \phi_1, \ldots, \phi_N \} \]
  is a BOS, w.r.t. the uniform discrete probability measure $\nu$.

- We construct a random sampling matrix with entries 
  \[ \tilde{F}_{l,\omega} := \frac{1}{\sqrt{m}} \phi_\omega(l) = \frac{1}{\sqrt{m}} e^{-2\pi i (t-1)(\omega-1)/N}, \]
  for all $\omega \in [N]$, and $l \in S$, where $|S| = m$ is a set of random rows from the full DFT matrix. That is, we randomly select $m$ rows independently from a DFT matrix according to $\nu$ (i.e., uniformly selecting).

- Theorem 1 from Lecture 19 tells us that $\tilde{F}$ will have $\varepsilon_2(\tilde{F}) \leq \varepsilon/4$ for any chosen $p, \varepsilon \in (0,1)$ and integers $M \geq k \geq 16 \ln \left( \frac{4M}{1-p} \right)$ with probability $\geq 1 - N^{-\ln^4 N}$, provided that $m \geq \tilde{C}k \ln^4 N$. Here, $\tilde{C}$ is universal constant.

- Form a diagonal random matrix, $D \in \mathbb{R}^{N \times N}$, with $\pm 1$ on the diagonal, each with probability 1/2:
  \[ D_{ii} = \begin{cases} 
  1, & \text{with prob. } \frac{1}{2} \\
  -1, & \text{with prob. } \frac{1}{2}
  \end{cases} \quad (1) \]

- Theorem 3 from Lecture 16 now tells us that $\tilde{F}D \in \mathbb{C}^{m \times N}$ will be a strict J-L embedding for any arbitrary set $P \subseteq \mathbb{R}^N$ having cardinality $|P| \leq M$ with probability $\geq p - N^{-\ln^3 N}$, provided that $m \geq \tilde{C}' \ln(\frac{4M}{1-p}) \ln^4 N$. Here $\tilde{C}'$ is an absolute constant.

**Theorem 1.** Let $P \subseteq \mathbb{R}^N$ have $|P| \leq M$, and $p, \varepsilon \in (0,1)$. Form $\tilde{F}D \in \mathbb{C}^{m \times N}$ as above. Then,
  \[ (1 - \varepsilon)||\bar{x}||_2^2 \leq ||\tilde{F}D\bar{x}||_2^2 \leq (1 + \varepsilon)||\bar{x}||_2^2, \]
  with hold for all $\bar{x} \in P$ with probability at least $p - N^{-\ln^3 N}$, provided that $\tilde{F}D$ has at least $m = \frac{\tilde{C}'}{\varepsilon^2} \ln(\frac{4M}{1-p}) \ln^4 N$ rows. Here $\tilde{C}'$ is a universal constant.
Proof: Follows from the argument above.

– Note that \( \tilde{F} \in \mathbb{C}^{m \times N} \) has a fast matrix-vector multiply, which is the whole point...

To compute \( \tilde{F}D\tilde{x} \) we can:

• Computer \( D\tilde{x} \) in \( O(N) \) multiplies.
• Take the DFT of \( D\tilde{x} \) with the FFT in \( O((N \log N)\)-operations

So \( \tilde{F}D \) has an \( O(N \log N) \) matrix-vector multiply!

3  The Overdetermined Least Squares Problem [1]

Compute

\[
\hat{y}_{\text{min}} := \arg \min_{\tilde{x} \in \mathbb{R}^n} \| A\tilde{x} - \tilde{b} \|,
\]

for \( A \in \mathbb{C}^{N \times n}, N \gg n, \) and \( \tilde{b} \in \mathbb{C}^{N}. \)

Standard deterministic solution approaches (e.g., via the QR-decomposition) use \( O(Nn^2) \) operations.

If \( n \leq N \) are both large, we want to solve this faster.

4  A Randomized Algorithm for Solving the Problem

Theorem 2. There exists a universal constant \( \tilde{C} \in \mathbb{R}^{+} \) such that a fast J-L embedding matrix \( \tilde{F} \in \mathbb{C}^{m \times N} \), with \( m = \tilde{C}(n + 1) \ln \left( \frac{33}{2n + \sqrt{1 - p}} \right) \ln^4 N \) rows, will satisfy

\[
\frac{1}{2} \| A\tilde{y} - \tilde{b} \|_2 \leq \| \tilde{F}DA\tilde{y} - \tilde{F}D\tilde{b} \|_2 \leq \frac{3}{2} \| A\tilde{y} - \tilde{b} \|_2,
\]

for all \( \tilde{y} \in \mathbb{R}^n \), with probability at least \( p - N^{-\ln^3 N} \).

Let

\[
\hat{y}_{\text{min}} := \arg \min_{\tilde{x} \in \mathbb{R}^n} \| \tilde{F}D(A\tilde{x} - \tilde{b}) \|_2.
\]

If Theorem 2 holds we have that

\[
\frac{1}{2} \| A\hat{y}_{\text{min}} - \tilde{b} \|_2 \leq \| \tilde{F}D(A\hat{y}_{\text{min}} - \tilde{b}) \|_2 \leq \| \tilde{F}D(A\hat{y}_{\text{min}} - \tilde{b}) \|_2 \leq \frac{3}{2} \| A\hat{y}_{\text{min}} - \tilde{b} \|_2.
\]

Therefore, \( \| A\hat{y}_{\text{min}} - \tilde{b} \|_2 \leq 3 \| A\hat{y}_{\text{min}} - \tilde{b} \|_2 \). This implies that \( \hat{y}_{\text{min}} \) is a decent approximation to the optimal solution \( \tilde{y}_{\text{min}} \).

– The computational cost of computing \( \hat{y}_{\text{min}} \) is:
1. Computing \( \tilde{FDA} \) and \( \tilde{FD\vec{b}} \) takes \( O(nN \log N) \)-time, using the FFT.

2. Solving for \( \vec{y}_{\min} \) takes \( O(mn^2) \) operations (e.g., via the QR-decomposition).

The total running time is \( O(nN \log(N) + n^3 \ln\left(\frac{1}{12n + 2 \sqrt{1 - p}}\right) \ln^4 N) \).

- If \( n = \Theta(\sqrt{N}) \), and \( p \) is considered at constant, the deterministic method takes \( O(N^2) \)-operations, while the randomized approach takes \( O(N^{1.5} \log^4 N) \)-operations. This is a clear improvement when \( N \) is large.

---

**Proof of Theorem 2:** Let \( \vec{a}_j \in \mathbb{R}^N \) be the \( j \)th column of \( A \). Consider the subspace \( S := \text{span}\{\vec{a}_1, \ldots, \vec{a}_n, \vec{b}\} \).

- \( S \) is \((n + 1)\)-dimensional subspace \( \subset \mathbb{C}^N \). The unit ball \( B \) in \( S \) is isomorphic to the unit ball in \( \mathbb{R}^{2n+2} \). Thus, \( C_{\varepsilon/8}(B) \leq (1 + 16/\varepsilon)^{2n+2} \) by Lemma 2 in Lecture 14.

- Apply the proof of Lemma 3 in Lecture 14 (subspace embedding) to strictly embed \( S \) with \( \tilde{FDA} \), setting \( \varepsilon = \frac{1}{2} \). Theorem 1 above guarantees that \( \tilde{FD} \) will embed \( B \) with high probability, etc.

- Note: Theorem 2 is only useful in practice if \( \tilde{FDA} \) is about as well conditioned as \( A \) was in the first place! If \( \tilde{v}_j \) is the \( j^{th} \) right singular vector of \( \tilde{FDA} \) we can see that

\[
\frac{\sigma_n(A)}{2} = \frac{\|A\tilde{v}_n\|_2}{2} \leq \frac{\|A\tilde{v}_n\|_2}{2} \leq \|\tilde{FDA}\tilde{v}_n\|_2 = \sigma_n(\tilde{FDA}) \leq \sigma_1(\tilde{FDA}) = \|\tilde{FDA}\tilde{v}_1\|_2 \leq \frac{3}{2} \|A\tilde{v}_1\|_2 \leq \frac{3}{2} \sigma_1(A).
\]

Thus, \( \kappa(\tilde{FDA}) \leq 3\kappa(A) \).
Reference [1] notes that one can use a pre-conditioner for $\tilde{F}DA$ to quickly construct a pre-conditioner for $A$. We can then boost relative accuracy from 3 to $\varepsilon$ in $O(\log(1/\varepsilon))$ steps of a pre-conditioned conjugate gradient method (see [1] for more info.).

References