Recitation #2: HW #1 from 2017 Review

TA: Rajesh Jayaram

September 13, 2019

High Probability Matrix Product and Embeddings

Recall: Given a $A \in \mathbb{R}^{n \times d}$, a random matrix $S \in \mathbb{R}^{k \times n}$ is said to be a ϵ-subspace embedding (SE) if with probability at least $2/3$, for every $x \in \mathbb{R}^d$ we have $\|SAx\|_2^2 = (1 \pm \epsilon)\|Ax\|_2^2$.

- One desirable property of this is that for some applications, instead of storing all of A, we can store the more compact matrix $SA \in \mathbb{R}^{k \times d}$.

Now suppose we want the SE guarantee to hold with much higher probability, say $1 - \delta$. We could increase k, but usually $k \propto 1/\delta$.

High Probability Matrix Product and Embeddings

Recall: Given a $A \in \mathbb{R}^{n \times d}$, a random matrix $S \in \mathbb{R}^{k \times n}$ is said to be a ϵ-subspace embedding (SE) if with probability at least $2/3$, for every $x \in \mathbb{R}^d$ we have $\|SAx\|_2^2 = (1 \pm \epsilon)\|Ax\|_2^2$.

- One desirable property of this is that for some applications, instead of storing all of A, we can store the more compact matrix $SA \in \mathbb{R}^{k \times d}$.

Now suppose we want the SE guarantee to hold with much higher probability, say $1 - \delta$. We could increase k, but usually $k \propto 1/\delta$.

- Instead, suppose we want the storage of our algorithm to be $\propto \log(1/\delta)$.

High Probability Matrix Product and Embeddings

Recall: Given a $A \in \mathbb{R}^{n \times d}$, a random matrix $S \in \mathbb{R}^{k \times n}$ is said to be a ϵ-subspace embedding (SE) if with probability at least $2/3$, for every $x \in \mathbb{R}^d$ we have $\|SAx\|_2^2 = (1 \pm \epsilon)\|Ax\|_2^2$.

- One desirable property of this is that for some applications, instead of storing all of A, we can store the more compact matrix $SA \in \mathbb{R}^{k \times d}$.

Now suppose we want the SE guarantee to hold with much higher probability, say $1 - \delta$. We could increase k, but usually $k \propto 1/\delta$.

- Instead, suppose we want the storage of our algorithm to be $\propto \log(1/\delta)$.
- Show how given $S^1 A, S^2 A, \ldots, S^t A$ for $t = \Theta(\log(1/\delta))$, we can find a $i \in [t]$ where S^i is a SE for A.

High Probability Matrix Product and Embeddings

Fist step:

- Each $S^i A$ is a SE independently with prob $2/3$.
- By Hoeffding inequality, with prob $1 - e^{-\Omega(t)} = 1 - \delta$, there is a set $T \subseteq [t]$ with $|T| \geq \frac{3t}{5}$ such that $S^i A$ is a SE for all $i \in T$.

So we have many subspace embeddings among

$$S^1 A,\ S^2 A,\ \ldots, S^t A$$

Unfortunately, looking only at a single $S^i A$, impossible to check if S^i is a subspace embedding.

High Probability Matrix Product and Embeddings

Problem rephrased: Given:

$$S^1A,\ S^2A,\ \ldots, S^tA$$

Such that at least a $3/5$ fraction of the S^i's are SE's, find one that is a SE.

High Probability Matrix Product and Embeddings

Problem rephrased: Given:

$$S^1 A,\ S^2 A,\ \ldots, S^t A$$

Such that at least a $3/5$ fraction of the S^i's are SE's, find one that is a SE.

- We have a set $T \subseteq [t]$ so that for $i \in T$ and all x:

  $$\|S^i Ax\|_2^2 = (1 \pm \epsilon)\|Ax\|_2^2$$

  and we want to find a $i \in T$.

High Probability Matrix Product and Embeddings

Main Idea: for $i \in T$,the $S^i A$'s are all *close* to A, thus if $i, j \in T$, $S^i A$ should also be *close* to $S^j A$.

- If $i \in T$:

$$\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$$

  for all $j \in T$

High Probability Matrix Product and Embeddings

Main Idea: for $i \in T$,the $S^i A$'s are all *close* to A, thus if $i, j \in T$, $S^i A$ should also be *close* to $S^j A$.

- If $i \in T$:
$$\|S^i A x\|_2^2 = (1 \pm \epsilon)^2 \|S^j A x\|_2^2$$
for all $j \in T$
- If for some $j \in T$ we have $\|S^i A x\|_2^2 = (1 \pm \epsilon)^2 \|S^j A x\|_2^2$ for all $x \in \mathbb{R}^d$, then S^i is a $O(\epsilon)$-SE

High Probability Matrix Product and Embeddings

Main Idea: for $i \in T$,the $S^i A$'s are all *close* to A, thus if $i, j \in T$, $S^i A$ should also be *close* to $S^j A$.

- If $i \in T$:
$$\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$$
for all $j \in T$
- If for some $j \in T$ we have $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all $x \in \mathbb{R}^d$, then S^i is a $O(\epsilon)$-SE

Algorithm: for each $i \in [t]$, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x for at least a $3/5$-fraction of the $S^j A$'s.

- If $S^i A$ satisfies the above property, then $S^i A$ is a $O(\epsilon)$-SE (why?).

High Probability Matrix Product and Embeddings

Algorithm: for each $i \in [t]$, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x for at least a $3/5$-fraction of the $S^j A$'s.

- **How to check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x?? Ideas?**

High Probability Matrix Product and Embeddings

Algorithm: for each $i \in [t]$, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x for at least a $3/5$-fraction of the $S^j A$'s.

- **How to check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x?? Ideas?**
  - **Option** 1, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x in an ϵ-net with $\epsilon = \Theta(\frac{1}{\max_k \|S^k A\|_2})$ (takes exponential time).

High Probability Matrix Product and Embeddings

Algorithm: for each $i \in [t]$, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x for at least a $3/5$-fraction of the $S^j A$'s.

- **How to check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x?? Ideas?**
  - **Option** 1, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x in an ϵ-net with $\epsilon = \Theta(\frac{1}{\max_k \|S^k A\|_2})$ (takes exponential time).
  - **Option** 2, If $\|S^i Ax\|_2^2 = \|S^j Ax\|_2^2 \pm 2\epsilon$ for all unit x, then

$$\begin{aligned} &\iff |x^T (S^i A)^T (S^i A) x - x^T (S^j A)^T (S^j A) x| \leq 2\epsilon \\ &\iff \|(S^i A)^T (S^i A) - (S^j A)^T (S^j A)\|_2 \leq 2\epsilon \end{aligned} \quad (1)$$

High Probability Matrix Product and Embeddings

Algorithm: for each $i \in [t]$, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x for at least a $3/5$-fraction of the $S^j A$'s.

- **How to check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x?? Ideas?**
  - **Option** 1, check if $\|S^i Ax\|_2^2 = (1 \pm \epsilon)^2 \|S^j Ax\|_2^2$ for all x in an ϵ-net with $\epsilon = \Theta(\frac{1}{\max_k \|S^k A\|_2})$ (takes exponential time).
  - **Option** 2, If $\|S^i Ax\|_2^2 = \|S^j Ax\|_2^2 \pm 2\epsilon$ for all unit x, then

$$\begin{aligned} &\iff |x^T (S^i A)^T (S^i A) x - x^T (S^j A)^T (S^j A) x| \leq 2\epsilon \\ &\iff \|(S^i A)^T (S^i A) - (S^j A)^T (S^j A)\|_2 \leq 2\epsilon \end{aligned} \tag{2}$$

So just check if $\|(S^i A)^T (S^i A) - (S^j A)^T (S^j A)\|_2 \leq 2\epsilon$ (note that the LHS is just the max singular value of $(S^i A)^T (S^i A) - (S^j A)^T (S^j A)$, so can compute with SVD).

High Probability Matrix Product and Embeddings

Option 2 (more detail) We said

$$\begin{aligned} \max_{x,\|x\|_2=1} |x^T(S^iA)^T(S^iA)x - x^T(S^jA)^T(S^jA)x| \leq 2\epsilon \\ \iff \|(S^iA)^T(S^iA) - (S^jA)^T(S^jA)\|_2 \leq 2\epsilon \end{aligned} \quad (3)$$

High Probability Matrix Product and Embeddings

Option 2 (more detail) We said

$$\max_{x, \|x\|_2=1} |x^T(S^iA)^T(S^iA)x - x^T(S^jA)^T(S^jA)x| \leq 2\epsilon$$
$$\iff \|(S^iA)^T(S^iA) - (S^jA)^T(S^jA)\|_2 \leq 2\epsilon \tag{4}$$

Proof: since $(S^iA)^T(S^iA) - (S^jA)^T(S^jA)$ is symmetric, can write

$$(S^iA)^T(S^iA) - (S^jA)^T(S^jA) = (V\Sigma^{1/2})(\Sigma^{1/2}V^T) = U^TU$$

in SVD where $U = \Sigma^{1/2}V^T$. Then

$$\max_{\|x\|_2=1} |x^T\left((S^iA)^T(S^iA) - (S^jA)^T(S^jA)\right)x| = \max_{\|x\|_2=1} |x^TU^TUx|$$
$$= \max_{\|x\|_2=1} \|Ux\|_2^2$$
$$= \|U\|_2^2 \tag{5}$$
$$= \|U^TU\|_2$$
$$= \|(S^iA)^T(S^iA) - (S^jA)^T(S^jA)\|_2 \qquad \square$$

High Probability Matrix Product and Embeddings

Recap: Given:

$$S^1 A, \; S^2 A, \; \ldots, S^t A$$

Such that at least a $3/5$ fraction of the S^i's are a ϵ-SE, to find a S^i that is a $O(\epsilon)$-SE

- For each $i \in [t]$, check if $\|(S^i A)^T (S^i A) - (S^j A)^T (S^j A)\|_2 \leq 2\epsilon$ for at least a $3/5$-fraction of all $j \in [t]$. If so, i is a $O(\epsilon)$-SE.
- The condition that a $3/5$ fraction of the S^i's are a ϵ-SE holds with prob $1 - \delta$, which is what we wanted.

High Probability Matrix Product and Embeddings

A similar problem Given $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$, there are families of random matrices $S \in \mathbb{R}^{n \times r}$ such that

$$\Pr[\|ASS^T B - AB\|_F \leq \epsilon \|A\|_F \|B\|_F] \geq 2/3$$

High Probability Matrix Product and Embeddings

A similar problem Given $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$, there are families of random matrices $S \in \mathbb{R}^{n \times r}$ such that

$$\Pr[\|ASS^T B - AB\|_F \leq \epsilon \|A\|_F \|B\|_F] \geq 2/3$$

Given

$$\left(A(S^1)(S^1)^T B\right), \left(A(S^2)(S^2)^T B\right), \ldots, \left(A(S^t)(S^t)^T B\right)$$

where $S^1, \ldots, S^t$ are independently drawn from such a family with $t = \Theta(\log(1/\delta))$, find a $i \in [t]$ such that

$$\|A(S^i)(S^i)^T B - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$$

with prob $1 - \delta$.
Important: We don't know A, B, just the sketches $A(S^i)(S^i)^T B$.

High Probability Matrix Product and Embeddings

Approach Let $Z^i = A(S^i)(S^i)^T B$. We know that $\|Z^i - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$ for a $3/5$ fraction of $i \in [t]$ with probability $1 - \delta$.

High Probability Matrix Product and Embeddings

Approach Let $Z^i = A(S^i)(S^i)^T B$. We know that $\|Z^i - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$ for a $3/5$ fraction of $i \in [t]$ with probability $1 - \delta$.

If $\|Z^i - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$ and $\|Z^j - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$, then $\|Z^i - Z^j\|_F \leq 2\epsilon \|A\|_F \|B\|_F$.

- For each $i \in [t]$, if $\|Z^i - Z^j\|_F \leq 2\epsilon \|A\|_F \|B\|_F$ for a $3/5$ fraction of $j \in [t]$, then return Z^i.

High Probability Matrix Product and Embeddings

Approach Let $Z^i = A(S^i)(S^i)^T B$. We know that $\|Z^i - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$ for a $3/5$ fraction of $i \in [t]$ with probability $1 - \delta$.

If $\|Z^i - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$ and $\|Z^j - AB\|_F \leq \epsilon \|A\|_F \|B\|_F$, then $\|Z^i - Z^j\|_F \leq 2\epsilon \|A\|_F \|B\|_F$.

- For each $i \in [t]$, if $\|Z^i - Z^j\|_F \leq 2\epsilon \|A\|_F \|B\|_F$ for a $3/5$ fraction of $j \in [t]$, then return Z^i.
- **Problem, don't know** $\|A\|_F \|B\|_F$.

High Probability Matrix Product and Embeddings

Problem Want to check for each $i \in [t]$, if $\|Z^i - Z^j\|_F \leq 2\epsilon\|A\|_F\|B\|_F$ for a $3/5$ fraction of $j \in [t]$, but we don't know don't know $\|A\|_F\|B\|_F$.

High Probability Matrix Product and Embeddings

Problem Want to check for each $i \in [s]$, if $\|Z^i - Z^j\|_F \leq 2\epsilon \|A\|_F \|B\|_F$ for a $3/5$ fraction of $j \in [s]$, but we don't know don't know $\|A\|_F \|B\|_F$.

Instead:

- Let $c_i = \text{median}_{j \in [t]} \|Z^i - Z^j\|_F$.

High Probability Matrix Product and Embeddings

Problem Want to check for each $i \in [s]$, if $\|Z^i - Z^j\|_F \leq 2\epsilon\|A\|_F\|B\|_F$ for a $3/5$ fraction of $j \in [s]$, but we don't know don't know $\|A\|_F\|B\|_F$.

Instead:

- Let $c_i = \text{median}_{j \in [t]} \|Z^i - Z^j\|_F$.
- If $\|Z^i - AB\|_F^2 \leq \epsilon\|A\|_F\|B\|_F$, then $c_i \leq 2\epsilon\|A\|_F\|B\|_F$ (why)?

High Probability Matrix Product and Embeddings

Problem Want to check for each $i \in [s]$, if $\|Z^i - Z^j\|_F \leq 2\epsilon\|A\|_F\|B\|_F$ for a $3/5$ fraction of $j \in [s]$, but we don't know don't know $\|A\|_F\|B\|_F$.

Instead:

- Let $c_i = \text{median}_{j \in [t]} \|Z^i - Z^j\|_F$.
- If $\|Z^i - AB\|_F^2 \leq \epsilon\|A\|_F\|B\|_F$, then $c_i \leq 2\epsilon\|A\|_F\|B\|_F$ (why)?
- If $c_i \leq 2\epsilon\|A\|_F\|B\|_F$, then $\|Z^i - AB\|_F^2 \leq 3\epsilon\|A\|_F\|B\|_F$ (why)?

High Probability Matrix Product and Embeddings

Problem Want to check for each $i \in [s]$, if $\|Z^i - Z^j\|_F \leq 2\epsilon\|A\|_F\|B\|_F$ for a $3/5$ fraction of $j \in [s]$, but we don't know don't know $\|A\|_F\|B\|_F$.

Instead:

- Let $c_i = \text{median}_{j\in[t]}\|Z^i - Z^j\|_F$.
- If $\|Z^i - AB\|_F^2 \leq \epsilon\|A\|_F\|B\|_F$, then $c_i \leq 2\epsilon\|A\|_F\|B\|_F$ (why)?
- If $c_i \leq 2\epsilon\|A\|_F\|B\|_F$, then $\|Z^i - AB\|_F^2 \leq 3\epsilon\|A\|_F\|B\|_F$ (why)?
- **Solution:** return i with $i = \arg\min_j c_j$, then we know $\|Z^i - AB\|_F^2 \leq 3\epsilon\|A\|_F\|B\|_F$.

Count-Sketch Preserves Frobenius Norm

Problem:
Let $k = O(1/\epsilon^2)$, and $A \in \mathbb{R}^{n \times d}$. Show that if $S \in \mathbb{R}^{k \times n}$ is a count-sketch matrix, then

$$\Pr[\|SA\|_F^2 = (1 \pm \epsilon)\|A\|_F^2] > 3/4$$

Recall, $S \in \mathbb{R}^{k \times n}$ is a sparse matrix defined by hash functions $h : [n] \to [k], \sigma : [n] \to \{1, -1\}$. S has one non-zero entry per column i, placed in a random row $h(i) \in [k]$, given a random sign $\sigma(i)$.

$$S = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}$$

Count-Sketch Preserves Frobenius Norm

Problem:
Let $k = O(1/\epsilon^2)$, and $A \in \mathbb{R}^{n \times d}$. Show that if $S \in \mathbb{R}^{k \times n}$ is a count-sketch matrix, then

$$\Pr[\|SA\|_F^2 = (1 \pm \epsilon)\|A\|_F^2] > 3/4$$

Recall, $S \in \mathbb{R}^{k \times n}$ is a sparse matrix defined by hash functions $h : [n] \to [k], \sigma : [n] \to \{1, -1\}$. S has one non-zero entry per column i, placed in a random row $h(i) \in [k]$, given a random sign $\sigma(i)$.

$$S = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}$$

Approach: Compute expectation + variance + Chebyshev's.

Count-Sketch Expectation

$S \in \mathbb{R}^{k\times n}$, $A \in \mathbb{R}^{n\times d}$. We have:

$$
\begin{aligned}
\mathbb{E}[\|SA\|_F^2] &= \mathbb{E}[\sum_{i,j} \langle S_{i,*}, A_{*,j}\rangle^2] \\
&= \sum_{i,j} \mathbb{E}\left[\left(\sum_{t \text{ s.t. } h(t)=i} \sigma(t) A_{t,j}\right)^2\right] \\
&= \sum_{i,j} \mathbb{E}\left[\sum_{t \text{ s.t. } h(t)=i} \sigma(t)^2 A_{t,j}^2 + \sum_{t,t' \text{ s.t. } h(t)=i} \sigma(t) A_{t,j} \sigma(t') A_{t',j}\right] \\
&= \sum_{i,j} \mathbb{E}\left[\sum_{t \text{ s.t. } h(t)=i} \sigma(t)^2 A_{t,j}^2\right] \\
&= \sum_{i,j} \sum_{t \text{ s.t. } h(t)=i} A_{t,j}^2 = \|A\|_F^2
\end{aligned}
\tag{6}
$$

Variance: $\mathbb{E}[\|SA\|_F^4]$, Try it yourself!

The Kronecker Product

The **Kronecker Product** $\otimes$ is a operation on vectors (or matrices). Given two matrices $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{p \times q}$, the Kronecker product $A \otimes B \in \mathbb{R}^{mp \times nq}$ is given by

$$A \otimes B = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

The matrix $A \otimes B$ contains all $nmpq$ possible combinations of products $a_{i,j} b_{r,s}$. The Kronecker product is a special case of a *tensor*-product.

The Kronecker Product

The **Kronecker Product** $\otimes$ is a operation on vectors (or matrices). Given two matrices $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}$, the Kronecker product $A \otimes B \in \mathbb{R}^{mp\times nq}$ is given by

$$\begin{bmatrix}
a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\
a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\
\vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\
a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\
\vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\
\vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\
a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\
a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\
\vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\
a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq}
\end{bmatrix}$$

The Kronecker Product

What does this mean for vectors $x \in \mathbb{R}^n, y \in \mathbb{R}^m$:

$$x \otimes y = \begin{bmatrix} x_1 y \\ x_2 y \\ \vdots \\ x_n y \end{bmatrix} = \begin{bmatrix} x_1 y_1 \\ x_1 y_2 \\ \vdots \\ x_1 y_m \\ x_2 y_1 \\ \vdots \\ x_2 y_m \\ \vdots \\ x_n y_1 \\ \vdots \\ x_n y_m \end{bmatrix} \in \mathbb{R}^{nm}$$

Tensor-sketch

TensorSketch is a fast way of applying a Count-Sketch to Kronecker products. Let S^1, S^2 be count-sketches with functions $h^1, h^2 : [n] \to [s], \sigma^1, \sigma^2 : [n] \to \{1, -1\}$. We have $T \in \mathbb{R}^{s \times n^2}$, and for $x, y \in \mathbb{R}^n$:

$$(T(x \otimes y))_i = \sum_{j+k=i \pmod s} (S^1 x)_j (S^2 y)_k$$

Tensor-sketch

TensorSketch is a fast way of applying a Count-Sketch to Kronecker products. Let S^1, S^2 be count-sketches with functions $h^1, h^2 : [n] \to [s], \sigma^1, \sigma^2 : [n] \to \{1, -1\}$. We have $T \in \mathbb{R}^{s \times n^2}$, and for $x, y \in \mathbb{R}^n$:

$$(T(x \otimes y))_i = \sum_{j+k=i \pmod s} (S^1 x)_j (S^2 y)_k$$

Problem 4.1: Show that T is a count-sketch matrix with hash functions $H(j, k) = h^1(j) + h^2(k) \pmod s$ and $\sigma(j, k) = \sigma^1(j)\sigma^2(k)$. Here we think of (j, k) as indexing into a entry in $[n^2]$.

Tensor-sketch

Problem 4.1: Show that T is a count-sketch matrix with hash functions $H(j,k) = h^1(j) + h^2(k) \pmod{s}$ and $\sigma(j,k) = \sigma^1(j)\sigma^2(k)$. Here we think of (j,k) as indexing into a entry in $[n^2]$.
Recall: T is said to be a count-sketch matrix with hash functions H, σ if T has one non-zero entry per column $(j,k) \in [n^2]$, which is in row $H((j,k))$, and has value $\sigma((j,k))$.

This problem: Show that if

$$(T(x \otimes y))_i = \sum_{j+k=i \pmod{s}} (S^1 x)_j (S^2 y)_k$$

for every $x, y \in \mathbb{R}^n$, then T is a count-sketch with $H(j,k) = h^1(j) + h^2(k) \pmod{s}$ and $\sigma(j,k) = \sigma^1(j)\sigma^2(k)$.

Tensor-sketch

Hint: Consider $(e_j \otimes e_k) = e_{(j,k)}$, where $e_j, e_k \in \mathbb{R}^n$ and $e_{(j,k)} \in \mathbb{R}^{n^2}$ are standard basis vectors. Then $T(e_j \otimes e_j))$ is just the $(j,k) - th$ column of T: $T(e_j \otimes e_k) = T_{*,(j,k)}$. By definition

$$(T(e_i \otimes e_j))_i = \sum_{a+b=i \pmod s} (S^1 e_j)_a (S^2 e_k)_b$$

Note that $S^1 e_j = \sigma^1(j) \cdot e_{h^1(j)}$ and $S^2 e_k = \sigma^2(k) \cdot e_{h^2(k)}$, so

$$(T(e_i \otimes e_j))_i = \sum_{a+b=i \pmod s} (\sigma^1(j) \cdot e_{h^1(j)})_a (\sigma^2(k) \cdot e_{h^2(k)})_b$$

Note $(\sigma^1(j) \cdot e_{h^1(j)})_a (\sigma^2(k) \cdot e_{h^2(k)})_b = 0$ unless $a = h^1(j)$ and $b = h^2(k)$, so

$$(T(e_i \otimes e_j))_i = \begin{cases} \sigma^1(j)\sigma^2(k) & h^1(j) + h^2(k) = i \pmod s \\ 0 & \text{otherwise} \end{cases}$$

Tensor-sketch

Hint: We have shown:

$$(T(e_i \otimes e_j))_i = \begin{cases} \sigma^1(j)\sigma^2(k) & h^1(j) + h^2(k) = i \pmod{s} \\ 0 & \text{otherwise} \end{cases}$$

but since $(T(e_i \otimes e_j)) = T_{*,(j,k)}$, this tells us exactly what the entries of the (j,k)-th column should be. Moreover, this holds for every column (j,k).

Can you show that this column above satisfies the definition required for T to be a count-sketch with hash functions H, σ?