Problem Set 1 Solutions

September 25, 2019

1 Problem 1

We will adapt the analysis for the flattening lemma seen in class. Write $A = U\Sigma V^T$ for the SVD of A. We first show that the rows of U have norm bounded by $\kappa\sqrt{d/n}$, where $\kappa = \frac{\sigma_1(A)}{\sigma_d(A)}$ is the condition number of A. Write $A = U\Sigma V$ in it's SVD. Note that for any i:

$$\begin{aligned} \|e_i U\|_2^2 &\leq \|e_i U\Sigma\|_2^2/\sigma_d^2 \\ &= \|e_i U\Sigma V^T\|_2^2/\sigma_d^2 \\ &= \|e_i A\|_2^2/\sigma_d^2 \\ &= 1/\sigma_d^2 \end{aligned} \tag{1}$$

where the last equality holds because all rows of A have unit norm by assumption. Similarly, we have

$$\begin{aligned} \|e_i U\|_2^2 &\geq \|e_i U\Sigma\|_2^2/\sigma_1^2 \\ &= \|e_i U\Sigma V^T\|_2^2/\sigma_1^2 \\ &= \|e_i A\|_2^2/\sigma_1^2 \\ &= 1/\sigma_1^2 \end{aligned} \tag{2}$$

Since $\sum_{i=1}^n \|e_u U\|_2^2 = \|U\|_F^2 = d$ (because U has d orthonormal, unit norm columns), and since by the above inequalities we know the ratio between any two values $\|e_i U\|_2^2$ and $\|e_j U\|_2^2$ is at most $\frac{\sigma_1^2}{\sigma_d^2} = \kappa^2$, it follows that $\max_{i\in[n]} \|e_i U\|_2^2 \leq \kappa^2 \frac{d}{n}$. In other words, the largest squared row norm of U is at most $\kappa^2(d/n)$.

Now to prove that S is a subspace embedding for the column span of A, first recall that because the subspace embedding property must hold for all $x \in \mathbb{R}^d$, we can assume that the columns of A are orthonormal, namely: it suffices to prove that $\|SUx\|_2 = (1 \pm \epsilon)\|x\|_2$ for all $x \in \mathbb{R}^d$. We now closely follow the argument based on the Matrix Chernoff inequality seen in class. Let Y_i be the i-th sampled row of U, and let $X = I_d - nY_i^T Y$. Then the X_i's are independent, symmetric random matrices. Let the j-th row of U be denoted by U_j. We have

$$\mathbb{E}[X_i] = I_d - n\sum_{j=1}^n (\frac{1}{n}) U_j^T U_j = I_d - U^T U = 0^{d\times d}$$

and

$$\|X_i\|_2 \leq \|I_d\|_2 + n\max_{j\in[n]} \|e_j U\|_2^2 \leq 1 + n\kappa^2 \frac{d}{n} = \Theta(\kappa^2 d)$$

We now bound $\|\mathbb{E}[X_i^T X_i]\|_2$, noting that $X_i = I_d - nY_i^T Y_i$, so:

$$\begin{aligned}\mathbb{E}[X_i^T X_i + I_d] &= 2I_d - 2n\mathbb{E}[Y_i^T Y_i] + n^2\mathbb{E}[Y_i^T Y_i Y_i^T Y_i] \\ &= 2I_d - 2I_d + n^2 \sum_j \frac{1}{n} U_I^T U_i U_i^T U_i \\ &= n\sum_{i=1}^n U_i^T U_i \cdot \|U_i\|_2^2\end{aligned} \tag{3}$$

Let $Z = \mathbb{E}[X_i^T X_i + I_d]$, which is a symmetric matrix so $\|Z\|_2 = \max_{x\in\mathbb{R}^d, \|x\|_2=1} |x^T Z x|$, which we can bound by

$$\begin{aligned}|x^T Z x| &= n\sum_{i=1}^n x^T U_i^T U_i x \cdot \|U_i\|_2^2 \\ &= n\sum_{i=1}^n \langle x, U_i\rangle^2 \|U_i\|_2^2 \\ &\leq \kappa^2 d \sum_{i=1}^n \langle x, U_i\rangle^2 \\ &= \kappa^2 d x^T \left(\sum_{i=1}^n U_i^T U_i\right) x \\ &= \kappa^2 d (x^T I_d x) \\ &\leq \kappa^2 d \|I_d\|_2 \\ &\leq \kappa^2 d\end{aligned} \tag{4}$$

Thus $\|\mathbb{E}[X_i^T X_i]\|_2 \leq \|\mathbb{E}[X_i^T X_i] + I_d\|_2 + \|I_d\|_2 = \|\mathbb{E}[X_i^T X_i] + I_d\|_2 + 1$, which is $\|Z\|_2 + 1 \leq \kappa^2 d + 1$, from which it follows $\|\mathbb{E}[X_i^T X_i]\|_2 = O(\kappa^2 d)$. Now observe that $W = \frac{1}{s}\sum_{i=1}^s X_i = I_d - (SU)^T(SU)$. By the Matrix Chernoff inequality, have have $\mathbf{Pr}[\|W\|_2 > \epsilon] < 2d \cdot e^{-s\epsilon^2/(\sigma^2 + \gamma\epsilon/3)}$, where $\sigma^2 = \|\mathbb{E}[X_i^T X]\|_2 \leq \kappa^2 d$ and $\|X_i\|_2 \leq \gamma \leq \kappa^2 d$, thus

$$\mathbf{Pr}\Big[\|I_d - (SU)^T(SU)\|_2 \geq \epsilon\Big] \leq 2d \cdot e^{-s\epsilon^2/\Theta(\kappa^2 d)}$$

Thus, setting $s = O(\frac{1}{\epsilon^2}\kappa^2 d\log(d))$, we have that $\mathbf{Pr}[\|I_d - (SA)^T(SA)\|_2 \geq \epsilon] \leq 1/10$. Thus for every unit vector $x \in \mathbb{R}^d$, we have $\|x^T x - x^T(SU)^T(SU)x| \leq \epsilon$, which implies that $x^T(SU)^T(SU)x = \|SUx\|_2^2 = (1 \pm \epsilon)$ for all unit vectors $x \in \mathbb{R}^d$, which gives the desired subspace embedding property.

2 Problem 2

2.1 2.1

Fix any $\text{nnz}(A), \text{nnz}(B)$, and WLOG we have $\text{nnz}(A) \leq \text{nnz}(B)$. Let $A^0 \in \mathbb{R}^{m\times n}$ be the matrix with $A^0_{i,j} = 1$ for the first $\text{nnz}(A)$ values of (i,j) (in row major order, assuming $\text{nnz}(A) > n$). Define $B^0 \in \mathbb{R}^{n\times r}$ similarly, but in column major order. Let S_A, S_B be the set of non-zero entries of A^0, B^0, respectively. Since $|S_A| \leq |S_B|$, there is a natural surjection $h : S_A \to S_B$ that comes from sending $(i,j) \to (j,i)$ (note that if $(i,j) \in S_A$ then $(j,i) \in S_B$ by construction).

Now for each $(i,j) \in S_A$, let $A^{i,j} = A^0 + (X-1)e_i e_j^T$ (note that $e_i e_j^T$ is the matrix with every entry equal to 0 except for the (i,j)-position, where it is equal to 1), for some large value X we will

later choose. Similarly define $B^{i,j} = B^0 + (X-1)e_i e_j^T$ for $(i,j) \in S_B$. Define Let $Z^0 = A^0 B^0$ and $Z^{i,j} = A^{i,j} B^{h(i,j)} = A^{i,j} B^{j,i}$. Notice that $\|Z^0 - Z^{i,j}\|_F \geq X^2 - n$, thus any algorithm that outputs a matrix $G \in \mathbb{R}^{m \times r}$ with $\|G - Z^0\|_F \leq \frac{1}{2}\|A^0\|_F \|B^0\|_F \leq \frac{1}{2} n\sqrt{mr}$ must have $\|G\|_\infty < 2n\sqrt{mr}$. But if this is the case, then setting $Z = 100n\sqrt{mr}$, we have:

$$\begin{aligned} \|G - Z^{i,j}\|_F &\geq Z^2 - 2n\sqrt{mr} \\ &\geq \frac{9}{10} Z^2 \\ &> \frac{1}{2}(Z + \sqrt{nm})(Z + \sqrt{nr}) \\ &\geq \frac{1}{2}\|A^{i,j}\|_F \|B^{i,j}\|_F \end{aligned} \tag{5}$$

It follows that if G is a correct solution on input (A^0, B^0), then G is not a correct solution on $(A^{i,j}, B^{j,i})$ for any $(i,j) \in S_A$. Conversely, if $\|G - Z^{i,j}\|_F \leq \frac{1}{2}\|A^{i,j}\|_F \|B^0\|_F \leq (3/5)Z^2$, then $\|G\|_\infty \geq (2/5)Z^2$, from which it follows that $\|G - Z^0\|_F \geq (2/5)Z^2 - n \geq 100n^2 mr > \frac{1}{2}\|A^0\|_F \|B^0\|_F$. if G is a correct solution on some input $(A^{i,j}, B^{j,i})$ for any $(i,j) \in S_A$, then G is not a correct solution on (A^0, B^0). It follows that the output of any algorithm can be correct on at most one of the types of input: either (A^0, B^0) or $(A^{i,j}, B^{j,i})$ for some $(i,j) \in S_A$.

Now let $\mathcal{X} = (A^0, B^0) \cup \{(A^{i,j}, B^{j,i}) \mid (i,j) \in S_A\}$ be the set of inputs. Let $\mathcal{D}$ be a distribution over $\mathcal{X}$ that gives probability $1/2$ to the input (A^0, B^0), and with the remaining probability is uniform over all inputs in $\{(A^{i,j}, B^{j,i}) \mid (i,j) \in S_A\}$. Thus any algorithm that correctly solves the approximate matrix product problem with probability $2/3$ must also distinguish between whether the input from the distribution is (A^0, B^0) or in $\{(A^{i,j}, B^{j,i}) \mid (i,j) \in S_A\}$ with probability $2/3$. By Yao's minimax principle, if there is a randomized algorithm that for each input reads $o(\text{nnz}(A))$ entries in expectation and is correct with probability $2/3$ (over the random coin flips of the algorithm), then there is a deterministic algorithm that reads $o(\text{nnz}(A))$ entries of the input in expectation and is correct with probability $2/3$ (where now the probability and expectation is over the randomness used to draw an input from the distribution $\mathcal{D}$). By Marov's inequality, there is a deterministic algorithm that *always* reads $o(\text{nnz}(A))$ entries of the input and solves the problem with probability $7/12$ (over the randomness in the distribution $\mathcal{D}$). So assume that such a deterministic algorithm $\mathcal{A}$ exists.

Now note that the $\mathcal{A}$ can only distinguish between the two types of input (either (A^0, B^0) or $(A^{i,j}, B^{j,i})$ for some $i, j \in S_A$) if it reads an entry of the input which is larger than 1. Define the set H_A, H_B to be the set of entries read by a deterministic algorithm on the input (A^0, B^0) (so $H_A \subset S_A$ and $H_B \subset S_B$). Let $H = H_A \cup H_B$. Note that (i,j) can be any entry in S_A, but (j,i) is restricted to being in the first $\text{nnz}(A)$ entries of S_B. So the probability that $(i,j) \in H_A$ when (i,j) is drawn uniformly from S_A is $\frac{|H_A|}{|S_A|}$ and the probability that $(j,i) \in H_B$ when (j,i) is drawn uniformly from the first $|S_A|$ entries of S_B is at most $\frac{|H_B|}{|S_A|}$. By a union bound, the probability that $(i,j) \in H_A$ or $(j,i) \in H_B$ for a given $(i,j) \sim S_A$ chosen uniformly at random is at most $\frac{|H_A| + |H_B|}{|S_A|} = \frac{H}{\text{nnz}(A)} = o(1)$ by assumption. Moreover, if $(i,j) \notin H$ and $(j,i) \notin H_B$, then $\mathcal{A}$ cannot distinguish $(A^{i,j}, B^{j,i})$ from (A^0, B^0), and moreover the set of entries read by $\mathcal{A}$ on input $(A^{i,j}, B^{j,i})$ must be precisely H, since all entries of the input restricted to H are the same in both cases. Thus with probability $1 - o(1)$ over the distribution $\mathcal{D}$, the algorithm $\mathcal{A}$ only sees entries with value 1 in either case. It follows that the algorithm can be correct with probability at most $(1 - o(1))/2 < 7/12$, which is a contradiction.

2.2 2.2

We use the fact that if $S \in \mathbb{R}^{n \times s}$ is a matrix of i.i.d. $\mathcal{N}(0, 1/s)$ random variables, then with constant probability, if $S = O(L/\epsilon^2)$, S is a ϵ-subspace embedding for any fixed L-dimensional row span of vectors. Setting $L = m + r$, it follows that S is a ϵ-subspace embedding for the union of the row spans of A and B^T. Now note that for any matrix $Z \in \mathbb{R}^{n \times m}$, we can write $\|Z\|_2 = \max_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} x^T Z y / \|x\|_2 \|y\|_2$. For one direction, setting y to be a unit vector in the direction of the the largest right singular vector of Z and $x = Zy/\|Zy\|_2$, we have $xZy = \|Z\|_2$. For the other direction, note that $xZy \leq \langle \|x\|_2 \|Zy\|_2 \leq \|x\|_2 \|Z\|_2$ by Cauchy-Schwartz. Thus we can write

$$\|ASS^T B - AB\|_2 = \max_{x,y \text{ unit vectors}} |x^T ASS^T By - x^T ABy|$$

Now for any vectors x, y in the row spans of the union of A and B^T. We have that $\|(x-y)S\|_2^2 = \|xS\|_2^2 + \|yS\|_2^2 - 2\langle xS, yS \rangle$, and $\|x-y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\langle x, y \rangle$. But $\|(x-y)S\|_2^2 = \|x-y\|_2^2 \pm \epsilon \|x-y\|_2^2 = \|x-y\|_2^2 \pm O(\epsilon(\|x\|_2^2 + \|y\|_2^2))$, and $\|xS\|_2^2 + \|yS\|_2^2 = \|x\|_2^2 + \|y\|_2^2 \pm 2\epsilon(\|x\|_2^2 + \|y\|_2^2)$, from which it follows that $\langle x, y \rangle = \langle xS, yS \rangle \pm O(\epsilon(\|x\|_2^2 + \|y\|_2^2))$. Given this, fix any unit vectors x, y, and let $x' = xA$ and $y' = By$. Then

$$|(x')^T SS^T y' - (x')^T y'| = O(\epsilon(\|x'\|_2^2 + \|y'\|_2^2))$$

Thus

$$\begin{aligned} \max_{x,y \text{ unit vectors}} |x^T ASS^T By - x^T ABy| &= \max_{x,y \text{ unit vectors}} O(\epsilon(\|xA\|_2^2 + \|By\|_2^2))) \\ &\leq O(\epsilon)(\|A\|_2^2 + \|B\|_2^2) \end{aligned} \tag{6}$$

Thus

$$\|ASS^T B - AB\|_2 \leq O(\epsilon)(\|A\|_2^2 + \|B\|_2^2)$$

which is the desired result after a rescaling of ϵ.

3 Problem 3

3.1 3.1

Let $A \in \mathbb{R}^{n \times d}$ have the $d \times d$ identity matrix as the first d rows, and set the rest of the entries of A to 0. Let $x_{i,j} = 2(e_i + e_j)$ for all $i \neq j \in [d]$. Since S hashes each of the d columns of A into k buckets, we argue that if $S \in \mathbb{R}^{k \times n}$ has $k = o(d^2)$ rows, then with super-constant probability, at least two of the columns i, j of A will be hashed to the same row of S. To see this, note that by the birthday problem, after $b = \Theta(\sqrt{k})$ balls, with probability at least $1/2$ we will have at least one collision. Since we have $d = rb = \Omega(b)$ balls (for some $r = \Omega(1)$), after r repetitions of throwing b balls, the probability of a collision will be at least $1 - (1/2)^r = 1 - o(1)$. Now fix a colliding $i \neq j$. Then by the definition of count-sketch, either $\|SAx_{i,j}\|_2^2 = 0$ or $\|SAx_{i,j}\|_2^2 = (2+2)^2 = 16$, whereas $\|Ax_{i,j}\|_2^2 = 2^2 + 2^2 = 8$. In each case, the subspace embedding property fails to hold for this $x_{i,j}$.

3.2 3.2

Let A be the same matrix as above ($d \times d$ identity followed by $(n-d)$ rows of 0's). Let g_i be the Gaussian random variable in the i-th column of S, and observe that $\|SAe_i\|_2^2 = g_i^2$. Note note that the pdf of a Gaussian is lower bounded by $e^{-\frac{(1/2)^2}{2}}(2\pi)^{-1/2} = e^{-1/8}(2\pi)^{-1/2} > .35$ in the

interval $(-1/2, 1/2)$. it follows that $\mathbf{Pr}[|g_i| < 1/2] \geq (1/2 - (-1/2))e^{-1/8}(2\pi)^{-1/2} > .35$. Thus with probability at most .65, we have $\|SAe_i\|_2^2 \geq 1/4$ for a given i, and so the probability that $\|SAe_i\|_2^2 \geq 1/4$ for all $i = 1, 2, \ldots, d$ is at most $(.65)^d = \exp(-\Theta(d))$. Conversely, the probability that there is at least one $i \in \{1, 2, \ldots, d\}$ such that $\|SAe_i\|_2^2 < 1/4$ is at least $1 - \exp(-\Theta(d))$, so for such an i we have $\|SAe_i\|_2^2 < 1/4 < \frac{1}{2}\|Ae_i\|_2^2 = 1/2$, so with this probability S fails to be a $(1 \pm 1/2)$ subspace embedding.

4 Problem 4

4.1 4.1

For any $(i, j) \in [n] \times [n]$, identify (i, j) with a unique point in $[n^2]$. Then if T is a count-sketch with hash functions H, σ, then for any $x \in \mathbb{R}^{n^2}$, $(Tx)_i = \sum_{(j,k)\in[n]\times[n]} \sigma(j, k)x_{(j,k)}\delta_{H(j,k)=i}$, where $\delta_{H(j,k)=i}$ indicates the event that $H(j, k) = i$. Specifically, $T \in \mathbb{R}^{s\times n^2}$ is a count-sketch matrix if $T_{i,(j,k)} = \delta_{H(j,k)=i} \cdot \sigma(j, k)$.

Now by definition, for any $x, y \in \mathbb{R}^n$, we have:

$$\begin{aligned}
(T(x \times y))_i &= \sum_{j,k \text{ with } (j+k=i \pmod s)} (S^1x)_j \cdot (S^2y)_k \\
&= \sum_{j,k \text{ with } (j+k=i \pmod s)} \left(\sum_{p=1}^{n} \delta_{h^1(p)=j}\sigma^1(p)x_p\right)\left(\sum_{q=1}^{n} \delta_{h^2(q)=j}\sigma^2(q)y_q\right) \\
&= \sum_{j,k\in[n]} \left(\sum_{p=1}^{n} \delta_{h^1(p)=j}\sigma^1(p)x_p\right)\left(\sum_{q=1}^{n} \delta_{h^2(q)=j}\sigma^2(q)y_q\right)\delta_{(j+k=i \pmod s)} \\
&= \sum_{j,k\in[n]} \sum_{p,q\in[n]} \sigma^1(p)\sigma^2(q)(x \otimes y)_{(p,q)}\delta_{h^2(q)=j}\delta_{h^1(p)=j}\delta_{(j+k=i \pmod s)} \\
&= \sum_{p,q\in[n]} \sigma(p, q)(x \otimes y)_{(p,q)}\delta_{(h^1(p)+h^2(q)=i \pmod s)} \\
&= \sum_{p,q\in[n]} \sigma(p, q)(x \otimes y)_{(p,q)}\delta_{H(p,q)=i}
\end{aligned} \tag{7}$$

We now show that for any $i \in [s]$ and $j, k \in [n]$, if (i, k) indexes into $[n^2]$ in the natural way, then

$$T_{i,(j,k)} = \begin{cases} \sigma(j, k) & \text{if } H(j, k) = i \\ 0 & \text{otherwise} \end{cases}$$

To see this, note that $T(e_j \otimes e_k) = T_{*,(j,k)}$, where $e_j, e_k \in \mathbb{R}^n$ are standard basis vectors and $T_{*,(j,k)} \in \mathbb{R}^s$ is the (j, k)-th column of T. Moreover, by the above formula for the entries of $(T(x\times y))_i$ for $i \in [s]$, we conclude that $T_{*,(j,k)} = \sigma(j, k)\cdot e_{H(j,k)} \in \mathbb{R}^s$, and since this holds for every $(j, k) \in [n^2]$, this completes the proof.

4.2 4.2

We first claim that H is 3-wise independent. To show this, we begin by showing that H is not 4-wise independent. Given $x_1, x_2, y_1, y_2 \in [n]$, and given the values of $H(x_1, y_1), H(x_2, y_1), H(x_2, y_2)$, we

can compute the value of $H(x_1, y_2)$ via $H(x_1, y_2) = H(x_1, y_1) - H(x_2, y_1) + H(x_2, y_2) \pmod s$, which is simply

$$\begin{aligned} &= (h^1(x_1) + h^2(y_1)) - (h^1(x_2) + h^2(y_1)) + (h^1(x_2) + h^2(y_2)) \pmod s \\ &= h^1(x_1) - h^1(x_2) + h^1(x_2) + h^2(y_2) \pmod s \\ &= h^1(x_1) + h^2(y_2) \pmod s \\ &= H(x_1, y_2) \end{aligned} \tag{8}$$

Thus given 3 values of $H(x, y)$, one can compute a fourth exactly, so H is not 4-wise independent. Similarly with σ, we have $\sigma(x_1, y_1) \cdot \sigma(x_2, y_1) \cdot \sigma(x_2, y_2) = \sigma(x_1, y_2)$, so σ is less than 4-wise independent.

To see that H is at least 3-wise independent, fix any $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, and $i_1, i_2, i_3 \in [s]$, and consider the event $\mathcal{E}$ that $H(x_j, y_j) = i_j$ for all $j \in [3]$. First, note that if any of the (x_j, y_j) is such that both $x_j \notin \{x_{j'}\}_{j' \neq j}$ and $y_j \notin \{y_{j'}\}_{j' \neq j}$, then the entire term $H(x_j, y_j)$ is independent of the other two by the 4-wise independence of h^1, h^2, and can be considered separately. Thus WLOG this is not the case. It follows that there exactly 4 distinct values in the set $\{x_j\}_{j \in [3]} \cup \{y_j\}_{j \in [3]}$. There are two ways for this to occur: either $x_1 = x_2 = x_3$ and y_1, y_2, y_3 are distinct, or the set can be written as $(x_1, y_1), (x_2, y_1), (x_2, y_2)$ for some x_1, x_2, y_1, y_2.

For the first case and any fixing of $h^1(x^1)$ (using 4-wise independence, it doesn't matter which value of $h^1(x^1) \in [s]$ we condition on), the probability of the event $\mathcal{E}$ is the probability that $h^2(y_j) = i_j - h^1(x_1)$ for every $j \in [3]$, which is $\frac{1}{s^3}$ by the 4-universality of h^2. For the second case, we can fix any value of $h^1(x^1)$ (again, by 4-wise independence, it doesn't matter which value we condition on), the event $\mathcal{E}$ becomes the event that $h^2(y_1) = i_1 - h^1(x^1)$, $h^1(x^2) = i_2 - (i_1 - h_1(x^1))$ and $h^2(y_2) = i_3 - (i_2 - (i_1 - h_1(x^1)))$, where each value on the RHS of each of the three equalities is now a fixed value. By 4-universality of the hash functions h^1, h^2's, we have that this probability is $\frac{1}{s^3}$. Thus in each case, $\mathbf{Pr}[\mathcal{E}] = \frac{1}{s^3}$, implying that H is 3 universal.

To show that σ is 3-universal, one applies the same argument just presented for H, replacing the subtraction $-$ operation with multiplication, which completes the proof.

4.3 4.3

Let for $v, u \in \mathbb{R}^n$, let $p_v(x) = \sum_{i=1}^n \sigma^1(i) v_i x^{h^1(i)}$ and $q_u(x) = \sum_{i=1}^n \sigma^2(i) u_i x^{h^2(i)}$, which are formal polynomials of degree at most s. Note that for any $i \in [2s]$, we have that the coefficient of x^i in the polynomial $f_{v,u}(x) = p_v(x) q_u(x)$ is given by $c_i = \sum_{j,k,j+k=i} v_j u_k \sigma^1(j) \sigma^2(k)$. Thus for any $i \in [s]$, we have $c_i + c_{i+s} = \sum_{j,k \in [n], j+k=i \pmod s} v_j u_k \sigma^1(j) \sigma^2(k) = \sum_{j,k \in [n]} \delta_{H(j,k)=i} \sigma(j,k)(v \otimes u)_{j,k}$. Thus given the coefficients $c_1, c_2, \ldots, c_{2s}$ of $f_{v,u}(x) = \sum_{\ell=1}^{2s} c_\ell x^\ell$, we have that $(T(u \times v))_i = c_i + c_{i+s}$ can be computed in $O(1)$ time for each i, thus $T(u \times v)$ can be computed in $O(s)$ time given $c_1, c_2, \ldots, c_{2s}$. Now the coefficients of the polynomials $p_v(x), q_u(x)$ can be computed in $\text{nnz}(v)$ and $\text{nnz}(u)$ time respectively, at which point the coefficents c_ℓ of $f_{v,u}(x)$ can be computed in $O(s \log(s))$ time via the Fast Fourier Transform. Altogether, the overall runtime is $\text{nnz}(v)+\text{nnz}(u) + O(s \log(s))$ as needed.