15-859 Algorithms for Big Data — Fall 2017

Problem Set 1

Due: 15:00, Thursday, September 28

Please see the following link for collaboration and other homework policies:
`http://www.cs.cmu.edu/afs/cs/user/dwoodruf/www/teaching/15859-fall17/grading.pdf`

Problem 1: High Probability Matrix Product and Embeddings (12 points)

(1) (6 points) We saw for several random families $\mathcal{F}$ of matrices S, given an $m \times n$ matrix A and an $n \times p$ matrix B, if S is an $n \times r$ matrix with $r = \Theta(1/\epsilon^2)$ columns, then

$$\Pr_S[\|ASS^TB - AB\|_F \le \epsilon \|A\|_F \|B\|_F] \ge 2/3,$$

that is, S satisfies the *approximate matrix product* property. As an application, S provides a means of dimensionality reduction, that is, instead of storing A and B, we can store AS and S^TB. For example, if we see the entries of A and B in a data stream, it suffices to maintain AS and S^TB, which is very space-efficient.

One issue with the approach above is that the success probability is only $2/3$. Suppose now we independently sample $S^1, \ldots, S^\ell$, $\ell = \Theta(\log(1/\delta))$, from $\mathcal{F}$. Then with probability at least $1-\delta$, there exists an $i^* \in \{1, 2, \ldots, \ell\}$ for which $\|A(S^{i^*})(S^{i^*})^TB - AB\|_F \le \epsilon \|A\|_F \|B\|_F$. Suppose we maintain $A(S^i)(S^i)^TB$ for each $i \in \{1, 2, \ldots, \ell\}$ in a stream. Show how, given only $A(S^1)(S^1)^TB, \ldots, A(S^\ell)(S^\ell)^TB$, we can output an i^* such that with probability at least $1 - \delta$, $\|A(S^{i^*})(S^{i^*})^TB - AB\|_F \le \epsilon \|A\|_F \|B\|_F$.

(2) (6 points) A related problem is that of obtaining a high probability subspace embedding. Recall from class that a $k \times n$ random matrix S is said to be a $(1+\epsilon)$-approximate *subspace embedding* if for any fixed $n \times d$ matrix A, $n > d$, we have that with probability at least $2/3$ over our random choice of S (from some family $\mathcal{F}$ of matrices), simultaneously for all vectors $x \in \mathbb{R}^d$, it holds that $\|SAx\|_2 = (1 \pm \epsilon)\|Ax\|_2$.

In some applications it is desirable to obtain a $(1+\epsilon)$-approximate subspace embedding succeeding with probability at least $1-\delta$, for a small $\delta > 0$. While in some cases one can achieve this by increasing the number of rows of S by a small amount, here we show a general technique: given $\ell = \Theta(\log(1/\delta))$ sketches $S^1A, S^2A, \ldots, S^\ell A$, where each S^i is an independent $(1+\epsilon)$-approximate subspace embedding for A succeeding with probability $2/3$, show how to find an $i^* \in \{1, 2, \ldots, \ell\}$ for which $S^{i^*}A$ is a $(1+\Theta(\epsilon))$-approximate subspace embedding for A with probability at least $1 - \delta$. You may assume, for simplicity, that $\text{rank}(A) = d$. The high level solution is similar to the previous part for approximate matrix multiplication.

Problem 2: Linear Dependence on ϵ in Regression (13 points) In class we saw an application of subspace embeddings to overconstrained least squares regression: $\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$, where we chose a random matrix S, computed SA and Sb, and returned the solution $x' \in \mathbb{R}^d$ to $\min_{x \in \mathbb{R}^d} \|SAx - Sb\|_2$. We showed if S is a subspace embedding for the $(d+1)$-dimensional space $[A, b]$, given here as the column span of an $n \times (d+1)$ matrix, then x' satisfies $\|Ax' - b\|_2 \leq (1+\epsilon) \min_{x \in \mathbb{R}^d} \|Ax - b\|_2$ with probability at least $2/3$. We saw several instantiations of S, from Gaussian, to Subsampled Randomized Hadamard Transform, to CountSketch matrices. Our arguments each required S to have at least d/ϵ^2 rows. Here we show that S can have only $O(d/\epsilon)$ rows.

(1) (1 point) Let $U \in R^{n \times r}$ be an orthonormal basis for the column span of A, where $r = \text{rank}(A)$. Show if $x' = \text{argmin}_{x \in \mathbb{R}^r} \|SUx - Sb\|_2$ satisfies $\|Ux' - b\|_2 \leq (1+\epsilon) \min_x \|Ux - b\|_2$, then $y' = \text{argmin}_{y \in \mathbb{R}^d} \|SAy - Sb\|_2$ satisfies $\|Ay' - b\|_2 \leq (1+\epsilon) \min_y \|Ay - b\|_2$.

(2) (2 point) We will thus focus on the problem of showing that $x' = \text{argmin}_{x \in \mathbb{R}^r} \|SUx - Sb\|_2$ is such that $\|Ux' - b\|_2 \leq (1+\epsilon) \min_x \|Ux - b\|_2$ for an S with $O(d/\epsilon)$ rows. Let $x^* = \text{argmin}_x \|Ux - b\|_2$. Argue that $\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$.

(3) (10 points) Show that if S is a $k \times n$ matrix of i.i.d. zero-mean Gaussian random variables with variance $1/k$, where $k = O(d/\epsilon)$, then $\|U(x' - x^*)\|_2^2 = O(\epsilon)\|Ux^* - b\|_2^2$.

    You will need to use that (1) S is a $(1 \pm 1/2)$-approximate subspace embedding for any fixed d-dimensional subspace (e.g., the column space of a given matrix with d columns) and (2) S satisfies approximate matrix product. You will also need to use the particular form of x' and x^*, namely, $x' = (SU)^- Sb$ and $x^* = U^T b$, and it might be helpful to use that for a matrix C with linearly independent columns, $C^- = (C^T C)^{-1} C^T$.

Problem 3: CountSketch Preserves Frobenius Norm (10 points) Let $r = O(1/\epsilon^2)$. Show that for any fixed $n \times d$ matrix A, if S is a $d \times r$ CountSketch matrix (in this case S has a single, uniformly random position which is non-zero in each row, and that position is 1 with probability 1/2 and -1 with probability 1/2), then $\Pr[\|AS\|_F^2 = (1 \pm \epsilon)\|A\|_F^2] \geq 9/10$.

Problem 4: Sketching Structured Regression Problems (15 points) We saw that if $x' = \text{argmin}_x \|SAx - Sb\|_2$ is the solution to a sketched regression problem for a CountSketch matrix S which has $O(d^2/\epsilon^2)$ rows, then with probability at least 2/3 we have $\|Ax' - b\|_2 \leq (1+\epsilon) \min_x \|Ax - b\|_2$. Note that SA and Sb can be computed in $O(\text{nnz}(A) + n)$ time, and assuming each row of A contains a non-zero entry (otherwise we can throw away some rows of A and corresponding entries of b), this leads to an overall $\text{nnz}(A) + \text{poly}(d/\epsilon)$ time algorithm for regression.

(1) (7 points) We say that a randomized algorithm has property $\mathcal{P}$ for a family $\mathcal{F}$ of pairs (A, b), where $A \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^{n \times 1}$, if for every $(A, b) \in \mathcal{F}$, the algorithm succeeds in outputting an $x' \in \mathbb{R}^d$ for which $\|Ax' - b\|_2 \leq 2 \min_x \|Ax - b\|_2$ with probability at least 3/4, over the algorithm's random coin tosses. In this problem, assume for

simplicity, that $n \geq 2d$. Argue that for every possible value of m, $n \leq m \leq nd/2$, there exists a family $\mathcal{F}_m$ of pairs (A, b), where $\text{nnz}(A) = m$ for every $(A, b) \in \mathcal{F}_m$, such that any algorithm which has property $\mathcal{P}$ for $\mathcal{F}_m$ reads $\Omega(m)$ entries of A in expectation for some pair $(A, b) \in \mathcal{F}_m$. You can also assume for simplicity, that d divides m. Here the expectation is taken over the algorithm's random coin tosses. Note that it is important to be formal in proving a lower bound. In recitation we will go over Yao's minimax principle which is one way of formalizing this. If you cannot attend the recitation, the slides for it will also be posted online, or also feel free to come to office hours to ask for help.

(2) (8 points) Now we consider specific *structured* matrices A for which one can solve regression in time faster than $\text{nnz}(A)$. Let A be the $n \times d$ Vandermonde matrix, $n \geq d$, for which $A_{i,j} = i^{j-1}$, and let b be an arbitrary vector. Show how to obtain an $x' \in \mathbb{R}^d$ in time $n \cdot \text{poly}(\log n) + \text{poly}(d(\log n)/\epsilon)$ for which $\|Ax' - b\|_2 \leq (1 + \epsilon) \min_x \|Ax - b\|_2$ with probability at least $2/3$. Notice that the running time of this algorithm is *sublinear* in $\text{nnz}(A)$. You can use the fact that for an $r \times s$ Vandermonde matrix V, and an r-dimensional vector x, one can compute $x^T \cdot V$ deterministically in $(r + s) \cdot \text{poly}(\log(rs))$ time.