15-859 Algorithms for Big Data — Fall 2020

Problem Set 2

Due: Tuesday, October 20, noon ET

Please see the following link for collaboration and other homework policies: `http://www.cs.cmu.edu/afs/cs/user/dwoodruf/www/teaching/15859-fall20/grading.pdf`

Problem 1: High Precision Regression for Square Matrices (16 points)
Given a full rank $n \times d$ matrix A, with $n \geq d$, and a $d \times 1$ vector b, let $\kappa = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$ denote the condition number of A. Consider the regression problem:

$$\min_{x \in \mathbb{R}^d} \|A^T A x - b\|_2.$$

Show how to find a solution x' for which

$$\|A^T A x' - b\|_2 \leq \epsilon \|b\|_2$$

with probabilty $\geq 2/3$, in $\tilde{O}((\text{nnz}(A) + d^2)\log(\kappa/\epsilon) + \text{poly}(d))$ time, where for a function f, $\tilde{O}(f) = f \cdot \text{poly}(\log f)$.

HINT: As in class, think of computing a sketch SA of the input matrix A and computing a preconditioner R such that SAR has orthonormal columns. Then consider the problem $\min_{x \in \mathbb{R}^d} \|R^T A^T A R x - R^T b\|_2$.

Problem 2: Operator Norm Estimation From Sampling (17 points)
For a given $n \times d$ matrix A, with $n \geq d$, define its *stable rank* to be $\rho = \frac{\|A\|_F^2}{\|A\|_2^2}$. The stable rank is always at most the rank of A, which is always at most d.

Define a distribution $q = (q_1, \ldots, q_n)$ on the rows A_i of A such that

$$q_i \geq \beta \frac{\|A_i\|_2^2}{\|A\|_F^2},$$

where $0 < \beta < 1$ is some parameter. Let $r = \tilde{\Theta}(\frac{\rho}{\beta \epsilon^2} \cdot \ln d)$. Show that if we sample r rows, creating an $r \times n$ sampling and rescaling matrix S, where for each $i = 1, 2, \ldots, r$, we have $S_{i,j} = \frac{1}{\sqrt{r q_j}}$ if row j of A is sampled in the i-th iteration, and $S_{i,j} = 0$ otherwise, then

$$\|A^T A - A^T S^T S A\|_2 \leq \epsilon \|A\|_2^2,$$

where for a matrix A, $\|A\|_2 = \sup_x \frac{\|Ax\|_2}{\|x\|_2}$ is its operator norm.

HINT: Use the Matrix Chernoff bound from class. You do not need the flattening lemma here, though you need to be careful to obtain a bound that depends on the stable rank rather than the rank.

Problem 3: Existence Results for Robust Low Rank Approximation (17 points) For a matrix $A \in \mathbb{R}^{n \times d}$, the $\|A\|_{1,2}$ norm is equal to $\sum_{i=1}^{n} \|A_i\|_2$, i.e., the sum of Euclidean norms of its rows. One can verify that $\|A\|_{1,2}$ is actually a norm, meaning that it satisfies the defining properties of a norm.

In certain applications one would like a low rank approximation of A with respect to this norm, i.e., to find an $n \times k$ matrix U and a $k \times d$ matrix V which minimize

$$\|U \cdot V - A\|_{1,2},$$

over all such U and V. This norm is often considered more "robust" than the Frobenius norm, since here we do not square the Euclidean norm of the difference between rows of $U \cdot V$ and corresponding rows of A. Thus, if, for example, A starts off as a rank-k matrix and then we perturb one entry by a very large amount, we will pay less attention to that entry (often called an "outlier") since we do not square the difference.

Show that if $R \in \mathbb{R}^{d \times r}$ is a matrix of i.i.d gaussian entries with mean 0 and variance $1/r$ for $r = O((k + \log n)/\epsilon^2)$, then with probability at least $2/3$, the column span of $A \cdot R$ contains a $(1+\epsilon)$-approximate low rank approximation, meaning there exists a rank-k matrix Y for which

$$\|ARY - A\|_{1,2} \leq (1 + \epsilon) \min_{U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{k \times d}} \|UV - A\|_{1,2}.$$

HINT: The techniques for the existence argument we did in class for Frobenius norm low rank approximation may be helpful here.