CSE 392: Matrix and Tensor Algorithms for Data Spring 2024

Homework 2

Due Date: 03-08-2024

Assignments are to be submitted through Canvas, and should be individual work. You can discuss the problems, but should submit individually. Preferably typewritten.

Problem 1. Sign-JL and JL for Inner Products

Let $\boldsymbol{S} \in \mathbb{R}^{m\times d}$ be a *sign Johnson-Lindenstrauss matrix* with i.i.d. $\pm 1/\sqrt{m}$ entries. That is, each matrix entry is $\pm 1/\sqrt{m}$ with probability 1/2.

(i) Show that for any $\boldsymbol{x} \in \mathbb{R}^d$:

$$\mathbb{E}[\|\boldsymbol{Sx}\|_2^2] = \|\boldsymbol{x}\|_2^2 \quad \text{and} \quad \mathrm{Var}[\|\boldsymbol{Sx}\|_2^2] \leq \frac{2}{m}\|\boldsymbol{x}\|_2^4.$$

(ii) Then use this to prove that for $m = O\left(\frac{1}{\delta\epsilon^2}\right)$:

$$\Pr[|\|\boldsymbol{Sx}\|_2^2 - \|\boldsymbol{x}\|_2^2| \geq \epsilon\|\boldsymbol{x}\|_2^2] \leq \delta.$$

This proves the distributed JL lemma for sign matrices.

(iii) Generalize the result for approximating inner products between two vectors. For vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, prove that:

$$\Pr[|\langle \boldsymbol{Sx}, \boldsymbol{Sy}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle| \geq \epsilon\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2] \leq \delta,$$

if we choose $m = O\left(\frac{1}{\delta\epsilon^2}\right)$

Problem 2. Approximate Matrix Multiplcation for SRHT

Prove the approximate matrix product property for the Subsampled Randomized Hadamard Transform. Recall that $\boldsymbol{S} = \boldsymbol{P}\cdot\boldsymbol{H}\cdot\boldsymbol{D}$, and given $\boldsymbol{A} \in \mathbb{R}^{n\times d}$; $\boldsymbol{B} \in \mathbb{R}^{d\times n}$; $\epsilon, \delta > 0$, AMM says with probability at least $1-\delta$:

$$\|\boldsymbol{BS}^\top\boldsymbol{SA} - \boldsymbol{BA}\|_F^2 \leq \epsilon^2/n\|\boldsymbol{A}\|_F^2\|\boldsymbol{B}\|_F^2.$$

Show that this statement holds if $\boldsymbol{S}$ has $O(n\log^2(nd/\delta)/\epsilon^2)$ rows.

Hints: Use the fact that $\boldsymbol{P}$ is uniformly sampling and approximating the product between matrices $\boldsymbol{BD}^\top\boldsymbol{H}$ and $\boldsymbol{HDA}$, and use the variance argument we used for AMM with sampling in Lecture 6. You will also need the SRHT mixing lemma from Lecture 8.

Problem 3. LinearTimeSVD: ℓ_2- norm error bound

Given $\boldsymbol{A} \in \mathbb{R}^{n\times d}$, let $\boldsymbol{H}_k$ be computed using the LinearTimeSVD algorithm we saw in Lecture 11. Then, show that

$$\|\boldsymbol{A} - \boldsymbol{H}_k\boldsymbol{H}_k^\top\boldsymbol{A}\|_2^2 \leq \|\boldsymbol{A} - \boldsymbol{A}_k\|_2^2 + 2\|\boldsymbol{AA}^\top - \boldsymbol{CC}^\top\|_2$$

For this,

(i) Let $\mathcal{H}_k = span(\boldsymbol{H}_k)$ and $\mathcal{H}_{n-k}$ be the orthogonal complement. Then, show that

$$\|\boldsymbol{A} - \boldsymbol{H}_k\boldsymbol{H}_k^\top \boldsymbol{A}\|_2 \leq \max_{\boldsymbol{z}\in\mathcal{H}_{n-k}, \|\boldsymbol{z}\|=1} \|\boldsymbol{z}^\top \boldsymbol{A}\|_2.$$

Hint: Use the definition of the spectral norm, and the fact that any vector $\boldsymbol{x} \in \mathbb{R}^n$ can be decomposed as $\boldsymbol{x} = \alpha\boldsymbol{y} + \beta\boldsymbol{z}$, where $\boldsymbol{y} \in \mathcal{H}_k, \boldsymbol{z} \in \mathcal{H}_{n-k}$ and $\alpha^2 + \beta^2 = 1$.

(ii) Next show that,

$$\|\boldsymbol{z}^\top \boldsymbol{A}\|_2^2 \leq \|\boldsymbol{A} - \boldsymbol{A}_k\|_2^2 + 2\|\boldsymbol{A}\boldsymbol{A}^\top - \boldsymbol{C}\boldsymbol{C}^\top\|_2$$

Hint: You may need to use the Hoffman-Wielandt inequality, which says for any $\boldsymbol{A}, \boldsymbol{E} \in \mathbb{R}^{n\times d}$, we have $\max_t |\sigma_t(\boldsymbol{A} + \boldsymbol{E}) - \sigma_t(\boldsymbol{A})| \leq \|\boldsymbol{E}\|_2$.

(iii) Combine the above two to get the desired bound.

Problem 4. Rank-k approximation and trace

For a PSD matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, let $\boldsymbol{A}_k$ be the best rank-k approximation to $\boldsymbol{A}$. Then, show that

$$\|\boldsymbol{A} - \boldsymbol{A}_k\|_F \leq \frac{1}{2\sqrt{k}} \operatorname{Tr}(\boldsymbol{A})$$

Hints:

(a) Use the definitions of trace and Frobenius norm of a PSD matrix in terms of its eigenvalues.

(b) Consider the vectors, $\boldsymbol{v} = [\lambda_1, \dots, \lambda_n]$ (vector of eigenvalues of $\boldsymbol{A}$) and $\boldsymbol{v}_k = [\lambda_1, \dots, \lambda_k, 0, \dots, 0]$ (vector of eigenvalues of $\boldsymbol{A}_k$).

(c) Express the trace and the Frobenius norm above in terms of the norms of these vectors.

(d) Use the relation between 2-norm, 1-norm and ∞-norm of vectors.

Problem 5. Leverage scores and preconditioning

(i) Implement the (approximate) leverage scores estimation algorithm we studied in Lecture 10 using Gaussian sketching matrices. I.e., use Gaussian sketch for both $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. You can use the `arrhythmia-clean.mat` dataset for testing your script.

(ii) Replace Gaussian sketch with Countsketch. What do you observe?

(iii) Use part of the above leverage score estimation algorithm to build a preconditioned least squares solver (e.g., the iterative refinement algorithm we saw in Lecture 10). Compare its performance to standard least squares regression on the Arrhythmia data set (you used in HW1).

(iv) Recall that this dataset was ill-conditioned. Does preconditioning help?

Submit your scripts along with the assignment.