Homework 5
APPM 5650 Fall 2021
Randomized Algorithms

Due date: Monday, Sept 27 2021 at 10:20 AM
Theme: Randomized Sketches **Instructor**: Prof. Becker

Instructions Collaboration with your fellow students is allowed and in fact recommended, although direct copying is not allowed. Please write down the names of the students that you worked with. The internet is allowed for basic tasks only, not for directly looking for solutions.

An arbitrary subset of these questions will be graded.

Problem 1: [READING and MATH] Read exercise 4.1.4 and Lemma 4.1.5 about approximate identities, from Vershynin's 2018 "High-dimensional probability" book.

Deliverable: Do exercise 4.1.6: if $A \in \mathbb{R}^{m\times n}$ and all the singular values of A are in the set $[1-\delta, 1+\delta]$ for some $\delta > 0$, then $\|A^T A - I_n\| \le 3\max\left(\delta, \delta^2\right)$. Note: you may assume $m \ge n$, otherwise the hypothesis cannot be true unless $\delta \ge 1$ (why?) and the result is not interesting.

Problem 2: [PROGRAMMING] Try various sketches, do they converge to the identity as number of repeated trials N increases to ∞? Each draw of a sketch S_i is $m\times n$, but unlike the above exercise, here $m < n$, and instead of looking at $S_i^T S_i \approx I_n$, we ask about

$$\left\| \frac{1}{N}\sum_{i=1}^{N} S_i^T S_i - I_n \right\| \tag{1}$$

as $N \to \infty$ (in either spectral or Frobenius norm).

Also consider a histogram of outcomes of sketching

a) $x \in \mathbb{R}^n$ where x is very sparse, and
b) $x \in \mathbb{R}^n$ where all the entries of x are similar in value.

Look at $\|Sx\|^2/\|x\|^2$.

Generate S in all of the following manners:

a) Gaussian
b) Haar (cf. HW 2). Note: this is really part of a Haar matrix – it should *not* be a full $n\times n$ orthogonal matrix. It should be $m\times n$ like all the other sketches, but with orthonormal rows.
c) Fast Johnson-Lindenstrauss, $S = RHD$ where R samples m of n rows uniformly at random, H is the discrete cosine transform, and D is diagonal with Rademacher random variables on the diagonal.
d) Simple sampling of m of n rows (e.g., $S = R$ in the above notation)
e) *at least one* of the following variants
   i. the approach from "How to Fake Multiply by a Gaussian Matrix" (Kapralov, Potluru, Woodruff, ICML 2016, https://arxiv.org/abs/1606.05732)

type	sketch	complexity
ℓ_1	Dense Cauchy [Sohler and Woodruff, 2011]	$\mathcal{O}(nd^2 \log n + d^3 \log d + d^{\frac{11}{2}} \log^{\frac{3}{2}} d/\epsilon^2)$
ℓ_1	Fast Cauchy [Clarkson et al., 2013]	$\mathcal{O}(nd \log n + d^3 \log^5 d + d^{\frac{17}{2}} \log^{\frac{9}{2}} d/\epsilon^2)$
ℓ_1	Sparse Cauchy [Meng and Mahoney, 2013a]	$\mathcal{O}(\mathrm{nnz}(A) \log n + d^7 \log^5 d + d^{\frac{19}{2}} \log^{\frac{11}{2}} d/\epsilon^2)$
ℓ_1	Reciprocal Exponential [Woodruff and Zhang, 2013]	$\mathcal{O}(\mathrm{nnz}(A) \log n + d^3 \log d + d^{\frac{13}{2}} \log^{\frac{5}{2}} d/\epsilon^2)$
ℓ_1	Lewis Weights [Cohen and Peng, 2015]	$\mathcal{O}(\mathrm{nnz}(A) \log n + d^3 \log d + d^{\frac{9}{2}} \log^{\frac{1}{2}} d/\epsilon^2)$
ℓ_2	Gaussian Transform	$\mathcal{O}(nd^2 + d^3 \log(1/\epsilon)/\epsilon)$
ℓ_2	SRHT [Tropp, 2011]	$\mathcal{O}(nd \log n + d^3 \log n \log d + d^3 \log(1/\epsilon)/\epsilon)$
ℓ_2	Sparse ℓ_2 embedding [Cohen, 2016]	$\mathcal{O}(\mathrm{nnz}(A) \log n + d^3 \log d + d^3 \log(1/\epsilon)/\epsilon)$
ℓ_2	Refinement Sampling [Cohen et al., 2015a]	$\mathcal{O}(\mathrm{nnz}(A) \log(n/d) \log d + d^3 \log(n/d) \log d + d^3 \log(1/\epsilon)/\epsilon)$

Figure 1: Types of sketches as listed in "Weighted SGD for ℓ_p Regression with Randomized Preconditioning" by Yang, Chow, Ré and Mahoney, SODA 2016, https://arxiv.org/abs/1502.03571
.

ii. the FJLT approach from the original FJLT paper, where R is not simple sub-sampling but rather the matrix R is such that each entry is 0 with a certain probability, otherwise drawn from an appropriately scaled Gaussian (see "Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform", Ailon and Chazelle, 2006, STOC, https://dl.acm.org/citation.cfm?id=1132597)

iii. The low-density parity check (LDPC) codes used for the R term in the FJLT, from "Low Rank Approximation using Error Correcting Coding Matrices" (Ubaru, Mazumdar, Saad, ICML 205, http://proceedings.mlr.press/v37/ubaru15.html).

iv. CountSketch ("Simple and deterministic matrix sketching", Edo Liberty, KDD 2013, https://arxiv.org/abs/1206.0594; or Clarkson and Woodruff, https://doi.org/10.1145/3019134)

v. the approach from "Very Sparse Random Projections" (Li, Hastie, Church, KDD 2006, https://dl.acm.org/citation.cfm?id=1150436), where each entry S_{ij} is iid with value $\sqrt{3}$ with probability 1/6, value $-\sqrt{3}$ with probability 1/6, and value 0 with probability 2/3.

vi. the $S = \text{const} \cdot THG\Pi HB$ approach in eq. (7) from "Fastfood – Approximating Kernel Expansions in Loglinear Time" (Le, Sarlós, Smola, ICML 2013, https://arxiv.org/abs/1408.3060)

vii. Fast Johnson-Lindenstrauss, where H is now a Hadamard transform instead of the DCT

viii. Anything else you can find in the literature, e.g., see Fig. 1

For each of the (at least 5) methods for S, plot the quantity in Eq. (1) as a function of N, and check that it behaves as you think it should.

Try with several choices of n and m to make sure your scaling is correct; most of the sketch matrices need a $\sqrt{n/m}$ or $\sqrt{1/m}$ scaling.

Deliverables: A plot of the quantity in Eq. (1) as a function of N for each type of sketch, and histograms of $\|Sx\|_2^2/\|x\|_2^2$ for the different types of sketches and the two types of x mentioned above. Include at least code snippets for how you generated each type of matrix.

Are these good metrics to look at? Can you think of other metrics to look at?

Hint: do the scaling slightly wrong on purpose, and make sure that your plots detect this error, so that you can be sure your true scaling really is correct. What kind of graph is best (e.g., log or linear scaling on the axes)? What do you expect the histogram to be centered around? Try changing m, n to catch bugs. Usually we want $n \gtrsim 100$ otherwise if it is very small you can get some weird effects.