Homework 8
APPM 4720/5720 Spring 2019
Randomized Algorithms

Due date: Friday, Mar 8 2019
Theme: Entry-wise sampling — Instructor: Stephen 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.

In this homework, we will consider an entry-wise sampling scheme. For a given $m \times n$ target matrix X, consider a sparse matrix $\hat{X}$ where each entry of $\hat{X}$ is chosen independently, and

$$\hat{X}_{i,j} = \begin{cases} X_{i,j}/p_{i,j} & \text{with probability } p_{i,j} \\ 0 & \text{with probability } 1 - p_{i,j} \end{cases} \quad \forall\, i = 1, \ldots, M,\ j = 1, \ldots, N. \tag{1}$$

We will use X to be the matrix from the MNIST data we used previously, so X is a 784×3000 matrix.

Problem 1: [READING] Skim the paper "Matrix Entry-wise Sampling: Simple is Best" (Achlioptas, Karnin, and Liberty, KDD 2013).

Compute the quantities involved in equation (9) of that paper for the matrix X from the MNIST data mentioned above.

Deliverable: the left-hand and right-side quantities in equation (9) of the paper, and a sentence describing if you expect ℓ_1 or ℓ_2 sampling to do better (based on equation (9)).

Problem 2: [MATH] For $\hat{X}$ sampled according to the above scheme,

a) Prove $\mathbb{E}[\hat{X}] = X$

b) Calculate $\mathbb{E}[\hat{X}\hat{X}^T]$, and discuss how we could make unbiased estimator of XX^T. (This would be usesful to estimate covariance matrices).

Problem 3: [CODING]

Sample $\hat{X}$ as discussed above, according to two schemes:

a) ℓ_1 sampling: take $p_{i,j} \propto |X_{i,j}|$, e.g., $p_{i,j} = \min\left(1, c|X_{i,j}|/(\sum_{i',j'} |X_{i',j'}|)\right)$ and adjust the constant c to vary the expected number of sampled entries.

b) ℓ_2 sampling: take $p_{i,j} \propto X_{i,j}^2$

Deliverable: for several sparsities (i.e., different values of c), record the error (in either spectral or Frobenius norm) of $\|X - \hat{X}\|/\|X\|$ for both ℓ_1 and ℓ_2 sampling schemes, and make a plot of sparsity vs. error. Discuss if the results agree with your prediction based on Problem 1.

Hint: to make sure your code is correct, you might want to run some experiments to verify $\mathbb{E}[\hat{X}] = X$, e.g., by looking at $\|1/N \sum_{k=1}^{N} \hat{X}_k - X\|$ for independent draws $\hat{X}_k$, and plotting this error on a log-log plot to make sure it decays to 0.

Bonus (not required, only for fun): try ℓ_2 sampling with a truncation scheme for small entries. Does this help? Or, try ℓ_1 or ℓ_2 sampling where the normalization is done per row or per column. Does this improve overall errors vs. compression?

Double bonus (not required, only for fun): under either sampling scheme, can you make an estimator that is an unbiased estimator of the covariance $1/nXX^T$? You may need to record a little more information than just $\hat{X}$.