

Homework 10

APPM 4720/5720 Spring 2019

Randomized Algorithms

Due date: Monday, April 1 2019 by 5 PM. I suggest you aim for Fri March 22 in order to exploit office hours.

Theme: Randomized SVD

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.

Please refer to “Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions” (Halko, Martinsson, Tropp; SIAM Review 2011; we’ll refer to this paper as “HMT”), and also Algo 3 of “Single-pass PCA of large high-dimensional data” (Yu, Gu, Li, Liu, Li; IJCAI 2017, [arXiv link](#))

Algorithm 1 Basic 2-pass randomized SVD of HMT, i.e., Algo 4.1 + Algo 5.1

```

1: procedure RANDOMIZED SVD( $A, k, p, q$ )  $\triangleright A$  is  $M \times N$ ,  $k$  is target rank,  $p$  oversampling,  $q \geq 0$  pwr iter
2:   Draw  $N \times \ell$  matrix  $\Omega$ , iid Normal entries ( $\ell = k + p$ )
3:    $Y = A\Omega$ 
4:    $QR = Y$   $\triangleright$  Thin QR factorization
5:   for  $i = 1, \dots, q$  do  $\triangleright$  Optional: power iterations for increased accuracy
6:      $B = A^T Q$ 
7:      $QR = B, B \leftarrow Q$ 
8:      $Y = AB$ 
9:      $QR = Y$ 
10:  end for
11:   $B = A^T Q$   $\triangleright$  So  $A \approx QQ^T A = QB^T$ 
12:   $\tilde{U}\Sigma V^T = B^T$   $\triangleright$  Thin svd
13:  Keep first  $k$  columns of  $\tilde{U}$  and  $V$ , first  $k$  cols and rows of  $\Sigma$ 
14:   $U = Q\tilde{U}$ 
15:  return  $\{U, \Sigma, V\}$  and/or  $\tilde{A} \stackrel{\text{def}}{=} U\Sigma V^T$ 
16: end procedure

```

Problem 1: [CODING] Compare the performance of

- a) the randomized SVD (fix the oversampling parameter, e.g., $p = 10$,
- b) the randomized SVD with a few power iterations, and
- c) the Lanczos method (`svds` in Matlab, or `scipy.sparse.linalg.svds` in Python).

Compare the algorithms on the following two matrices (one dense, the other sparse). In Python, you don’t have to do exactly this, but make similar matrices.

Matrix 1:

```

1 M = 1e3; N = 2*M;
2 A = 1/sqrt(M*N)*randn(M,N)*diag(logspace(0,-5,N))*randn(N,N);

```

Algorithm 2 1-pass randomized SVD of HMT section 5.5

```
1: procedure ONE-PASS RANDOMIZED SVD( $A, k, p$ )  $\triangleright A$  is  $M \times N$ ,  $k$  is target rank,  $p$  oversampling
2:   Draw  $N \times \ell$  matrix  $\Omega$ , iid Normal entries ( $\ell = k + p$ )
3:   Draw  $M \times \ell$  matrix  $\tilde{\Omega}$ , iid Normal entries
4:   Multiply  $Y = A\Omega$  and  $\tilde{Y} = A^T\tilde{\Omega}$   $\triangleright$  In 1-pass over data
5:    $QR = Y$  and  $\tilde{Q}\tilde{R} = \tilde{Y}$   $\triangleright$  Thin QR factorization
6:    $B = \operatorname{argmin}_{B \in \mathbb{R}^{\ell \times \ell}} \|Q^TY - B(\tilde{Q}^T\Omega)\|_F^2 + \|(\tilde{Q}^T\tilde{Y})^T - (Q^T\tilde{\Omega})^TB\|_F^2$   $\triangleright$  Use LSQR
7:    $\tilde{U}\Sigma\tilde{V}^T = B$   $\triangleright$  Thin svd
8:   Keep first  $k$  columns of  $\tilde{U}$  and  $\tilde{V}$ , first  $k$  cols and rows of  $\Sigma$ 
9:    $U = Q\tilde{U}$ ,  $V = \tilde{Q}\tilde{V}$ 
10:  return  $\{U, \Sigma, V\}$  and/or  $\tilde{A} \stackrel{\text{def}}{=} U\Sigma V^T$ 
11: end procedure
```

Algorithm 3 1-pass randomized SVD of Yu et al. '17

```
1: procedure BETTER ONE-PASS RANDOMIZED SVD( $A, k, p$ )  $\triangleright A$  is  $M \times N$ ,  $k$  is target rank,  $p$ 
   oversampling
2:   Draw  $N \times \ell$  matrix  $\Omega$ , iid Normal entries ( $\ell = k + p$ )
3:   Multiply  $Y = A\Omega$  and  $B = A^TY$   $\triangleright$  In 1-pass over data
4:    $QR = Y$  and  $\tilde{Q}\tilde{R} = \tilde{Y}$   $\triangleright$  Thin QR factorization
5:    $B \leftarrow BR^{-1}$   $\triangleright$  so  $B = A^TQ$ ; this step must be modified if  $R$  is rank deficient!
6:    $\tilde{U}\Sigma\tilde{V}^T = B^T$   $\triangleright$  Thin svd
7:   Keep first  $k$  columns of  $\tilde{U}$  and  $V$ , first  $k$  cols and rows of  $\Sigma$ 
8:    $U = Q\tilde{U}$ 
9:   return  $\{U, \Sigma, V\}$  and/or  $\tilde{A} \stackrel{\text{def}}{=} U\Sigma V^T$ 
10: end procedure
```

Matrix 2:

```
1 M = 2e3; N = 2*M;
2 A = sprandn(M,N,0.05);
```

Deliverable: Plot the errors in spectral and Frobenius norm, and the timing, for the 3 methods as well as for a direct dense SVD (`svd` in Matlab, `numpy.linalg.svd` in Python), for both matrices. Overall, that is 6 plots (2 errors + timing \times 2 matrices).

Hints: computing the spectral norm of the error can be the most time-consuming part of your code, so you can estimate this using power iterations. Matlab does it for you if you have a sparse matrix (`normest`) but we really need it for a sparse matrix minus a low-rank matrix. I've adapted `normest` to do this and put it on Github, called `my_normest.m`, and you can similarly adapt one for Python.

In Matlab, to export figures, I recommend `export_fig`. Make sure your graphs are readable if you print them in black and white!

Problem 2: [CODING] Compare the two versions of one-pass SVD algorithms, Algorithms 2 and 3 (also include the two-pass Algorithm 1 as a benchmark). Do this on a similar matrix to problem 1, but one that has faster decaying singular values:

```
1 M = 1e3; N = 2*M;
2 A = 1/sqrt(M*N)*randn(M,N)*diag(logspace(0,-5,N).^4)*randn(N,N);
```

Deliverable: for $k = 150$, what is the Frobenius norm error for all 3 randomized SVD algorithms? Comment on the accuracy of the different algorithms. You can choose how much oversampling p to use.

Hint: for Algorithm 2, you find B as the solution to a least-squares problem. However, it's not a typical least-squares problem because it is a matrix variable. It is equivalent to a standard least-squares problem in the variable $b = \text{vec}(B) \stackrel{\text{def}}{=} B(:)$ [Matlab notation], with a linear operator

$$\tilde{\mathcal{A}}(b) = \mathcal{A}(B = \text{mat}(b)) = \begin{bmatrix} \text{vec}(BD) \\ \text{vec}(FB) \end{bmatrix}, \quad D \stackrel{\text{def}}{=} \tilde{Q}^T \Omega, \quad F \stackrel{\text{def}}{=} (Q^T \tilde{\Omega})^T$$

and right-hand-side $\begin{bmatrix} \text{vec}(Q^T Y) \\ \text{vec}((\tilde{Q}^T \tilde{Y})^T) \end{bmatrix}$. Here, $\text{mat}(b)$ reshapes b into an $\ell \times \ell$ matrix, in whatever order it needs to be consistent with vec so that $\text{mat} \circ \text{vec}$ is the $\ell^2 \times \ell^2$ identity matrix.

You definitely want to use an iterative method, because making \mathcal{A} explicit is a large matrix and it would be inefficient to apply ($\mathcal{O}(\ell^4)$ flops, compared to $\mathcal{O}(\ell^3)$ flops if you do it as above). In matlab, use `lsqr`, and in python use `scipy.sparse.linalg.lsqr`. For python, the routine will expect you to pass in a `LinearOperator` object.

For an iterative method, you need to know the transpose/adjoint of $\tilde{\mathcal{A}}$. Use the fact that the order of composition of linear operators flips when you do the transpose, and that mat and vec are transposes of each other, and the transpose of $B \mapsto BD$ is $Z \mapsto ZD^T$, etc. If you're confused about this part, please come to office hours!

Note that Algorithm 3 will fail if you apply it to a matrix that has true rank less than ℓ . You can make the algorithm robust by computing the SVD of B , determining what the effective rank is, then removing columns of Y and B . You do *not need to implement this!*.

Problem 3: [CODING] Run Algorithm 3 on the 73 GB file `/rc_scratch/stbe1590/hugeSquareMatrix.mat` (on the research computing cluster; refer to HW 3 and HW 4) which contains a $10^5 \times 10^5$ matrix, for $k = 150$. You may want to check your algorithm using the smaller matrix `/rc_scratch/stbe1590/largeSquareMatrix.mat` which is only 752 MB and contains a $10^4 \times 10^4$ matrix.

Deliverable: return a plot of the estimated top 150 singular values (note that it is not easy to report the error, as this requires either a second pass through the data, or a randomized estimate), and also report the time it took to read in the data and the time it took to do the computation.

Hint: to compute something like $Y = A\Omega$ and $B = A^T Y$ in 1-pass over the data, assuming the entire data matrix A is too large to store in memory, load A by rows and for simplicity, assume we can split A into 2 blocks of rows, each of which fits into memory (and it will soon be clear how to generalize to more than 2 blocks), eg., $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$, and similarly, $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} A_1 \Omega \\ A_2 \Omega \end{bmatrix}$.

Then $B = A^T Y = \begin{bmatrix} A_1^T & A_2^T \end{bmatrix} \cdot \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = A_1^T Y_1 + A_2^T Y_2$. Thus the idea is as follows: load A_1 into memory, compute $Y_1 = A_1 \Omega$, then compute $B = A_1^T Y_1$. Now throw away A_1 and load A_2 into memory, compute $Y_2 = A_2 \Omega$, and then update $B \leftarrow B + A_2^T Y_2$.