Fast Matrix Algorithms for Data Analytics: Problem Set 3

1. The purpose of this exercise is to prove the equivalence of subspace iteration and the "power" version of the RSVD. Suppose you are given an $m \times n$ matrix $\mathbf{A}$ and an $n \times k$ matrix $\mathbf{G}$ of full rank. Then set

$$\mathbf{Y} = (\mathbf{A}\mathbf{A}^*)^q \mathbf{A}\mathbf{G}$$

for some positive integer q. Also define $\mathbf{Z}$ as the output of the iteration

$$\begin{aligned}
&\mathbf{Z} \leftarrow \text{orth}(\mathbf{A}\mathbf{G}) \\
&\textbf{for } i = 1 : q \\
&\quad \mathbf{Z} \leftarrow \text{orth}(\mathbf{A}^*\mathbf{Z}) \\
&\quad \mathbf{Z} \leftarrow \text{orth}(\mathbf{A}\mathbf{Z}) \\
&\textbf{end}
\end{aligned}$$

The output of orth($\mathbf{A}$) for a matrix $\mathbf{A}$ is a matrix $\mathbf{Z}$ with orthonormal columns such that ran($\mathbf{Z}$) = ran($\mathbf{A}$). Show that ran($\mathbf{Y}$) = ran($\mathbf{Z}$).

2. Suppose we would like to perform the `RSVD algorithm` on a matrix $\mathbf{A}$, but we do not know in advance the rank of $\mathbf{A}$. In this case, to find a low-rank approximation of $\mathbf{A}$, for phase A of `RSVD` we wish to construct $\mathbf{Q}$ such that $\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^*\mathbf{A}\| < \epsilon$, where ϵ is a user-defined parameter. An algorithm for accomplishing this, published as Algorithm 4.2 of the paper "Finding Structure with Randomness" by Halko, Martinsson, and Tropp, is given in Algorithm 1 below. To make sure we understand why this algorithm works, we'll focus in on lines 7-9.

   Specifically, consider the iteration:
   Input: $\mathbf{Q} \in \mathbb{R}^{m \times r}$, $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}_r$.
   Iteration:

$$\begin{aligned}
\bar{\mathbf{y}} &= (\mathbf{I} - \mathbf{Q}\mathbf{Q}^*)\mathbf{y} \\
\mathbf{q} &= \frac{\bar{\mathbf{y}}}{\|\bar{\mathbf{y}}\|} \\
\bar{\mathbf{Q}} &= [\mathbf{Q}, \mathbf{q}]
\end{aligned}$$

   Show that for this procedure, the output $\bar{\mathbf{Q}}$ is such that ran($\bar{\mathbf{Q}}$) = ran($[\mathbf{Q} \quad \mathbf{y}]$) and $\bar{\mathbf{Q}}^*\bar{\mathbf{Q}} = \mathbf{I}_{r+1}$.

Algorithm 1 Adaptive Randomized Range Finder

Given an $m \times n$ matrix $\mathbf{A}$, a tolerance ϵ, and an integer r (e.g., $r = 10$), the following scheme computes an orthonormal matrix $\mathbf{Q}$ such that (4.2) holds with probability at least $1 - \min\{m, n\}10^{-r}$.

Draw standard Gaussian vectors $\omega^{(1)}, \ldots, \omega^{(r)}$ of length n.
$i = 1, 2, \ldots, r$, compute $\mathbf{y}^{(i)} = \mathbf{A}\omega^{(i)}$.
$j = 0$.
$\mathbf{Q}^{(0)} = [\,]$, the $m \times 0$ empty matrix.
while $\max\left\{\|\mathbf{y}^{(j+1)}\|, \|\mathbf{y}^{(j+2)}\|, \ldots, \|\mathbf{y}^{(j+r)}\|,\right\} > \epsilon/(10\sqrt{2/\pi})$,
    $j = j + 1$.
    Overwrite $\mathbf{y}^{(j)}$ by $(\mathbf{I} - \mathbf{Q}^{(j-1)}(\mathbf{Q}^{(j-1)})^*)\mathbf{y}^{(j)}$.
    $\mathbf{q}^{(j)} = \mathbf{y}^{(j)}/\|\mathbf{y}^{(j)}\|$.
    $\mathbf{Q}^{(j)} = [\mathbf{Q}^{(j-1)}\ \mathbf{q}^{(j)}]$.
    Draw a standard Guassian vector $\omega^{(j+r)}$ of length n.
    $\mathbf{y}^{(j+r)} = (\mathbf{I} - \mathbf{Q}^{(j)}(\mathbf{Q}^{(j)})^*)\mathbf{A}\omega^{(j+r)}$.
    for $i = (j+1), (j+2), \ldots, (j+r-1)$,
        Overwrite $\mathbf{y}^{(i)}$ by $\mathbf{y}^{(i)} - \mathbf{q}^{(j)}\langle\mathbf{q}^{(j)}, \mathbf{y}^{(i)}\rangle$.
    end for
end while
$\mathbf{Q} = \mathbf{Q}^{(j)}$.

3. Let $\mathbf{R}$ be an $m \times n$ random matrix. Assume the entries of $\mathbf{R}$ are independent, and $\mathbb{E}[\mathbf{R}_{ij}] = 0$ and $\mathrm{Var}(\mathbf{R}_{ij}) = 1\ \forall i, j$. Let $\mathbf{x} \in \mathbb{R}^n$. Show that $\mathbb{E}[\|\mathbf{R}\mathbf{x}\|^2] = m\|\mathbf{x}\|^2$.

4. Let k and ℓ be positive integers such that $k < \ell$, and let $\mathbf{G}$ be a $k \times \ell$ random matrix whose every entry is drawn independently from a normalized Gaussian distribution. It is known that when ℓ is "sufficiently" much larger than k, the rows of $\mathbf{G}$ will be close to orthogonal, and that $\mathbf{G}$ will be well conditioned. In this exercise, we will walk through an argument that provides an indication of how this could be proven.

   Let $N(0,1)$ denote a normalized Gaussian distribution and let $S(k)$ denote the distribution for the square root of random variable with a χ_k^2 distribution. In other words, if $\mathbf{g}$ is a vector of length k whose entries are drawn independently from $N(0,1)$, then $\|\mathbf{g}\| \in S(k)$.

   (a) Assume we have used entries in the first row of $\mathbf{G}$ to build a Householder reflector that you apply from the right of $\mathbf{G}$ to obtain a matrix of the form

$$\mathbf{H}_1 = \begin{bmatrix} h_{11} & 0 & 0 & 0 & 0 & \cdots \\ g & g & g & g & g & \cdots \\ g & g & g & g & g & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Show that each entry marked by "g" is drawn from $N(0,1)$ and $h_{11} \in S(\ell)$.

(b) Now suppose we have used the entries below the diagonal in the first column to build a Householder reflector that upon application from the left maps $\mathbf{H}_1$ to a matrix of the form

$$\mathbf{H}_2 = \begin{bmatrix} h_{11} & 0 & 0 & 0 & 0 & \cdots \\ h_{21} & g & g & g & g & \cdots \\ 0 & g & g & g & g & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Show that each entry marked by "g" is drawn from $N(0,1)$, that $h_{11} \in S(\ell)$, and that $h_{21} \in S(k-1)$.

(c) Suppose we continue building Householder reflectors in the pattern outlined in (a) and (b) to drive $\mathbf{G}$ all the way to a bidiagonal matrix

$$\mathbf{H} = \begin{bmatrix} h_{11} & 0 & 0 & 0 & 0 & \cdots \\ h_{21} & h_{22} & 0 & 0 & 0 & \cdots \\ 0 & h_{32} & h_{33} & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Show that each diagonal entry $h_{ii} \in S(\ell - i + 1)$ and that each sub-diagonal entry $h_{i+1,i} \in S(k-i)$.

(d) The expectation of a random variable h drawn from distribution $S(n)$ is $\mathbb{E}[h] = \sqrt{2}\Gamma((n+1)/2)/\Gamma(n/2)$. Show that $\mathbb{E}[\mathbf{H}]$ is well-conditioned (*i.e.* the singular values are bounded away from 0) when ℓ is sufficiently much larger than k. [1]

Hint: Use that $\mathbf{H}$ *is diagonally dominant, and apply the Gershgorin circle theorem.*

[1] Warning: If you work out all the details, this gets messy! I'd recommend that you only work out enough details to convince yourself that the singular values can be bounded away from 0.