Fast Matrix Algorithms for Data Analytics: Problem Set 1 Solutions

1. Assume that for $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{Q} \in \mathbb{R}^{m \times k}$ where $\mathbf{Q}$ has orthonormal columns, $\text{range}(\mathbf{Q}) = \text{range}(\mathbf{A})$. Prove that $\mathbf{A} = \mathbf{Q}\mathbf{Q}^*\mathbf{A}$.

   Hint: A linear operator $P : X \to Y$ between two vector spaces X and Y is a projection iff it satisfies $P^2 = P$. Projections satisfy the property that for all $x \in \text{range}(P)$, $Px = x$. (this follows from the definition; if you don't see why, prove it!) If you don't see how to proceed with the proof, try an approach that takes advantage of this information.

   Solution: First, observe that since $(\mathbf{Q}\mathbf{Q}^*)^2 = \mathbf{Q}(\mathbf{Q}^*\mathbf{Q})\mathbf{Q}^* = \mathbf{Q}\mathbf{I}\mathbf{Q}^* = \mathbf{Q}\mathbf{Q}^*$, the matrix $\mathbf{Q}\mathbf{Q}^*$ is a projection. Therefore, for all $\mathbf{x} \in \text{ran}(\mathbf{Q}\mathbf{Q}^*), \mathbf{Q}\mathbf{Q}^*\mathbf{x} = \mathbf{x}$.

   So, if $\text{ran}(\mathbf{Q}\mathbf{Q}^*) = \text{ran}(\mathbf{Q})$, we're done, because then $\text{ran}(\mathbf{Q}\mathbf{Q}^*) = \text{ran}(\mathbf{A})$, so

$$\begin{aligned}
&\mathbf{a}_i \text{ is a column of } \mathbf{A} \\
&\implies \mathbf{a}_i \in \text{ran}(\mathbf{A}) = \text{ran}(\mathbf{Q}\mathbf{Q}^*) \\
&\implies \mathbf{Q}\mathbf{Q}^*\mathbf{a}_i = \mathbf{a}_i \\
&\implies \mathbf{Q}\mathbf{Q}^*\mathbf{A} = \mathbf{A}.
\end{aligned}$$

   We'll now show that $\text{ran}(\mathbf{Q}\mathbf{Q}^*) = \text{ran}(\mathbf{Q})$. Assume $\mathbf{U}\mathbf{D}\mathbf{V}^*$ is the SVD of $\mathbf{Q}$. Then

$$\mathbf{Q}\mathbf{Q}^* = \mathbf{U}\mathbf{D}\mathbf{V}^*\mathbf{V}\mathbf{D}\mathbf{U}^* = \mathbf{U}\mathbf{D}^2\mathbf{U}^*.$$

   Therefore $\mathbf{U}\mathbf{D}^2\mathbf{U}^*$ is the SVD of $\mathbf{Q}\mathbf{Q}^*$, and in particular $\mathbf{Q}$ and $\mathbf{Q}\mathbf{Q}^*$ have the same left singular vectors. The left singular vectors of a matrix form a basis for the range of the matrix, so since the ranges of $\mathbf{Q}$ and $\mathbf{Q}\mathbf{Q}^*$ share the same basis, they are in fact identical, *i.e.* $\text{ran}(\mathbf{Q}\mathbf{Q}^*) = \text{ran}(\mathbf{Q})$.

   Alternate Solution: Let $\mathbf{T} = \mathbf{Q}\mathbf{Q}^*\mathbf{A}$, and let $\mathbf{t}_i$ be the column vectors of $\mathbf{T}$. Then by carrying out the matrix multiplication, we get

$$\mathbf{t}_i = \sum_{j=1}^{k} (\mathbf{q}_j^*\mathbf{a}_i)\mathbf{q}_j,$$

   where $\mathbf{a}_i$ and $\mathbf{q}_j$ are the column vectors of $\mathbf{A}$ and $\mathbf{Q}$, respectively. Since $\text{ran}(\mathbf{Q}) = \text{ran}(\mathbf{A})$, the columns of $\mathbf{Q}$ form an orthonormal basis for $\text{ran}(\mathbf{A})$, so any vector $\mathbf{x} \in \text{ran}(\mathbf{A})$ can be written as $\mathbf{x} = \sum_{j=1}^{k} (\mathbf{q}_j^* \mathbf{x}) \mathbf{q}_j$. In particular,

$$\mathbf{a}_i = \sum_{j=1}^{k} (\mathbf{q}_j^* \mathbf{a}_i) \mathbf{q}_j = \mathbf{t}_i, \quad 1 \leq i \leq n,$$

so $\mathbf{T} = \mathbf{Q}\mathbf{Q}^* \mathbf{A} = \mathbf{A}$.

2. (a) Let $\mathbf{A}$ be an $m \times n$ matrix, set $p = \min(m, n)$, and suppose that the singular value decomposition of $\mathbf{A}$ takes the form

$$\underset{m \times n}{\mathbf{A}} = \underset{m \times p}{\mathbf{U}} \; \underset{p \times p}{\mathbf{D}} \; \underset{p \times n.}{\mathbf{V}^*} \tag{1}$$

Recall the definition of the spectral norm of $\mathbf{A}$:

$$\|\mathbf{A}\| = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \sup_{\|\mathbf{x}\|_2 = 1} \|\mathbf{A}\mathbf{x}\|_2.$$

Let k be an integer such that $1 \leq k < p$ and let $\mathbf{A}_k$ denote the truncation of the SVD to the first k terms:

$$\mathbf{A}_k = \mathbf{U}(:, 1:k)\mathbf{D}(1:k, 1:k)\mathbf{V}(:, 1:k)^*.$$

Prove directly from the definition of the spectral norm that

$$\|\mathbf{A} - \mathbf{A}_k\| = \sigma_{k+1}. \tag{2}$$

Solution: First, partition the factorization $\mathbf{U}\mathbf{D}\mathbf{V}^*$ as

$$\begin{array}{cc} m & \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \\ & \begin{array}{cc} k & p-k \end{array} \end{array} \begin{array}{c} k \\ p-k \end{array} \begin{array}{c} \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} \\ \begin{array}{cc} k & p-k \end{array} \end{array} \begin{array}{c} k \\ p-k \end{array} \begin{array}{c} \begin{bmatrix} \mathbf{V}_1^* \\ \mathbf{V}_2^* \end{bmatrix} \\ n \end{array} .$$

Then observe that $\mathbf{U}_1 = \mathbf{U}(:, 1:k)$, $\mathbf{D}_1 = D(1:k, 1:k)$, and $\mathbf{V}_1^* = V(:, 1:k)^*$, so that $\mathbf{A}_k = \mathbf{U}_1\mathbf{D}_1\mathbf{V}_1^*$. By carrying out block multiplication on the partitioned factorization, we see that

$$\mathbf{A} = \mathbf{U}_1\mathbf{D}_1\mathbf{V}_1^* + \mathbf{U}_2\mathbf{D}_2\mathbf{V}_2^* = \mathbf{A}_k + \mathbf{U}_2\mathbf{D}_2\mathbf{V}_2^*,$$

so

$$\mathbf{A} - \mathbf{A}_k = \mathbf{U}_2\mathbf{D}_2\mathbf{V}_2^*. \tag{3}$$

Let $\mathbf{x} \in \mathbb{R}^n$ be any vector such that $\|\mathbf{x}\| = 1$. We will show that $\|(\mathbf{A} - \mathbf{A}_k)\mathbf{x}\| \leq \sigma_{k+1}$. We establish the notation that $\mathbf{v}_i$ and $\mathbf{u}_i$ are the columns of $\mathbf{V}$ and $\mathbf{U}$, respectively. Since the columns of $\mathbf{V}$ are orthonormal we can construct an orthonormal basis of $\mathbb{R}^n$: $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{v}_{p+1}, \ldots, \mathbf{v}_n\}$ (note that vectors $\mathbf{v}_{p+1}$ through $\mathbf{v}_n$ are not actually columns of $\mathbf{V}$ but are simply used to construct the basis), and thus

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{v}_n$$

for some $c_i, i = 1, \ldots, n$. Now, we have that

$$(\mathbf{A} - \mathbf{A}_k)\mathbf{x} = \mathbf{U}_2\mathbf{D}_2\mathbf{V}_2^*\mathbf{x}.$$

Since the i-th entry of $\mathbf{V}_2^*\mathbf{x}$ is $\langle \mathbf{v}_i, \mathbf{x} \rangle$, and since $\mathbf{x}$ is a linear combination of the orthonormal basis $\{\mathbf{v}_i\}_{i=1}^n$,

$$\mathbf{V}_2^*\mathbf{x} = \begin{bmatrix} c_{k+1} \\ c_{k+2} \\ \vdots \\ c_p \end{bmatrix},$$

and so

$$\mathbf{U}_2\mathbf{D}_2\mathbf{V}_2^*\mathbf{x} = \sum_{i=k+1}^{p} c_i \sigma_i \mathbf{u}_i$$

which implies

$$\|\mathbf{U}_2\mathbf{D}_2\mathbf{V}_2^*\mathbf{x}\| \leq \sigma_{k+1} \| \sum_{i=k+1}^{p} c_i \mathbf{u}_i \|.$$

Finally, by the orthonormality of $\{\mathbf{u}_i\}_{i=1}^p$,

$$\| \sum_{i=k+1}^{p} c_i \mathbf{u}_i \|^2 = \sum_{i=k+1}^{p} c_i^2,$$

and by the orthonormality of $\{\mathbf{v}_i\}_{i=1}^n$,

$$1 = \|\mathbf{x}\|^2 = \| \sum_{i=1}^{n} c_i \mathbf{v}_n \|^2 = \sum_{i=1}^{n} c_i^2 \implies \sum_{i=k+1}^{p} c_i^2 \leq 1,$$

and therefore

$$\|(\mathbf{A} - \mathbf{A}_k)\mathbf{x}\| \leq \sigma_{k+1} \| \sum_{i=k+1}^{p} c_i \mathbf{u}_i \| \leq \sigma_{k+1}.$$

Thus, we have shown that $\|\mathbf{A} - \mathbf{A}_k\| \leq \sigma_{k+1}$. Next, we observe that for $\mathbf{x} = \mathbf{v}_{k+1}$,

$$\|(\mathbf{A} - \mathbf{A}_k)\mathbf{x}\| = \|\sigma_{k+1}\mathbf{u}_{k+1}\| = \sigma_{k+1}\|\mathbf{u}_{k+1}\| = \sigma_{k+1},$$

so since $\|\mathbf{v}_{k+1}\| = 1$, $\|\mathbf{A} - \mathbf{A}_k\| \geq \sigma_{k+1}$. Therefore,

$$\|\mathbf{A} - \mathbf{A}_k\| = \sigma_{k+1}.$$

(b) In phase A of the `RSVD` algorithm, we seek a matrix $\mathbf{Q} \in \mathbb{R}^{m\times k}$ with orthonormal columns such that $\mathbf{A} \approx \mathbf{Q}\mathbf{Q}^*\mathbf{A}$. Since $\operatorname{rank}(\mathbf{Q}\mathbf{Q}^*\mathbf{A}) \leq k$, the Eckart-Young Theorem assures us that

$$\inf_{\mathbf{Q}\in\mathbb{R}^{m\times k}} \|\mathbf{A} - \mathbf{Q}\mathbf{Q}^*\mathbf{A}\| \geq \sigma_{k+1}.$$

Show that we can achieve this bound by choosing $\mathbf{Q} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k\end{bmatrix}$, where $\{\mathbf{u}_i\}_{i=1}^k$ are the k leading left singular vectors of $\mathbf{A}$. That is, show that for such a $\mathbf{Q}$, we have

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^*\mathbf{A}\| = \sigma_{k+1}.$$

Solution: Let $p = \min(m, n)$ and assume $\mathbf{U}\mathbf{D}\mathbf{V}^*$ is the SVD of $\mathbf{A}$. We have that $\mathbf{Q} = \mathbf{U}_k = \mathbf{U}(:, 1 : k) = \mathbf{U}\mathbf{P}$, where $\mathbf{P}$ is a $p \times k$ matrix defined by

$$\mathbf{P} = \begin{bmatrix}\mathbf{I}_k \\ \mathbf{0}\end{bmatrix}.$$

Then we have that

$$\mathbf{A} - \mathbf{Q}\mathbf{Q}^*\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^* - \mathbf{U}\mathbf{P}\mathbf{P}^*\mathbf{U}^*\mathbf{U}\mathbf{D}\mathbf{V}^* = \mathbf{U}(\mathbf{D} - \mathbf{P}\mathbf{P}^*\mathbf{D})\mathbf{V}^*.$$

Carrying out the matrix multiplication, we see that

$$\mathbf{P}\mathbf{P}^*\mathbf{D} = \begin{bmatrix}\mathbf{D}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} =: \hat{\mathbf{D}}_k.$$

Thus, since the spectral norm is invariant under multiplication by matrices with orthonormal columns or rows, we have

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^*\mathbf{A}\| = \|\mathbf{U}(\mathbf{D} - \hat{\mathbf{D}}_k)\mathbf{V}^*\| = \|\mathbf{D} - \hat{\mathbf{D}}_k\| = \sigma_{k+1}.$$

3. Suppose $\mathbf{A}$ is a real symmetric $n \times n$ matrix with eigenpairs $\{\lambda_j, \mathbf{v}_j\}_{j=1}^n$, ordered so that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. Define a sequence of vectors $\mathbf{x}_p = \mathbf{A}^p \mathbf{v}_0$, where $\mathbf{v}_0$ is an $n \times 1$ random vector whose entries are drawn independently from a standard Gaussian distribution.

(a) Set $\beta = |\lambda_2|/|\lambda_1|$ and $\mathbf{y}_p = (1/\|\mathbf{x}_p\|)\mathbf{x}_p$. Assume $\lambda_1 = 1$ and $\beta < 1$. Prove that as $p \to \infty$, the vectors $\{\mathbf{y}_p\}$ converge either to $\mathbf{v}_1$ or $-\mathbf{v}_1$.

Solution: Since $\mathbf{A}$ is symmetric, we may assume without loss of generality that the eigenvectors $\{\mathbf{v}_i\}_{i=1}^n$ of $\mathbf{A}$ form an orthonormal basis for $\mathbb{R}^n$. Thus, we can write

$$\mathbf{x}_0 = \sum_{i=1}^n c_i \mathbf{v}_i,$$

for some $c_i \in \mathbb{R}, i = 1, 2, \ldots, n$, which implies

$$\mathbf{x}_p = \sum_{i=1}^n c_i \lambda_i^p \mathbf{v}_i = c_1 \mathbf{v}_1 + \sum_{i=2}^n c_i \lambda_i^p \mathbf{v}_i$$

and since $\mathbf{x}_0$ is chosen from an i.i.d. Gaussian distribution, each c_i is nonzero with probability 1. Next, after observing $\beta < 1$ implies that $|\lambda_i| < 1$ for $i > 1$, we have that

$$\lim_{p\to\infty} c_i \lambda_i^p \mathbf{v}_i = \mathbf{0}, \quad i = 2, 3, \ldots, n.$$

Therefore,

$$\begin{aligned}
\lim_{p\to\infty} \mathbf{y}_p &= \frac{\lim_{p\to\infty} \mathbf{x}_p}{\|\lim_{p\to\infty} \mathbf{x}_p\|} \\
&= \frac{c_1 \mathbf{v}_1 + \lim_{p\to\infty} \sum_{i=2}^n c_i \lambda_i^p \mathbf{v}_i}{\|c_1 \mathbf{v}_1 + \lim_{p\to\infty} \sum_{i=2}^n c_i \lambda_i^p \mathbf{v}_i\|} \\
&= \frac{c_1 \mathbf{v}_1}{|c_1| \|\mathbf{v}_1\|} \\
&= \operatorname{sign}(c_1) \mathbf{v}_1 = \pm \mathbf{v}_1.
\end{aligned}$$

(b) What is the speed of convergence of $\{\mathbf{y}_p\}$?

Solution: Since the convergence of $\{\mathbf{y}_p\}$ is controlled by how quickly the $c_i \lambda_i^p \mathbf{v}_i$ terms go to zero as p increases, the convergence will be geometric, with the largest $|\lambda_i|$, $i = 2, 3, \ldots, n$ controlling the speed. Therefore, the error is reduced by roughly $|\lambda_2|$ each iteration.

(c) Assume again that $\beta < 1$, but now drop the assumption that $\lambda_1 = 1$. Prove that your answers in (a) and (b) are still correct, with the exception that if $\lambda_1 < 0$, then it is the vector $(-1)^p \mathbf{y}_p$ that converges instead.

Solution: As before, note that we can write

$$\mathbf{x}_p = \sum_{i=1}^{n} c_i \lambda_i^p \mathbf{v}_i.$$

Then $\beta < 1$ implies that

$$\lim_{p\to\infty} c_i \left(\frac{\lambda_i}{|\lambda_1|}\right)^p \mathbf{v}_i = 0, \quad i = 2, 3, \ldots, n.$$

Therefore, we have that

$$\begin{aligned}
\lim_{p\to\infty} (-1)^p \mathbf{y}_p &= \frac{\lim_{p\to\infty} (-1)^p \frac{\mathbf{x}_p}{|\lambda_1|^p}}{\| \lim_{p\to\infty} \frac{\mathbf{x}_p}{|\lambda_1|^p} \|} \\
&= \frac{\sum_{i=1}^n \lim_{p\to\infty} (-1)^p c_i \left(\frac{\lambda_i}{|\lambda_1|}\right)^p \mathbf{v}_i}{\| \sum_{i=1}^n \lim_{p\to\infty} c_i \left(\frac{\lambda_i}{|\lambda_1|}\right)^p \mathbf{v}_i \|} \\
&= \frac{\lim_{p\to\infty} (-1)^p c_1 \left(\frac{\lambda_1}{|\lambda_1|}\right)^p \mathbf{v}_1}{\| \lim_{p\to\infty} c_1 \left(\frac{\lambda_1}{|\lambda_1|}\right)^p \mathbf{v}_1 \|} \\
&= \frac{\lim_{p\to\infty} (-1)^{2p} c_1 \mathbf{v}_1}{|c_1| \|\mathbf{v}_1\|} \\
&= \operatorname{sign}(c_1) \mathbf{v}_1 = \pm \mathbf{v}_1.
\end{aligned}$$

If $\beta > 1$, then we similarly have

$$\lim_{p\to\infty} \mathbf{y}_p = \pm \mathbf{v}_1.$$

Furthermore, the convergence will again be geometric with rate β since the term that controls the rate of the convergence is $\frac{|\lambda_2|}{|\lambda_1|}$, so the error is reduced by roughly $\frac{|\lambda_2|}{|\lambda_1|}$ each iteration.

4. Consider the "single pass algorithm" for a non-Hermitian $m \times n$ matrix $\mathbf{A}$, the essentials of which are reiterated below. Suppose our matrix $\mathbf{T}$ is such that $\mathbf{Q}^*\mathbf{Y} = \mathbf{T}(\mathbf{W}^*\mathbf{G})$ and $\mathbf{W}^*\mathbf{Z} = \mathbf{T}^*(\mathbf{Q}^*\mathbf{H})$ hold *exactly* (this is not usually the case in practice!). Show that in this case, the output of the algorithm is exact,

i.e. $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^*$.

"Single Pass" RSVD

- Generate random matrices $\mathbf{G} \in \mathbb{R}^{n\times\ell}$ and $\mathbf{H} \in \mathbb{R}^{m\times\ell}$. For this problem, choose $\ell = \min(m,n)$, though in practice we choose $\ell < \min(m,n)$.
- Compute sample matrices $\mathbf{Y} = \mathbf{A}\mathbf{G}$ and $\mathbf{Z} = \mathbf{A}^*\mathbf{H}$.
- Find ON matrices $\mathbf{Q}$ and $\mathbf{W}$ such that $\mathbf{Y} = \mathbf{Q}\mathbf{Q}^*\mathbf{Y}$ and $\mathbf{Z} = \mathbf{W}\mathbf{W}^*\mathbf{Z}$.
- Solve for $\mathbf{T}$ the linear systems $\mathbf{Q}^*\mathbf{Y} = \mathbf{T}(\mathbf{W}^*\mathbf{G})$ and $\mathbf{W}^*\mathbf{Z} = \mathbf{T}^*(\mathbf{Q}^*\mathbf{H})$.
- Factor $\mathbf{T}$ so that $\mathbf{T} = \hat{\mathbf{U}}\mathbf{D}\hat{\mathbf{V}}^*$.
- Form $\mathbf{U} = \mathbf{Q}\hat{\mathbf{U}}$ and $\mathbf{V} = \mathbf{W}\hat{\mathbf{V}}$.

Solution: For now, assume $n \leq m$ so that $\mathbf{G} \in \mathbb{R}^{n\times n}$. From the various matrix definitions in the algorithm, we have that

$$\mathbf{A}\mathbf{G} = \mathbf{Y} = \mathbf{Q}\mathbf{Q}^*\mathbf{Y} = \mathbf{Q}\mathbf{T}\mathbf{Q}^*\mathbf{G} = \mathbf{U}\mathbf{D}\mathbf{V}^*\mathbf{G}.$$

$\mathbf{G}$ is $n \times n$ and as discussed in lecture, its columns are therefore linearly independent with probability 1. So $\mathbf{G}^{-1}$ exists and therefore $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^*$. If instead $m \leq n$, then we simply follow an analogous path using $\mathbf{H}$, $\mathbf{W}$, *etc.*