University of Edinburgh
INFR11156: Algorithmic Foundations of Data Science (2025)
Solutions 2

Problem 1: Compute the right-singular vectors v_i, the left-singular vectors u_i, the singular values σ_i and hence find the *Singular value decomposition* of

1. $A = \begin{pmatrix} 1 & 1 \\ 0 & 3 \\ 3 & 0 \end{pmatrix}$;

2. $A = \begin{pmatrix} 0 & 2 \\ 2 & 0 \\ 1 & 3 \\ 3 & 1 \end{pmatrix}$.

Solution: Throughout the solution we will make use of the following lemma.

Lemma 1. *Let a and b be two real numbers satisfying $a^2 + b^2 = 1$ and $a \geq 0$. The product ab is maximised when $a = b = \frac{\sqrt{2}}{2}$.*

Proof. Using the initial conditions we can rewrite $a = \sqrt{1 - b^2}$. Hence maximising the product ab reduces to maximising the function $f(x) = x\sqrt{1 - x^2}$. A point x_0 maximises $f(x)$ if $x_0 \geq 0$ and $f'(x_0) = 0$. We have that

$$f'(x) = \sqrt{1 - x^2} + \frac{-x^2}{\sqrt{1 - x^2}} = \frac{1 - 2x^2}{\sqrt{1 - x^2}}.$$

We conclude that $x_0 = \frac{\sqrt{2}}{2}$ which gives $a = b = \frac{\sqrt{2}}{2}$. □

1. For finding the first right-singular vector v_1, we look at any vector $v = \begin{pmatrix} a \\ b \end{pmatrix}$ such that $\|v\| = 1$ and v maximises $\|Av\|$. Without loss of generality we can also assume that $a \geq 0$. Firstly, note that maximising $\|Av\|$ is equivalent to maximising $\|Av\|^2$. We also have that:

$$\|Av\|^2 = \left\| \begin{pmatrix} a + b \\ 3b \\ 3a \end{pmatrix} \right\|^2 = (a + b)^2 + 9b^2 + 9a^2.$$

Since $\|v\| = 1$, we have that $a^2 + b^2 = 1$. Therefore $\|Av\|^2 = 10(a^2 + b^2) + 2ab = 10 + 2ab$. We see that $\|Av\|^2$ is maximised if and only if ab is maximised. Using Lemma 1 that happens when $a = b = \frac{1}{\sqrt{2}}$. So the first right-singular vector $v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and the first singular value is $\sigma_1 = \|Av_1\| = \sqrt{11}$. For the first left-singular vector u_1 we compute

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{22}} \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}.$$

For the second right-singular vector v_2, we look at vectors $v = \begin{pmatrix} a' \\ b' \end{pmatrix}$ such that $\|v\| = 1$, $v \perp v_1$ and v maximises $\|Av\|$. Without loss of generality we can assume $a' \geq 0$. Since $v \perp v_1$ this implies that $a' + b' = 0$. Solving $a'^2 + b'^2 = 1$ gives us that $a' = \frac{1}{\sqrt{2}}$. Hence $v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Moreover, the second singular value is $\sigma_2 = \|Av_2\| = 3$. The second left-singular vector u_2 is computed by

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

The singular value decomposition of A is

$$A = UDV^T = \begin{pmatrix} \frac{2}{\sqrt{22}} & 0 \\ \frac{3}{\sqrt{22}} & \frac{-1}{\sqrt{2}} \\ \frac{3}{\sqrt{22}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{11} & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

2. Again, for finding v_1 we look at any vector $v = \begin{pmatrix} a \\ b \end{pmatrix}$ such that $\|v\| = 1$ and v maximises $\|Av\|$. Without loss of generality we can assume $a \geq 0$.

$$\|Av\|^2 = \left\| \begin{pmatrix} 2b \\ 2a \\ a+3b \\ 3a+3b \end{pmatrix} \right\|^2 = 4a^2 + 4b^2 + (a+3b)^2 + (3a+b)^2 = 14(a^2+b^2) + 12ab.$$

Using that $\|v\| = 1$ we have that $\|Av\|^2 = 14 + 12ab$ which, by Lemma 1, is maximised for $a = b = \frac{1}{\sqrt{2}}$. Therefore we have that $v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\sigma_1 = \sqrt{20}$. The first left-singular vector u_1 is given by

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{2\sqrt{10}} \begin{pmatrix} 2 \\ 2 \\ 4 \\ 4 \end{pmatrix} = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix}.$$

A similar reasoning to the previous part tells us that $v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\sigma_2 = \|Av_2\| = \sqrt{8}$. We also have that

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{4} \begin{pmatrix} -2 \\ 2 \\ -2 \\ 2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}.$$

Hence, the singular value decomposition of A is

$$A = UDV^T = \begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{1}{2} \\ \frac{1}{\sqrt{10}} & \frac{1}{2} \\ \frac{2}{\sqrt{10}} & -\frac{1}{2} \\ \frac{2}{\sqrt{10}} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \sqrt{20} & 0 \\ 0 & \sqrt{8} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Problem 2: Consider the matrix

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 2 \\ 1 & -2 \\ -1 & -2 \end{pmatrix}.$$

1. Run the *power method* starting from $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ for $k = 3$ steps. What does this give as estimates for v_1 and σ_1?
2. What are the actual values of v_i's, σ_i's and u_i's? You might find it helpful to first compute the eigenvalues and eigenvectors of $B = A^\intercal A$.

Solution:

1. Recall that the power method computes a sequence of vectors $\{x_i\}$ such that $x_i = Bx_{i-1}$ for all $1 \le i \le k$, where the matrix $B = A^\intercal A$. In our case we have that

$$B = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 2 & 2 & -2 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 2 \\ 1 & -2 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 16 \end{pmatrix}$$

After $k = 3$ runs of the power method, we obtain a vector

$$x_3 = B^3 x = \begin{pmatrix} 4 & 0 \\ 0 & 16 \end{pmatrix}^3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 64 & 0 \\ 0 & 4096 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 64 \\ 4096 \end{pmatrix}$$

The estimate for v_1 is given by

$$\tilde{v_1} = \frac{x_3}{\|x_3\|} \simeq \begin{pmatrix} 0.0039 \\ 0.9998 \end{pmatrix}.$$

Also, the estimate for σ_1 is given by

$$\tilde{\sigma_1} = \|A\tilde{v_1}\| \simeq 3.9992.$$

2. Since the matrix B is already in diagonal form, its eigenvaues are simply the entries on the diagonal. Thus we have that $\lambda_1 = 16$ and $\lambda_2 = 4$. Recall that the eigenvalues of B are the squares of the singular values of the matrix A, therefore $\sigma_1 = 4$ and $\sigma_2 = 2$. Moreover, we know that the right-singular vectors v_i are the eigenvectors of B corresponding to λ_i. One has that $v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. For the left-singular vectors u_i we compute

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{4} \begin{pmatrix} 2 \\ 2 \\ -2 \\ -2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

and

$$u_1 = \frac{1}{\sigma_2} A v_2 = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

Problem 3: Let $v \in \mathbb{R}^n$ such that $\|v\| = 1$. Sample uniformly $x \in \{-1, 1\}^n$, and define $S = \langle x, v \rangle$. Prove that

$$\mathbf{E}\left[S^4\right] = 3\sum_{i=1}^{n} v_i^2 - 2\sum_{i=1}^{n} v_i^4 \leq 3.$$

Solution: We have that

$$\begin{aligned}
\mathbf{E}\left[\,S^4\,\right] &= \mathbf{E}\left[\left(\sum_{i=1}^{n} x_i v_i\right)^4\right] \\
&= \mathbf{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} x_i x_j x_k x_\ell v_i v_j v_k v_\ell\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} \mathbf{E}\left[\,x_i x_j x_k x_\ell\,\right] v_i v_j v_k v_\ell \\
&= \sum_{i=1}^{n} \mathbf{E}\left[\,x_i^4\,\right] v_i^4 + \frac{1}{2}\binom{4}{2}\sum_{i \neq j} \mathbf{E}\left[\,x_i^2 x_j^2\,\right] v_i^2 v_j^2 \\
&= \sum_{i=1}^{n} v_i^4 + 3\sum_{i \neq j} v_i^2 v_j^2 \\
&= 3\left(\sum_{i=1}^{n} v_i^2\right)\left(\sum_{j=1}^{n} v_j^2\right) - 2\sum_{i=1}^{n} v_i^4 \\
&= 3\,\|v\|^4 - 2\sum_{i=1}^{n} v_i^4 \\
&\leq 3.
\end{aligned}$$

In the third line we used the linearity of the expectation. The equality in the fourth line comes from the fact that under expectation, all products of x_i's vanish when at least one factor has odd power. Finally the last inequality comes from the fact that we chose v to be a unit vector.

Problem 4: Let $A \in \mathbb{R}^{n \times n}$ be a symmetric and PSD matrix. Show that the power method can be applied to approximately compute the smallest eigenvalue of A.

Solution: Suppose A has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, counting multiplicities. First, we can run the power method to find a good approximation of the largest eigenvalue of A, say d is the approximated largest eigenvalue of λ_1. Using this, we can upper bound λ_1 by a constant, say $2d$. Consider the matrix $B = 2D - A$, where D is a diagonal matrix with each diagonal entry being equal to d. Notice that this ensures that matrix B is a PSD matrix. We claim that for every eigenvalue λ_i of A with corresponding eigenvector v_i, $2d - \lambda_i$ is an eigenvalue of B. Indeed we have that

$$Bv_i = (2D - A)v_i = 2Dv_i - Av_i = 2dv_i - \lambda_i v_i = (2d - \lambda_i)v_i.$$

Also note that the smallest eigenvalue of A, i.e. λ_n, corresponds to the largest eigenvalue of B, which is $2d - \lambda_n$. Hence we can run the power method for B to get an estimate for $2d - \lambda_n$ and subtract it from d to get an estimate of λ_n.

Problem 5: Let u be a fixed vector. Show that maximising $x^\intercal u u^\intercal (1 - x)$ subject to $x_i \in \{0, 1\}$ is equivalent to partitioning the coordinates of u into two subsets where the sum of the elements in both subsets are as equal as possible.

***Solution*:** Suppose that the vectors x and u are n-dimensional. Let $f(x) = x^\intercal u u^\intercal (1 - x)$. We have that

$$\begin{aligned} f(x) &= \left(\sum_{i=1}^{n} x_i u_i\right)\left(\sum_{j=1}^{n} u_j(1-x_j)\right) \\ &= \sum_{i,j=1}^{n} x_i(1-x_j)u_i u_j \\ &= \sum_{i:x_i=1}\sum_{j:x_j=0} x_i(1-x_j)u_i u_j \\ &= \left(\sum_{i:x_i=1} u_i\right)\left(\sum_{j:x_j=0} u_j\right). \end{aligned}$$

Let $a = \left(\sum_{i:x_i=1} u_i\right)$ and $b = \left(\sum_{j:x_j=0} u_j\right)$. Note that $a + b = \sum_{i=1}^{n} u_i = c$ for some constant c since the vector u is fixed. Therefore, the problem of maximising $f(x)$ subject to x, is equivalent to maximising the product ab, subject to the constraint $a + b = c$. A similar argument as the one in Lemma 1 of Problem 1 can be used to show that ab is maximised for $a = b$. In our case a and b take discrete values over the random sampling of x, hence $f(x)$ is maximised when $|a - b|$ is minimised. In other words, when we can partition the entries of u into two sets such that the sum of entries in the two sets is as equal as possible.