MTH786, Semester A, 2023/24

Solutions of coursework 1 **N. Perra**

Problem 1. Let A be a 2 matrix

$$A = \begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix}.$$

(i) Find eigenvalues, eigenvectors and eigenvalue decomposition of matrix A.

(ii) Let $\vec{x}$ be a two-dimensional column-vector. Write the product $A\vec{x}$ in terms of eigenvectors of matrix A.

(iii) Find singular values, right and left singular vectors and singular value decomposition of matrix A.

(iv) Let $\vec{x}$ be a two-dimensional column-vector. Write the product $A\vec{x}$ in terms of singular vectors of matrix A.

Solutions:

(i) The eigenvalues of matrix A can be found by solving $\det(A - \lambda I) = 0$. In our case one has

$$\det(A - \lambda I) = \det \begin{pmatrix} 3-\lambda & 4 \\ 0 & 5-\lambda \end{pmatrix} = (3-\lambda)(5-\lambda) = 0 \Rightarrow \lambda_{1,2} = 3, 5.$$

Corresponding eigenvectors can be found by solving $Au = \lambda u$ for the values of λ found above.

$$Au^{(1)} = 3u^{(1)} \Rightarrow \begin{cases} 3u_1^{(1)} + & 4u_2^{(1)} = 3u_1^{(1)} \\ & 5u_2^{(1)} = 3u_2^{(1)} \end{cases} \Rightarrow u^{(1)} = (1, 0)^T.$$

$$Au^{(2)} = 5u^{(1)} \Rightarrow \begin{cases} 3u_1^{(2)} + & 4u_2^{(2)} = 5u_1^{(2)} \\ & 5u_2^{(2)} = 5u_2^{(2)} \end{cases} \Rightarrow u^{(2)} = (2, 1)^T.$$

One can build an eigenvalue decomposition of a matrix A by writing

$$A = Q\Lambda Q^{-1},$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix whose diagonal elements are just eigenvalues of matrix A and Q is the matrix whose i-th column is an eigenvector corresponding to λ_i. In our case one has

$$\Lambda = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad Q^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

Therefore, the eigenvalue decomposition of matrix A has a form

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

(ii) Let v be an arbitrary vector. Then, because eigenvalues of matrix A are linearly independent, one can find α and β such that

$$v = \alpha u^{(1)} + \beta u^{(2)}. \tag{1}$$

Matrix A when applied to vector v produces

$$Av = \alpha \lambda_1 u^{(1)} + \beta \lambda_2 u^{(2)}.$$

We are only left with finding coefficients α, β. By multiplying vector 1 once by $u^{(1)}$ and once by $u^{(2)}$ we can get

$$\begin{cases} \alpha \left\| u^{(1)} \right\|^2 + \beta \left\langle u^{(1)}, u^{(2)} \right\rangle & = \left\langle v, u^{(1)} \right\rangle \\ \alpha \left\langle u^{(1)}, u^{(2)} \right\rangle + \beta \left\| u^{(2)} \right\|^2 & = \left\langle v, u^{(2)} \right\rangle \end{cases}.$$

If one introduces the so-called Gram matrix

$$\Gamma = Q^T Q \begin{pmatrix} \left\| u^{(1)} \right\|^2 & \left\langle u^{(1)}, u^{(2)} \right\rangle \\ \left\langle u^{(1)}, u^{(2)} \right\rangle & \left\| u^{(2)} \right\|^2 \end{pmatrix},$$

then the solution of above system of equations can be written as

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \Gamma^{-1} \begin{pmatrix} \left\langle v, u^{(1)} \right\rangle \\ \left\langle v, u^{(2)} \right\rangle \end{pmatrix}. \tag{2}$$

Combining the above, one can write

$$Av = \lambda_1 \left(\Gamma^{-1} \begin{pmatrix} \left\langle v, u^{(1)} \right\rangle \\ \left\langle v, u^{(2)} \right\rangle \end{pmatrix} \right)_1 u^{(1)} + \lambda_2 \left(\Gamma^{-1} \begin{pmatrix} \left\langle v, u^{(1)} \right\rangle \\ \left\langle v, u^{(2)} \right\rangle \end{pmatrix} \right)_2 u^{(2)}.$$

Remark: You may see that the expression above is similar to the one in case of Singular Value Decomposition (SVD), but is much more complicated. The problem here is the non-orthogonality of eigenvectors, which is a consequence of matrix A being non-symmetric. If matrix A would be symmetri then most likely Γ would be just a diagonal matrix, that is easy to invert.

(iii) The singular values of matrix A are defined as positive square roots of eigenvalues of $A^T A$. Let

$$B := A^T A = \begin{pmatrix} 9 & 12 \\ 12 & 41 \end{pmatrix}.$$

Then corresponding singular values could be found by solving

$$\det \left(B - \sigma^2 I \right) = \det \begin{pmatrix} 9 - \sigma^2 & 12 \\ 12 & 41 - \sigma^2 \end{pmatrix} = \left(\sigma^2 - 41 \right) \left(\sigma^2 - 9 \right) - 144 = 0.$$

Solutions are then given by $\sigma = \sqrt{5}, 3\sqrt{5}$. Right sigular vectors are found by solving $A^T A u = \sigma^2 u$, while left ones are found by solving $A A^T v =$

$\sigma^2 v$, for each value of σ. Normalised singular vectors are then equal to:

$$
\begin{aligned}
A^T A u^{(1)} &= 5u^{(1)} \Rightarrow u^{(1)} = \left(-\frac{3\sqrt{10}}{10} \; \frac{\sqrt{10}}{10}\right)^T, \\
A^T A u^{(2)} &= 45u^{(2)} \Rightarrow u^{(2)} = \left(\frac{\sqrt{10}}{10} \; \frac{3\sqrt{10}}{10}\right)^T, \\
A A^T v^{(1)} &= 5v^{(1)} \Rightarrow v^{(1)} = \left(-\frac{\sqrt{2}}{2} \; \frac{\sqrt{2}}{2}\right)^T, \\
A A^T v^{(2)} &= 45v^{(2)} \Rightarrow v^{(2)} = \left(\frac{\sqrt{2}}{2} \; \frac{\sqrt{2}}{2}\right)^T.
\end{aligned}
$$

Using the above we can rewrite the matrix A in a form of SVD as

$$
A = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 \\ 0 & 3\sqrt{5} \end{pmatrix} \begin{pmatrix} -\frac{3\sqrt{10}}{10} & \frac{\sqrt{10}}{10} \\ \frac{\sqrt{10}}{10} & \frac{3\sqrt{10}}{10} \end{pmatrix}.
$$

(iv) Finally, for any vector w one can write

$$
Aw = \sqrt{5}v^{(1)} \left\langle w, u^{(1)} \right\rangle + 3\sqrt{5}v^{(2)} \left\langle w, u^{(2)} \right\rangle.
$$

Problem 2. Let the function $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$
f\left(\mathbf{x} = (x_1, x_2)^\top\right) = \frac{1}{2} \langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle + \langle \mathbf{u}, \mathbf{x} \rangle,
$$

where $\mathbf{A}$ is a real, symmetric, positive definite 2×2 matrix, and $\mathbf{u}$ is a real vector of length 2. In other words

$$
f\left(\mathbf{x} = (x_1, x_2)^\top\right) = \frac{1}{2} a x_1^2 + b x_1 x_2 + \frac{1}{2} c x_2^2 + v x_1 + w x_2,
$$

where

$$
\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \text{ and } \mathbf{u} = (v, w)^\top.
$$

Our goal is to find such a vector $\mathbf{x}^*$ that minimises $f(\mathbf{x})$.

1. Show that the gradient ∇f is given by $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{u}$.
2. Thus proof that the gradient is zero at $\mathbf{x}^* = -\mathbf{A}^{-1}\mathbf{u}$.
3. Evaluate $f(\mathbf{x}^*)$.
4. Calculate the Hessian $H_f(\mathbf{x}^*)$ and show it is positive. This will finish the proof of $\mathbf{x}^*$ being a minimizer point of $f(\mathbf{x})$.

Problem 3. Consider the following function of real arguments ω_0, ω_1.

$$
f(\omega_0, \omega_1) = \frac{1}{2s} \sum_{k=1}^{s} (\omega_0 + \omega_1 x_k - y_k)^2, \tag{3}
$$

where $x_1, x_2, \ldots, x_s$ and $y_1, y_2, \ldots, y_s$ are real-valued constants. This function measures the mean squared error of a linear approximation for the data points $\{(x_i, y_i)\}_{k=1}^{s}$.

(i) Find partial derivatives $\frac{\partial}{\partial\omega_0}f(\omega_0,\omega_1)$, $\frac{\partial}{\partial\omega_1}f(\omega_0,\omega_1)$, $\frac{\partial^2}{\partial\omega_0^2}f(\omega_0,\omega_1)$, $\frac{\partial^2}{\partial\omega_0\partial\omega_1}f(\omega_0,\omega_1)$, $\frac{\partial^2}{\partial\omega_1^2}f(\omega_0,\omega_1)$.

Hint: you should be able to show that

$$\frac{\partial}{\partial\omega_0}f(\omega_0,\omega_1) = \omega_0 + \overline{\mathbf{x}}w_1 - \overline{\mathbf{y}}, \qquad \frac{\partial}{\partial\omega_1}f(\omega_0,\omega_1) = \overline{\mathbf{x}}\omega_0 + \overline{\mathbf{x^2}}w_1 - \overline{\mathbf{xy}}.$$

where

$$\overline{\mathbf{x}} = \frac{1}{s}\sum_{k=1}^{s} x_k, \quad \overline{\mathbf{x^2}} = \frac{1}{s}\sum_{k=1}^{s} x_k^2, \quad \overline{\mathbf{y}} = \frac{1}{s}\sum_{k=1}^{s} y_k, \quad \overline{\mathbf{xy}} = \frac{1}{s}\sum_{k=1}^{s} x_k y_k.$$

(ii) Using the above results, find values ω_0^*, ω_1^* such that

$$\nabla f(\omega_0^*, \omega_1^*) = 0$$

. These should be functions of $x_1, x_2, \ldots, x_s$ and $y_1, y_2, \ldots, y_s$.

Hint: you should obtain

$$\omega_0^* = \frac{\overline{\mathbf{y}}\cdot\overline{\mathbf{x^2}} - \overline{\mathbf{x}}\cdot\overline{\mathbf{xy}}}{\overline{\mathbf{x^2}} - \overline{\mathbf{x}}^2}, \quad \omega_1^* = \frac{\overline{\mathbf{xy}} - \overline{\mathbf{x}}\cdot\overline{\mathbf{y}}}{\overline{\mathbf{x^2}} - \overline{\mathbf{x}}^2}.$$

(iii) Using the expressions of second order derivatives obtained in (i) find the value of Hessian of the function f for $\omega_0 = \omega_0^*$ and $\omega_1 = \omega_1^*$. Prove it is positive definite and thus show that $f(\omega_0,\omega_1)$ attains its minimum value at $\omega_0 = \omega_0^*$ and $\omega_1 = \omega_1^*$.

(iv) Find $\min\limits_{\omega_0,\omega_1\in\mathbb{R}} f(\omega_0,\omega_1)$.

Solutions:

(i) `Partial derivatives are equal to`

$$\begin{aligned}
\frac{\partial f}{\partial\omega_0} &= \frac{1}{s}\sum_{k=1}^{s}(\omega_0 + \omega_1 x_k - y_k) = \omega_0 + \omega_1\overline{\mathbf{x}} - \overline{\mathbf{y}}, \\
\frac{\partial f}{\partial\omega_1} &= \frac{1}{s}\sum_{k=1}^{s} x_k(\omega_0 + \omega_1 x_k - y_k) = \omega_0\overline{\mathbf{x}} + \omega_1\overline{\mathbf{x^2}} - \overline{\mathbf{xy}}, \\
\frac{\partial^2 f}{\partial\omega_0^2} &= \left(\frac{\partial}{\partial\omega_0}\omega_0 + \omega_1\overline{\mathbf{x}} - \overline{\mathbf{y}}\right) = 1, \\
\frac{\partial^2 f}{\partial\omega_0\partial\omega_1} &= \frac{\partial}{\partial\omega_1}(\omega_0 + \omega_1\overline{\mathbf{x}} - \overline{\mathbf{y}}) = \overline{\mathbf{x}}, \\
\frac{\partial^2 f}{\partial\omega_1^2} &= \frac{\partial}{\partial\omega_1}\left(\omega_0\overline{\mathbf{x}} + \omega_1\overline{\mathbf{x^2}} - \overline{\mathbf{xy}}\right) = \overline{\mathbf{x^2}},
\end{aligned}$$

`where we have introduced the following averages`

$$\overline{\mathbf{x}} = \frac{1}{s}\sum_{k=1}^{s} x_k, \quad \overline{\mathbf{x^2}} = \frac{1}{s}\sum_{k=1}^{s} x_k^2, \quad \overline{\mathbf{y}} = \frac{1}{s}\sum_{k=1}^{s} y_k, \quad \overline{\mathbf{xy}} = \frac{1}{s}\sum_{k=1}^{s} x_k y_k.$$

(ii) Extremal points can be found by solving $\nabla f\left(\omega_0^*, \omega_1^*\right) = 0$. This is equivalent to

$$\begin{cases} \omega_0^* + & \omega_1^* \overline{\mathbf{x}} = \overline{\mathbf{y}} \\ \omega_0^* \overline{\mathbf{x}} + & \omega_1^* \overline{\mathbf{x^2}} = \overline{\mathbf{xy}} \end{cases}.$$

Solution of the above system of linear equations is given by

$$\omega_0^* = \frac{\overline{\mathbf{y}} \cdot \overline{\mathbf{x^2}} - \overline{\mathbf{x}} \cdot \overline{\mathbf{xy}}}{\overline{\mathbf{x^2}} - \overline{\mathbf{x}}^2}, \quad \omega_1^* = \frac{\overline{\mathbf{xy}} - \overline{\mathbf{x}} \cdot \overline{\mathbf{y}}}{\overline{\mathbf{x^2}} - \overline{\mathbf{x}}^2}.$$

(iii) The Hessian of function f is equal to

$$H_f\left(\omega_0, \omega_1\right) = \begin{pmatrix} 1 & \overline{\mathbf{x}} \\ \overline{\mathbf{x}} & \overline{\mathbf{x^2}} \end{pmatrix}.$$

The 2×2 matrix is positive definite if and only if its both top left element and the determinant are positive. For the above Hessian we thus only need to check its determinant. It is equal to

$$\det H_f\left(\omega_0, \omega_1\right) = \overline{\mathbf{x^2}} - \overline{\mathbf{x}}^2,$$

which is non-negative due to AM-QM inequality. This proves that f attains its minimum at $\left(\omega_0^*, \omega_1^*\right)$.

Remark: the above Hessian doesn't depend on values of ω_0, ω_1 and is non-nega
This proves that the function f is convex. We will discuss convexity
in details in Weeks 2-3 as this plays a crucial role in optimisation problem

(iv) Using the above notations we can rewrite function f as

$$f\left(\omega_0, \omega_1\right) = \frac{1}{2}\left(\omega_0^2 - 2\omega_0 \overline{\mathbf{y}} + \overline{\mathbf{y^2}} + \omega_1^2 \overline{\mathbf{x^2}} + 2\omega_0\omega_1 \overline{\mathbf{x}} - 2\omega_1 \overline{\mathbf{xy}}\right).$$

Plugging the values we have found above one obtains (after some simple algebraic manipulations)

$$f\left(\omega_0^*, \omega_1^*\right) = \frac{1}{2} \frac{\left(\overline{\mathbf{x^2}} - \overline{\mathbf{x}}^2\right)\left(\overline{\mathbf{y^2}} - \overline{\mathbf{y}}^2\right) - \left(\overline{\mathbf{xy}} - \overline{\mathbf{x}} \cdot \overline{\mathbf{y}}\right)^2}{\overline{\mathbf{x^2}} - \overline{\mathbf{x}}^2}.$$

The above is the minimal possible value of f.

Problem 4. As you may have seen in the Lecture 2, the function considered in previous question is a mean-squared error function of a linear regression for data set $\{x_k, y_k\}_{k=1}^s$. Corresponding values ω_0^*, ω_1^* are the coefficients of a linear regression model for data set $\{x_k, y_k\}_{k=1}^s$. Using the above result find the linear regression, i.e. coefficients ω_0^*, ω_1^* for the following people's height/weight data.

Weight	162.31	183.93	154.34	187.50	187.06	173.42
Height	68.78	68.79	68.50	68.62	68.25	68.49

Now let us add one more data point and recalculate the regression.

Weight	162.31	183.93	154.34	187.50	187.06	173.42	192.34
Height	68.78	68.79	68.50	68.62	68.25	68.49	68.14

Can you explain the origin of such a difference between two results?

Solutions: Let us use the formula from previous question to find an equation of linear regression. As we have previously seen we just need to calculate corresponding averages and plug their values into the formulas. The calculation is easy to present in the form of a table:

							Average
Weight	162.31	183.93	154.34	187.50	187.06	173.42	174.76
Height	68.78	68.79	68.50	68.62	68.25	68.49	68.57
Height^2	4730.69	4732.06	4692.25	4708.70	4658.06	4690.88	4702.11
Height x Weight	11163.68	12652.54	10572.29	12866.25	12766.85	11877.54	11983.19

Corresponding linear regression coefficients could be then evaluated as

$$\omega_0 = 950.07, \quad \omega_1 = -11.31.$$

After one more point is added, we obtain a new table

								Average
Weight	162.31	183.93	154.34	187.50	187.06	173.42	192.34	177.27
Height	68.78	68.79	68.50	68.62	68.25	68.49	68.14	68.51
Height^2	4730.69	4732.06	4692.25	4708.70	4658.06	4690.88	4643.06	4693.67
Height x Weight	11163.68	12652.54	10572.29	12866.25	12766.85	11877.54	13106.05	12143.60

with corresponding linear regression coefficients

$$\omega_0 = 1825.70, \quad \omega_1 = -24.06.$$

We observe a dramatic change in the slope and shift coefficients. This is due to the fact that all samples present people with approximately the same height. Thus the denominator in our formulas for ω_0, ω_1 becomes small and the resulting values of parameters become unstable to a small change in data. We will discuss the ways to overcome such a problem later in the module.