MA398 MATRIX ANALYSIS AND ALGORITHMS: ASSIGNMENT 2

Please submit your solutions to this assignment via Moodle by **noon on Monday October 24$^{\text{th}}$**. Make sure that your submission is clearly marked with your name, university number, course and year of study. Note that some of the questions are practical questions that are concerned with real-world problems and require coding in Python.

- The written part of the solutions may preferably be typed in LaTeX, otherwise written on paper and subsequently scanned/photographed provided the images are clearly legible. You are required to deliver a single document entitled *MA398_Assignment1_FirstnameLastname.pdf*, outlining your solutions and explaining your interpretation and arguments to the questions .
- The Python code scripts relevant to each question should be submitted as *MA398_ Assignment1_ ExerciseN.ipynb* where you need to make sure that you document the environment in which I should be able to run your code.

To avoid losing marks unnecessarily, please make sure that,

- you show all intermediate steps of your calculations.
- you include all the coding that you have to do in order to obtain your answers and according to the instruction given above.
- include any plots you generate and label them appropriately.
- when you provide an answer to a question make sure that you justify your answer and provide details of any mathematical calculations that are required.

⚠Only in an emergency or if the Moodle submission is unavailable because of a general outage, the assignment should in the respective case be submitted by email to Randa.Herzallah@warwick.ac.uk, and olayinka.ajayi@warwick.ac.uk.

---

1. (Singular value decompositions).

   (a) Consider a non zero matrix $A \in \mathbb{R}^{n\times m}$ with rank r. The singular value decomposition of the matrix A is given by $A = U\Sigma V^T$ where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with entries $\sigma_1 > \sigma_2 > \cdots > \sigma_r$. Denoting by $u_i, i = 1,\ldots,m$ the column vectors of U and by $v_i, i = 1,\ldots,n$ the column vectors of v, show that,

$$Av_i = \left\{ \begin{array}{ll} \sigma_i u_i, & \text{for } i = 1,2,\ldots,r \\ 0, & \text{for } i = r+1, r+2,\ldots,m \end{array} \right\}$$

$$A^T u_i = \left\{ \begin{array}{ll} \sigma_i v_i, & \text{for } i = 1,2,\ldots,r \\ 0, & \text{for } i = r+1, r+2,\ldots,n \end{array} \right\}$$

   Answer: Since V is an orthogonal matrix we have that $V^T V = I$, hence

$$AV = U\Sigma$$

   Also from the definition of the SVD we have that,

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{n\times m}$$

and $\sigma_1 > \sigma_2 > \cdots > \sigma_r$. Hence writing U in terms of its columns,

$$U = \left(\begin{array}{c|c|c|c} u1 & u_2 & \dots & u_n \end{array} \right)$$

we get,

$$\begin{aligned} U\Sigma &= \left(\begin{array}{c|c|c|c} u_1 & u_2 & \dots & u_n \end{array} \right) \begin{pmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0 \end{pmatrix} \\ &= \left(\begin{array}{c|c|c|c|c|c|c} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & 0u_{r+1} & \dots & 0 \end{array} \right). \end{aligned}$$

Similarly, writing AV in terms of the columns of V we get,

$$AV = \left(\begin{array}{c|c|c|c} Av_1 & Av_2 & \dots & Av_m \end{array} \right)$$

Since these are equal, comparing the columns we have,

$$Av_i = \left\{ \begin{array}{ll} \sigma_i u_i, & \text{for } i = 1, 2, \dots, r \\ 0, & \text{for } i = r+1, r+2, \dots, m \end{array} \right\}$$

as claimed.
By transposing $A = U\Sigma V^T$ we get,

$$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$$

Since U is orthogonal we have $U^T U = I$. Thus,

$$A^T U = V\Sigma^T$$

Following the same argument as above we get,

$$\begin{aligned} A^T U &= \left(\begin{array}{c|c|c|c} A^T u_1 & A^T u_2 & \dots & A^T u_n \end{array} \right) \\ &= \left(\begin{array}{c|c|c|c|c|c|c} \sigma_1 v_1 & \sigma_2 v_2 & \dots & \sigma_r v_r & 0u_{r+1} & \dots & 0 \end{array} \right) \\ &= V\Sigma^T \end{aligned}$$

Comparing the columns gives us that,

$$A^T u_i = \left\{ \begin{array}{ll} \sigma_i v_i, & \text{for } i = 1, 2, \dots, r \\ 0, & \text{for } i = r+1, r+2, \dots, n \end{array} \right\}$$

as stated.

(b) Let $A \in \mathbb{R}^{n\times m}$, $n \geq m$, $\operatorname{rank}(A) = m$ with singular value decomposition $A = U\Sigma V^T$.

  i. Determine the singular value decomposition of the matrix $(A^T A)^{-1}$.

  ii. Hence evaluate its Euclidean norm $\|(A^T A)^{-1}\|_2$.

Answer: Noting that by assumption A is full rank, means that $A^T A$ is invertible, thus $(A^T A)^{-1}$ is well defined.
Let us first evaluate $A^T A$. Since $A = U\Sigma V^T$, we have,

$$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$$

as the singular value decomposition of A^T. This in turn yields,

$$\begin{aligned} A^T A &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T \Sigma V^T \end{aligned}$$

as the singular value decomposition of $A^T A$.
Note that,
$$\Sigma^T\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2) \in \mathbb{R}^{m\times m}$$
which implies that,
$$(\Sigma^T\Sigma)^{-1} = \mathrm{diag}(\sigma_1^{-2}, \sigma_2^{-2}, \dots, \sigma_m^{-2}) \in \mathbb{R}^{m\times m}$$
Hence,
$$\begin{aligned}(A^T A)^{-1} &= (V\Sigma^T\Sigma V^T)^{-1}\\ &= (V^T)^{-1}(\Sigma^T\Sigma)^{-1}V^{-1}\\ &= V(\Sigma^T\Sigma)^{-1}V^T\end{aligned}$$
However, this is not yet the singular value decomposition of $(A^T A)^{-1}$ since the singular values are currently in ascending order, as opposed to descending. Thus, taking a matrix $P \in \mathbb{R}^{m\times m}$ to be an antidiagonal matrix where all the entries are zero except those on the diagonal going from the lower left corner to the upper right corner being 1, which is orthogonal,
$$P = \begin{pmatrix} 0 & 0 & \dots & 0 & 1\\ 0 & 0 & \dots & 1 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & 0 & \dots & 0 & 0\end{pmatrix}.$$
Thus $P^T = P$, and $P^T P = PP = I$. Using this P matrix we can write,
$$\begin{aligned}(A^T A)^{-1} &= VPP^T(\Sigma^T\Sigma)^{-1}PP^TV^T\\ &= VP(P^T(\Sigma^T\Sigma)^{-1}P)(VP)^T\end{aligned}$$
where,
$$(P^T(\Sigma^T\Sigma)^{-1}P) = \mathrm{diag}(\sigma_m^2, \sigma_{m-1}^2, \dots, \sigma_1^2)$$
and VP is orthogonal since both V and P are orthogonal. Therefore, this is the desired singular value decomposition since the singular values are now in descending order.

2. (Polynomial (Over-)Fitting). Polynomial models can be used to provide predictions for a large number of real world problems. Here the main objective is to identify the coefficients $a_0, a_1, \dots, a_d$ of a polynomial function of the form,
$$y = f(x,a) = \sum_{i=0}^{d} a_i x^i + \epsilon \tag{1}$$
where ϵ is a random component that represents the residual error, and d is the degree of the polynomial. Given $n+1$ data points we can write a linear system of equations as follows,
$$\underbrace{\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_{n+1}\end{pmatrix}}_{b} = \underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^d\\ 1 & x_2 & x_2^2 & \dots & x_2^d\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_{n+1} & x_{n+1}^2 & \dots & x_{n+1}^d\end{pmatrix}}_{V}\underbrace{\begin{pmatrix} a_0\\ a_1\\ \vdots\\ a_d\end{pmatrix}}_{a}$$
The matrix V is called a Vandermonde matrix. In first year linear algebra, you probably proved the qualitative result that V is invertible if and only if $n = d+1$ and all the x_i are distinct.

(a) Numerically analyse the invertibility of the Vandermonde matrix through the following steps,

- Write a Python function Vandermonde(x,d) which assembles the above matrix, where `d = size(x)-1`, resulting in a square matrix.
- Generate equidistant points in $[-1,1]$, and numerically calculate the condition number $\kappa(V)$ of V for several representative values $d = n \leq 100$ in some matrix norm.

(b) Assuming $d = 3$, and $n = 5$, formulate the least square solution for the identification of the polynomial coefficients.

Answer: To find the optimal values of the model parameters, we define the sum of square error function by summing over all the data points and their corresponding output values,

$$E = \frac{1}{2}\sum_{l=1}^{n}[f(\mathbf{x}_l, a) - y_l]^2. \tag{2}$$

Using Equation (1) in (2), the sum of square error function can be written in the form,

$$E = \frac{1}{2}\sum_{l=1}^{n}\left\{\sum_{i=0}^{d} a_i x_l^i - y_l\right\}^2. \tag{3}$$

Rewrite the above equation in a matrix form,

$$E = \frac{1}{2}\|Va - b\|_2^2, \tag{4}$$

where, V, a, and b being as defined above. Now evaluating the derivative with respect to the unknown parameters a yields the normal equation,

$$V^T Va = V^t b$$

(c) This part of the question is concerned with predicting weather conditions using polynomial models. In particular given daily measurements of pressure and humidity provided in the file "WeatherData.xlsx" the objective is to identify a polynomial model that can predict the humidity based on pressure measurements. Use all data from day 1 to day 90 to identify the coefficients of your polynomial model and use everything from day 90 as test data.

i. Plot the humidity measurements against their corresponding pressure measurements. (Think of axes labels, scale of the axes, fontsize, plotmarkers, etc.)

ii. Starting from $d = 0$ to $d = 5$ identify the coefficients of the 6 polynomial models and state their values.

iii. Compute the relative forward errors of the polynomial models you identified in part ii and then and plot them against their corresponding polynomial order.

iv. Compute the relative forward errors of the polynomial models you identified in part ii on the test data and then and plot them against their corresponding polynomial order.

v. What is the optimal order of polynomial that fits your weather data. Justify your answer.

vi. What are the best estimates for the last five days calculated by your best polynomial model?