

# Numerical Linear Algebra Homework Project 2: Least Squares, Orthogonalization, and the SVD

Catalano Giuseppe, Cerrato Nunzia

## Problem 1

(1) Suppose we are given  $m$  pairs of data points,  $(x_1, y_1), \dots, (x_m, y_m)$ . We want to find a linear combination of prescribed functions  $\phi_1, \dots, \phi_n$  whose values at the points  $x_i \in [a, b]$ ,  $1 \leq i \leq m$ , approximate the values  $y_1, \dots, y_m$  as well as possible. More precisely, the problem is to find a function of the form  $f(x) = \alpha_1 \phi_1(x) + \dots + \alpha_n \phi_n(x)$  such that

$$\sum_{i=1}^m [y_i - f(x_i)]^2 \leq \sum_{i=1}^m [y_i - g(x_i)]^2 \quad \forall g \in \text{Span}(\phi_1, \dots, \phi_n), \quad (1)$$

where, usually,  $m > n$ . It is possible to rephrase the problem as:

$$f = \arg \min_{f \in \text{Span}(\phi_1, \dots, \phi_n)} \sum_{i=1}^m [y_i - f(x_i)]^2. \quad (2)$$

Now we can define a column vector  $\mathbf{z} \in \mathbb{R}^n$  such that:

$$[\mathbf{z}]_i = \alpha_i \quad (3)$$

and a matrix  $A$  such that:

$$[A\mathbf{z}]_i = f(x_i) = \alpha_1 \phi_1(x_i) + \dots + \alpha_n \phi_n(x_i). \quad (4)$$

In this way, the element of the  $i$ -th row and  $j$ -th column of the matrix  $A$  is:

$$[A]_{ij} = a_{ij} = \phi_j(x_i). \quad (5)$$

Finally, defining a column vector  $\mathbf{b} \in \mathbb{R}^n$  such that:

$$[\mathbf{b}]_i = y_i \quad (6)$$

we can rewrite the (2) as follows:

$$\tilde{\mathbf{z}} = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{z}\|_2^2 = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{z}\|_2, \quad (7)$$

where the function  $f$  can be built from  $\tilde{\mathbf{z}}$ .

(2) Now we suppose to take  $\phi_k = x^{k-1}$ ,  $1 \leq k \leq n$ . Under this assumption, the matrix  $A$  takes the form:

$$A = \begin{bmatrix} x_1^0 & \dots & x_1^{n-1} \\ \vdots & \ddots & \vdots \\ x_m^0 & \dots & x_m^{n-1} \end{bmatrix}. \quad (8)$$

We want to prove that, assuming that  $x_i \neq x_j$  for  $i \neq j$ ,  $A$  has full rank:  $\text{rank}(A) = n$ .

Proof: Proving that  $\text{rank}(A) = n$  is equivalent to prove that  $\dim(\ker(A)) = 0$ , that means

that  $\nexists \mathbf{v} \in \mathbb{R}^n$  s.t.  $\mathbf{v} \in \ker(A)$ . We want to prove this statement by contraddiction, therefore, we look for a vector  $\mathbf{v} \in \mathbb{R}^n$ , with  $\mathbf{v} \neq \underline{0}$ , such that  $A\mathbf{v} = \underline{0}$ , that means:

$$\begin{cases} v_1 x_1^0 + \dots + v_n x_1^{n-1} = 0 \\ \vdots \\ v_1 x_m^0 + \dots + v_n x_m^{n-1} = 0 \end{cases} \quad (9)$$

Defining the polynomial

$$p_{\mathbf{v}}^{(n-1)}(x) = \sum_{i=1}^n v_i x^{i-1} \quad (10)$$

we can observe that, for any choice of  $\mathbf{v} \neq \underline{0}$ ,  $p_{\mathbf{v}}^{(n-1)}(x)$  admits at most  $n - 1$  different roots, therefore  $\nexists \mathbf{v} \neq \underline{0}$  such that  $A\mathbf{v} = \underline{0}$ . This concludes the proof.  $\square$

**(3)** Consider the problem of finding the best fit with a quadratic function  $f(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2$  for the following data:

$x_i$	8	10	12	16	20	30	40	60	100
$y_i$	0.88	1.22	1.64	2.72	3.96	7.66	11.96	21.56	43.16

In the following we report the code that solves the normal equations:

$$A^T A \mathbf{v} = A^T \mathbf{b} \quad (11)$$

using the Cholesky factorization algorithm and then compares the result with the one found using the QR factorization of the matrix  $A$ .

\*\*\* INSERIRE CODICE \*\*\*

Using the previous functions we have computed the solution to the minimization problem, obtaining the following results:

Cholesky:  $x = [-1.91914925269909, 0.278213536291725, 0.001739400875055]$

QR factorization:  $x = [-1.91914925269904, 0.278213536291722, 0.001739400875055]$

From these results we can observe that, for this problem, both the algorithms perform in a similar way. In fact, the results differ at most in the 15th digit. In figure 1 we show the input data and the solutions to the least square problem.

**(4)** The following code computes the residual  $\mathbf{r} = \mathbf{d} - C\hat{\mathbf{x}}$  where  $\hat{\mathbf{x}} = [-1.919, 0.2782, 0.001739]$  is the approximate solution of the least squares problem.

\*\*\*INSERIRE CODICE\*\*\*

The results that we obtain are:

Residual:  $\mathbf{r} = [0.009503999999993, 0.716895999999906, 62.090847999905236]$

Norm 2 of the residual:  $\|\mathbf{r}\|_2 = 62.09498720144942$

The value of the residual may seem strange, in fact, if we compute the relative error we find:

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} = 7.728184292875672\text{e-}05. \quad (12)$$

However, we can observe that, from the relation

$$\frac{1}{k_2(C)} \frac{\|\mathbf{r}\|_2}{\|\mathbf{d}\|_2} \leq \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq k_2(C) \frac{\|\mathbf{r}\|_2}{\|\mathbf{d}\|_2} \quad (13)$$

we can obtain the following relation for  $\mathbf{r}$ :

$$\frac{\|\mathbf{d}\|_2}{k_2(C)} \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq \|\mathbf{r}\|_2 \leq k_2(C) \|\mathbf{d}\|_2 \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2}. \quad (14)$$

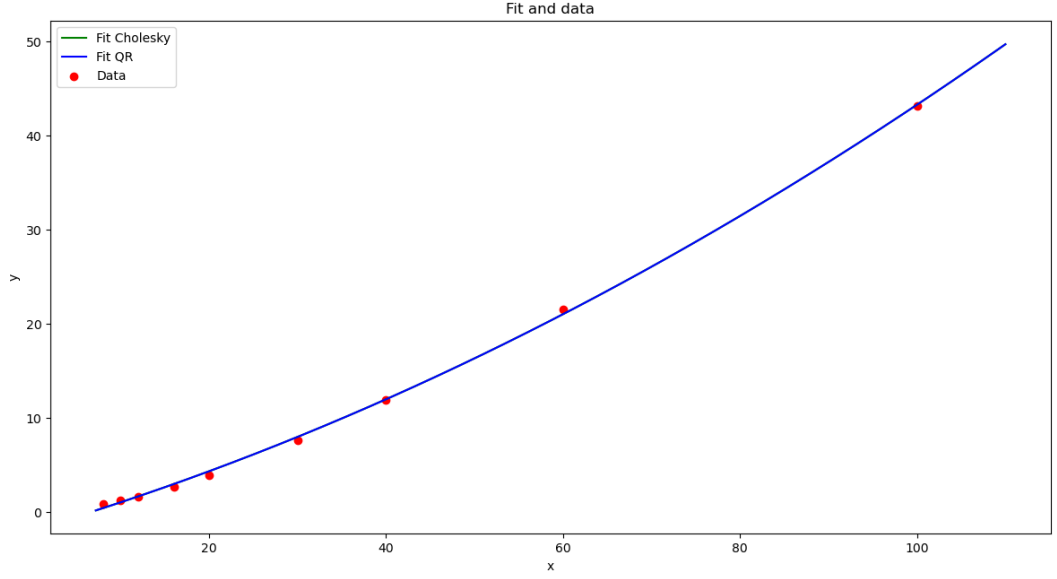


Figure 1

With the following python code we have computed the bounds to the residual. \*\*\* INSERIRE CODICE\*\*\*

From the output of the previous code we know that:

$$5 \times 10^{-7} \lesssim \|\mathbf{r}\|_2 \lesssim 3 \times 10^9 \quad (15)$$

this big range is due to the value of  $k_2(C) \simeq 8 \times 10^7$  that suggests us that we should not use the residual to measure the accuracy of the solution when the problem is ill-conditioned as in this case.

## Problem 2

(1) Let  $A \in \mathbb{R}^{m \times n}$ , with  $\text{rank}(A) = n$ , let  $A = QR$  be the (full) QR factorization of  $A$ , with  $Q \in \mathbb{R}^{m \times m}$  orthogonal and  $R \in \mathbb{R}^{m \times n}$  upper trapezoidal. Also, let  $A = Q_1 R_1$  be the reduced QR factorization of  $A$  with  $Q_1 \in \mathbb{R}^{m \times n}$  having orthonormal columns and  $R_1 \in \mathbb{R}^{m \times m}$  upper triangular. Show that  $R_1$  is nonsingular, and that the columns  $\mathbf{q}_1, \dots, \mathbf{q}_n$  of  $Q_1$  form an orthonormal basis for  $\text{Ran}(A)$ , the column space of  $A$ . Also, find an orthonormal basis for  $\text{Null}(A^T)$ , the null space of  $A^T$ .

We start showing that  $R_1$  is nonsingular. Since  $A$  has full rank, we know that:

$$A\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x} = \mathbf{0} \quad (16)$$

and multiplying both sides for  $Q_1^T$  knowing that  $Q_1^T Q_1 = I_n$  we obtain

$$R_1 \mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x} = \mathbf{0}, \quad (17)$$

that concludes the proof.

Now we want to show that the columns  $\mathbf{q}_1, \dots, \mathbf{q}_n$  of  $Q_1$  form an orthonormal basis for  $\text{Ran}(A)$ . We start observing that

$$\forall \mathbf{y} \in \text{Ran}(A) \exists \mathbf{x} \in \mathbb{R}^n : \mathbf{y} = A\mathbf{x} = Q_1 R_1 \mathbf{x} \quad (18)$$

and we know this from the definition of range of a matrix. In a similar way, knowing that  $R_1$  is nonsingular and therefore is a bijective map from  $\mathbb{R}^n$  to  $\mathbb{R}^n$ , we can put  $\mathbf{x}' = R_1\mathbf{x}$  and say that:

$$\forall \mathbf{y} \in \text{Ran}(A) \exists \mathbf{x} \in \mathbb{R}^n : \mathbf{y} = A\mathbf{x} = Q_1\mathbf{x}', \quad (19)$$

that means that  $\text{Ran}(A) = \text{Ran}(Q_1) = \text{Span}\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ .  
Now we want to find an orthonormal basis for  $\text{Null}(A^T)$

### Problem 3

(1) Let  $A \in \mathbb{R}^{m \times n}$ , with singular value decomposition  $A = U\Sigma V^T = \sum_{i=1}^n \sigma_i \mathbf{u}_i \mathbf{v}_i^T$  with  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$  and  $\text{rank}(A) = n$ . Express the singular values and singular vectors of the following matrices in terms of those of  $A$ .

(a)  $(A^T A)^{-1}$

$$(A^T A)^{-1} = (V^T)^{-1} \Sigma^{-2} V^{-1} = V \Sigma^{-2} V^{-1}$$

- Both left and right singular vectors of  $(A^T A)^{-1}$  are equal to the right singular vectors of  $A$ ;
- Singular values of  $(A^T A)^{-1}$  are equal to singular values of  $A$  raised to the  $-2$  power.

(b)  $(A^T A)^{-1} A^T$

$$(A^T A)^{-1} A^T = V \Sigma^{-2} V^T V \Sigma U^T = V \Sigma^{-1} U^{-1}$$

- Left singular vectors of  $(A^T A)^{-1} A^T$  are equal to the right singular vectors of  $A$ ;
- Right singular vectors of  $(A^T A)^{-1}$  are equal to the left singular vectors of  $A$ ;
- Singular values of  $(A^T A)^{-1}$  are equal to the inverse of singular values of  $A$ .

(c)  $A(A^T A)^{-1}$

$$A(A^T A)^{-1} = U \Sigma V^T V \Sigma^{-2} V^{-T} = U \Sigma^{-1} V^T$$

- Left and right singular vectors of  $(A^T A)^{-1}$  coincide with that of  $A$ ;
- Singular values of  $(A^T A)^{-1}$  are equal to the inverse of singular values of  $A$ .

(d)  $A(A^T A)^{-1} A^T$

$$A(A^T A)^{-1} A^T = U \Sigma (V^T) V \Sigma^{-2} V^T V \Sigma U^T = \mathbb{1}$$

(2)

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \quad (20)$$

$$A^T A = \begin{bmatrix} 1 & 2 \\ 2 & 8 \end{bmatrix} \quad (21)$$

$$\det(A^T A - \lambda \mathbb{1}) = 0 \quad (22)$$

$$\lambda_{1,2} = \frac{9 \pm \sqrt{65}}{2} \quad (23)$$

$$k_2(A) = \frac{\sigma_{\max}}{\sigma_{\min}} = \sqrt{\frac{9 + \sqrt{65}}{9 - \sqrt{65}}} \quad (24)$$