

# CS6210 - Homework/Assignment-4

Arnab Das(u1014840)

November 6, 2016

---

**1: Chapter-6, question-4**

---

(a) The techniques discussed in this chapter are for polynomial data fitting and not exponential data fitting, hence cannot be applied directly to  $u(t)$ .

(b) Given approximation of the form

$$u(t) = \gamma_1 \exp(\gamma_2 t)$$

and provided data points as  $(t_1, z_1), (t_2, z_2), \dots, (t_m, z_m)$ , where  $z_i > 0, i = 1, 2, \dots, m$  and  $m > 0$ . Considering instead the following approximation:

$$v(t) = \ln u(t) = (\ln \gamma_1) + \gamma_2 t$$

such that the data points become  $(t_1, b_1), (t_2, b_2), \dots, (t_m, b_m)$  where  $b_i = \ln z_i$  and

$$v(t) = x_1 + x_2 t$$

such that  $x_1 = \ln \gamma_1$  and  $x_2 = \gamma_2$  and  $v(t) = \ln u(t)$

Then for solving  $A^T A x = A^T b$ , we define the following matrices:

$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \dots & \dots \\ 1 & t_m \end{bmatrix} \text{ leading to } B = A^T A = \begin{bmatrix} \sum_{i=1}^m 1 = m & \sum_{i=1}^m t_i \\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2 \end{bmatrix} \text{ and } b = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_m \end{bmatrix} \text{ leading to } A^T b = \begin{bmatrix} \sum_{i=1}^m b_i \\ \sum_{i=1}^m t_i b_i \end{bmatrix} \text{ Thus solving}$$

$$B = A^T A x = A^T b$$

we get the following two equations:

$$x_1 + x_2 = 1 \tag{1}$$

and

$$3x_1 + 5x_2 = 4.9 \tag{2}$$

Solving (i) and (ii), we get:  $x_1 = \ln \gamma_1 = 0.5$  and  $x_2 = \gamma_2 = 0.95$

Thus, we have,

$$\begin{aligned} v(t) &= \ln u(t) = \ln(0.5) + 0.95t \\ \Rightarrow u(t) &= 0.5 \times \exp(0.95t) \end{aligned}$$

(Answer).

---

**2: Chapter-6, question-5**

---

(a) For tall and skinny matrices,  $A$ , of the system of equations,  $Ax = b$ , the number of rows is much larger than the number of columns. Let the number of rows be  $m$  and the number of columns  $n$ , then in such cases generally  $m \gg n$ . These systems are overdetermined and  $b$  is generally not in the range space of  $A$ . Thus applying LU does not give a unique solution to  $Ax=b$ . Instead, a better way to approach such problems is to minimize the residual  $\|b - Ax\|$ , such that within the tolerance of the residual we have the best solution for  $x$ . When transforming to the normal equation,  $A^T A x = A^T b$ , here one can use LU decomposition since we have  $n \times n$  matrix, however cholesky decomposition is a better choice here since the matrix  $A^T A$  is symmetric positive definite. In case of the QR decomposition, which has an upper triangular part in  $R$ , however the  $Q$  matrix allows us to extract the upper-triangular system of equation whose solution leads to a solution of the least square problem. Here, the orthonormal behaviour of  $Q$  is used to transform into

equivalent set of equations such that the **norms are not affected**. Finally, SVD is used more in cases where A is rank deficient or nearly rank deficient, in which cases LU cannot be used. Thus, in all the cases LU directly doesn't fit the scenario for application except that LU requires to be slightly modified so that for every column it zeroes out, the vector of its remaining elements is orthonormal to the previous columns. With this introduction of orthonormality, it can be used in QR decomposition.

(b) In the normal equation: the way condition number came to be  $K(B) = K^2(A)$ . This was because of the following derivation:

In normal equation we were solving:

$$A^T A x = A^T b$$

and

$$(A^T A)^{-1} A^T = V S^{-2} V^T V \sum^T U^T = V (S^{-1} 0) U^T$$

where  $A = U \sum V^T$ , V and U are orthogonal matrices and  $\sum$  is a diagonal matrix with singular values of A along the diagonal, and hence  $A^T A = V \sum U^T U \sum V^T = V \sum^2 V^T$ . Thus  $\|(A^T A)^{-1} A^T\| = \|(S^{-1} 0)\|$  and  $\|((A^T A)^{-1} A^T)^{-1}\| = \|V(S 0) U^T\|$ . Thus the condition number becomes:

$$K(B = A^T A) = \frac{\lambda_1}{\lambda_n} = \frac{\sigma_1^2}{\sigma_n^2} = K^2(A)$$

Now, for QR, we write  $x = (A^T A)^{-1} A^T b$ , and we have represented  $A=QR$  by a QR decomposition where Q is orthonormal and of same dimension as A. The following translation results in the final form:

$$x = (A^T A)^{-1} A^T b = (R^T Q^T Q R)^{-1} R^T Q^T b = (R^T R)^{-1} R^T Q^T b = R^{-1} Q^T b$$

Hence, we get the following relation,

$$x = R^{-1} Q^T b$$

Notice, that multiplication by orthogonal matrices do not affect the norms, thus,

$$\|R^{-1} Q^T\| = \|R^{-1}\| = \|A^{-1}\|$$

and also,

$$\|(Q^T)^{-1} R\| = \|R\| = \|A\|$$

Thus the condition number in this case comes to be  $\|A\| \|A^{-1}\| = K(A)$ .

The main saving comes due to usage of the orthonormal decomposition for the transformations, since the transformations using q only transform them in space without affecting the norm values, while for the previous normal equation, the singular values were getting multiplied during creation  $A^T A$ , introducing change in norm values during the transformations.

### 3: Chapter-8, question-8

(a) Given a rank deficient matrix, we analyse here its effect on the Gram-Schmidt process. Consider the span of three vectors  $v_1, v_2$  and  $v_3$ , such that  $v_3$  is linearly dependent on  $v_1$  and  $v_2$ . We can write  $v_3$  as a linear combination of  $v_1$  and  $v_2$ , as:

$$\mathbf{v}_3 = a\mathbf{v}_1 + b\mathbf{v}_2$$

Let  $\mathbf{u}_i$  denote the orthonormal unit vectors we generate along the Gram-Schmidt process. Then, for the first vector,  $\mathbf{v}_1$ , we have

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}$$

And the projection of  $\mathbf{v}_2$  on  $\mathbf{v}_1$  is

$$\mathbf{y}_2 = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1) \mathbf{u}_1$$

Then, the next orthonormal vector,  $\mathbf{u}_2$  will be along  $\mathbf{y}_2$ , such that  $\mathbf{u}_2 = \frac{\mathbf{y}_2}{|\mathbf{y}_2|}$ .

Now the projection of  $\mathbf{v}_3$ , say  $\mathbf{y}_3$ , on the span( $\mathbf{u}_1, \mathbf{u}_2$ ), will be,

$$\mathbf{y}_3 = \mathbf{v}_3 - \left( (\mathbf{v}_3 \cdot \mathbf{u}_1) \mathbf{u}_1 + (\mathbf{v}_3 \cdot \mathbf{u}_2) \mathbf{u}_2 \right)$$

$$\mathbf{y}_3 = \mathbf{v}_3 - \left( ((\mathbf{a}\mathbf{v}_1 + \mathbf{b}\mathbf{v}_2) \cdot \mathbf{u}_1) \mathbf{u}_1 + ((\mathbf{a}\mathbf{v}_1 + \mathbf{b}\mathbf{v}_2) \cdot \mathbf{u}_2) \mathbf{u}_2 \right)$$

$$\mathbf{y}_3 = \mathbf{v}_3 - \left( \mathbf{a}\mathbf{v}_1 + (\mathbf{b}\mathbf{v}_2 \cdot \mathbf{u}_1) \mathbf{u}_1 + (\mathbf{a}\mathbf{v}_1 \cdot \mathbf{u}_2) \mathbf{u}_2 + (\mathbf{b}\mathbf{v}_2 \cdot \mathbf{u}_2) \mathbf{u}_2 \right)$$