Problem set 2: Solutions

Workout 0.1 The definition of condition number for an square and nonsingular matrix is $\mathrm{cond}(A) = \|A\| \cdot \|A^{-1}\|$. First we compute A^{-1}:

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 3 & -1 & 4 \end{bmatrix} \quad \Longrightarrow \quad A^{-1} = \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 2.5 & -0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix}$$

Then we have (please refer to definition of norms in Lecture 5)

$$\begin{aligned} \mathrm{cond}_\infty(A) &= \|A\|_\infty \|A^{-1}\|_\infty = 8 \times 3.5 = 28 \\ \mathrm{cond}_1(A) &= \|A\|_1 \|A^{-1}\|_1 = 6 \times 4.5 = 27 \\ \mathrm{cond}_F(A) &= \|A\|_F \|A^{-1}\|_F \doteq 5.83 \times 3.20 \doteq 18.67 \end{aligned}$$

Workout 0.2 (1) First form the matrix A and then $A^T A$ and $A^T \boldsymbol{y}$:

$$A = \begin{bmatrix} 1 & 9 \\ 1 & 10 \\ 1 & 11 \\ 1 & 12 \\ 1 & 13 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 5 & 55 \\ 55 & 615 \end{bmatrix}, \quad A^T \boldsymbol{y} = \begin{bmatrix} 18 \\ 188 \end{bmatrix}$$

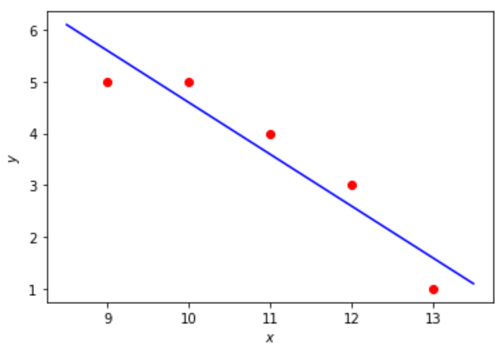
The normal equation then is:

$$\begin{bmatrix} 5 & 55 \\ 55 & 615 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 18 \\ 188 \end{bmatrix}.$$

Solving this equation (for example using `numpy.linalg.solve`) gives $\boldsymbol{a} = [a_0, a_1]^T = [14.6, -1]^T$. Plug in to the ansatz:

$$p_1(x) = a_0 + a_1 x = 14.6 - x$$

The plot of data and the least squares line:



(2) The residual vector is

$$\boldsymbol{e} = \boldsymbol{y} - A\boldsymbol{a} = \begin{bmatrix} 5 \\ 5 \\ 4 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 9 \\ 1 & 10 \\ 1 & 11 \\ 1 & 12 \\ 1 & 13 \end{bmatrix} \begin{bmatrix} 14.6 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.6 \\ 0.4 \\ 0.4 \\ 0.4 \\ -0.6 \end{bmatrix}$$

The residual vector tells that the vertical distance between the value of first data point i.e. 5.0 and its corresponding predicted value i.e. $p_1(9) = a_0 - 9a_1$ is -0.6. Similarly, for the second data point the error is 0.4, and so on. The norm 2 of residual vector is $residual = \|\boldsymbol{e}\|_2 \doteq 1.0954$. The is the minimum possible norm 2 we can obtain.

(3) The figure illustrates two basis vectors, $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, representing the columns of matrix A, along with vector $\boldsymbol{y}$ which serves as the right-hand side vector. In the context of the least squares problem, our objective is to find the best (closest) approximation for $\boldsymbol{y}$ within the subspace defined by the columns of A (depicted as the gray plane in the figure). When measuring distance using the 2 norm (Euclidean distance), the closest vector corresponds to the red vector. This vector is chosen such that the difference between $\boldsymbol{y}$ and the chosen vector is perpendicular to the plane, giving rise to the normal equation $A^T A\boldsymbol{x} = \boldsymbol{y}$. The dashed red line in the figure represents the residual vector. For a detailed mathematical formulation, refer to the lecture notes (Lecture 6).

Workout 0.3

1. The mean of data is $\overline{x} = (9 + 10 + 11 + 12 + 13)/5 = 11$, so the centred (shifted) data are

x	-2	-1	0	1	2
y	5.0	5.0	4.0	3.0	1.0

We form the new matrices:

$$A = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}, \quad A^T \boldsymbol{y} = \begin{bmatrix} 18 \\ -10 \end{bmatrix}$$

The normal equation then is:

$$\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 18 \\ -10 \end{bmatrix}.$$

Solving this equation gives $\boldsymbol{a} = [a_0, a_1]^T = [3.6, -1]^T$. Plug in to the ansatz:

$$p_1(x) = a_0 + a_1(x - 11) = 3.6 - (x - 11) = 14.6 - x$$

which is the same as before.

2. We compute the condition numbers in norm 2 with command `np.linalg.cond(A.T@A,2)`. (The default syntax computes the condition number in norm Frobenius). We get:

$$\text{cond}_2(A^T A) = 7686.0, \quad \text{cond}_2(A^T_{shift} A_{shift}) = 2.0$$

Observe the significant difference!

3. The rule of thumb for linear systems says:

$$(\textit{relative error in solution}) \approx (\textit{condition number}) \times (\textit{relative error in data})$$

In these cases we have

$$\begin{aligned} \text{Original data}: &\ (\textit{relative error in } \boldsymbol{a}) \approx 7686 \cdot 10^{-3} = 7.686 \\ \text{Centred data}: &\ (\textit{relative error in } \boldsymbol{a}) \approx 2 \cdot 10^{-3} = 0.002 \end{aligned}$$

The percentage of error in the first case is about 768.6% while for the second case it is about 0.2%.

Workout 0.4 The matrix A for the quadratic ansatz is $A = [\mathbf{1}, \boldsymbol{x}, \boldsymbol{x}^2]$ (columns). For the shifted data above we have

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}, \quad A^T \boldsymbol{y} = \begin{bmatrix} 18 \\ -10 \\ 32 \end{bmatrix}$$

The normal equation then is:

$$\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 18 \\ -10 \\ 32 \end{bmatrix}.$$

Workout 0.5 For linear regression we compute the reduced QR factorization using the following commands:

```
A = np.array([[1,9],[1,10],[1,11],[1,12],[1 ,13]])
Q, R = np.linalg.qr(A, mode='reduced')
print('Q = ', np.round(Q,2))
print('R = ', np.round(R,2))
```

The outputs are:

$$Q = \begin{bmatrix} -0.45 & -0.63 \\ -0.45 & -0.32 \\ -0.45 & 0.00 \\ -0.45 & 0.32 \\ -0.45 & 0.63 \end{bmatrix}, \quad R = \begin{bmatrix} -2.24 & -24.6 \\ 0. & 3.16 \end{bmatrix}$$

Then we form and solve triangular system $R\boldsymbol{x} = Q^T\boldsymbol{y}$ (Here Q and R are just reduced forms).

```
y = np.array([5,5,4,3,1])
a = scipy.linalg.solve_triangular(R,Q.T@y, lower=False)
print('a = ', np.round(a,2))
```

The solution is `a = [14.6, -1.0]`.

Part 2 is similar, just replace the matrix A by the shifted matrix.

Workout 0.6 Use the second column of $A^{(1)}$ and let

$$\boldsymbol{x} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

to get

$$\boldsymbol{u} = \boldsymbol{x} + \operatorname{sign}(x_1)\|\boldsymbol{x}\|_2\boldsymbol{e}_1 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} + \sqrt{9+1+4}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \doteq \begin{bmatrix} 6.742 \\ 1 \\ 2 \end{bmatrix}$$

The (3×3) Householder matrix corresponding to this $\boldsymbol{u}$ is:

$$\tilde{H} = I - \frac{2}{\boldsymbol{u}^T\boldsymbol{u}}\boldsymbol{u}\boldsymbol{u}^T \doteq \begin{bmatrix} -0.802 & -0.267 & -0.535 \\ -0.267 & 0.96 & -0.079 \\ -0.535 & -0.079 & 0.841 \end{bmatrix}.$$

The full Householder matrix is:

$$H = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{H} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.802 & -0.267 & -0.535 \\ 0 & -0.267 & 0.96 & -0.079 \\ 0 & -0.535 & -0.079 & 0.841 \end{bmatrix}.$$

Finally, we multiply H by $A^{(1)}$ to get:

$$A^{(2)} = HA^{(1)} \doteq \begin{bmatrix} 2. & 1. & 3. \\ 0. & -3.742 & 2.673 \\ 0. & -0. & 1.545 \\ 0. & -0. & -2.91 \end{bmatrix}$$

In Python we can write:

```
x = np.array([3.0,1.0,2.0])
u = x
u[0] = x[0] + np.linalg.norm(x,2)
Ht = np.eye(3) - 2/np.dot(u,u)*np.outer(u,u)
H = np.block([[np.ones(1),np.zeros(3)],[np.zeros((3,1)), Ht]])
print('A2 =', np.round(H@A1,3))
```

Be careful to define the array `x` with at least one floating point number such as `3.0` instead of `3`. Otherwise `u` would be an integer array.

Non-mandatory workouts

Workout 0.7 We assume that the molecular weight of nitrogen is x and that of oxygen is y. We have 6 equations form the table:

$$\begin{cases} x + y & = 30.006 \\ 2x + y & = 44.013 \\ x + 2y & = 46.006 \\ 2x + 3y & = 76.012 \\ 2x + 5y & = 108.010 \\ 2x + 4y & = 92.001 \end{cases} \implies \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \\ 2 & 3 \\ 2 & 5 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 30.006 \\ 44.013 \\ 46.006 \\ 76.012 \\ 108.010 \\ 92.001 \end{bmatrix}$$

Using normal equations: the normal system $A^T A\boldsymbol{x} = A^T \boldsymbol{b}$ has the following form:

$$\begin{bmatrix} 18 & 29 \\ 29 & 56 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 716.084 \\ 1302.121 \end{bmatrix}$$

From this we find the molecular weight of

nitrogen: 14.007 g/mol,

oxygen: 15.998 g/mol,

which is really close to the truth.

Using QR factorization: The reduced QR factors of A (from Python) are

$$Q_1 = \begin{bmatrix} -0.236 & -0.201 \\ -0.471 & -0.73 \\ -0.236 & 0.128 \\ -0.471 & -0.073 \\ -0.471 & 0.584 \\ -0.471 & 0.255 \end{bmatrix}, \quad R = \begin{bmatrix} -4.243 & -6.835 \\ 0. & 3.046 \end{bmatrix}$$

The least squares solution then is obtained by solving the triangular system $R_1\boldsymbol{x} = Q_1^T\boldsymbol{b}$:

$$\begin{bmatrix} -4.243 & -6.835 \\ 0. & 3.046 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -168.783 \\ 48.730 \end{bmatrix}$$

The solution is $\boldsymbol{x} = [x, y] = [14.007, 15.998]$ which is consistent with the solution we have already obtained using the normal equation.

Workout 0.8 Please refer to lecture notes (Lecture 6) for an explanation of how this non-linear ansatz can be addressed through linear regression. The approach involves taking the natural logarithm of the data values r and subsequently applying linear regression to the transformed dataset $(t_k, \ln r_k)$:

t	0	0.50	1.00	2.00	4.00
$\ln r$	0.698	0.582	0.351	−0.01	−0.868

The matrix A is:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0.5 \\ 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}.$$

Normal equation $A^T A\boldsymbol{a} = A^T\boldsymbol{b}$ then reads as (here $\boldsymbol{b} = \ln \boldsymbol{r}$)

$$\begin{bmatrix} 5 & 7.5 \\ 7.5 & 21.25 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 0.753 \\ -2.848 \end{bmatrix}$$

Solving this system gives $a_0 \doteq 0.747$ and $a_1 \doteq -0.398$. Finally, $\alpha = e^{a_0} \doteq 2.112$ and $\beta = a_1 = -0.398$ and the model is $r(t) = 2.112 \cdot e^{-0.398t}$.

Workout 0.9 We consider the proposed model $p = c/r^2$ as an ansatz with a single parameter c, applied to the dataset (r_k, p_k) for $k = 1, \ldots, N$, representing N different apartment in Uppsala (distances from center and prices). By introducing the variable change $x = 1/r^2$, we transform the model into the linear form $p(x) = c \cdot x$. Solving the corresponding normal equations for the transformed data pairs (x_k, p_k) results in solution c. Note that our solution vector is reduced to a scalar c in this scenario. Clearly, determining the value of c for Uppsala needs collecting N pairs (r_k, p_k) from the Uppsala housing market.

Workout 0.10.

1. Hint: use the fact that $\|I\| = 1$ (any norm) and complete the proof.
2. From submultiplicity property of a norm we have $1 = \|I\| = \|AA^{-1}\| \leqslant \|A\| \cdot \|A^{-1}\| = \mathrm{cond}(A)$.
3. Hint: use the fact that for $\alpha \neq 0$ we have $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$, and complete the proof.
4. Hint: use the facts that $\|D\| = \max |d_k|$ and $D^{-1} = \mathrm{diag}\{1/d_1, \dots, 1/d_n\}$, and complete the proof.
5. Hint: use the facts that $Q^{-1} = Q^T$ and $\|Q\|_2 = \sqrt{\lambda_{max}(Q^T Q)} = \sqrt{\lambda_{max}(I)} = \sqrt{1} = 1$, and complete the proof.
6. In this case we have $A = QR$ and $A^{-1} = R^{-1}Q^{-1} = R^{-1}Q^T$. Thus, $\|A\|_2 = \|QR\|_2 = \|R\|_2$ and $\|A^{-1}\|_2 = \|R^{-1}Q^T\|_2 = \|R^{-1}\|_2$. (orthogonal matrices preserve the norm 2). These show that $\mathrm{cond}_2(A) = \mathrm{cond}_2(R)$.