

**Matrices and Statistics with Applications**  
**Computer Assignment**  
**Least Squares and Orthogonal Transformations**

---

**Instructions:** This assignment must be done *individually!* You may of course discuss the assignment but not copy someone else's code.

**Matlab users:** Use the Matlab function `publish` (on the toolbar of the Matlab editor) to prepare the report.

**R users:** Use the reporting tool of R.

**Purpose:** The assignment is an introduction to basic matrix concepts and to using orthogonal transformations and decompositions.

---

1. Solve the following least squares problems using the normal equations and QR decomposition. In both cases, plot the data and the polynomial approximation.

- (a) Fit a third degree polynomial to the data

$$\begin{array}{c|cccc} t & 1 & 2 & 3 & 4 \\ \hline f(t) & 1 & 2 & 2 & 4 \end{array}$$

- (b) Fit a third degree polynomial to the data

$$\begin{array}{c|ccccc} t & 1 & 2 & 3 & 4 & 5 \\ \hline f(t) & 1 & 2 & 2 & 4 & 5 \end{array}$$

Explain the differences between (a) and (b).

2. Fit the third degree polynomial  $P(t) = a_0 + a_1t + a_2t^2 + a_3t^3$  to the following data, using normal equations.

$$\begin{array}{c|cccc} t & 501 & 502 & 503 & 504 \\ \hline f(t) & 1 & 2 & 2 & 4 \end{array}$$

What is the condition number of the matrix in the normal equations? Use a better model and compute the condition number of the normal equations.

3. (a) Compute the solution of the least squares problem

$$\min_{\beta} \|X\beta - y\|, \quad X = \begin{pmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

for the values  $\epsilon = 1, 10^{-1}, \dots, 10^{-8}, 10^{-9}$  using QR decomposition and normal equations. Compare the results. Which are more accurate?

- (b) Illustrate the condition numbers of the data matrix and the normal equations as a function of  $\epsilon$ .

4. At the URL

[http://users.mai.liu.se/larel04/kurser/book\\_weight.data](http://users.mai.liu.se/larel04/kurser/book_weight.data)

there are weight data for 15 books. It is assumed that the following model is valid:

$$w = b_0 + b_1v + b_2a + \eta,$$

where  $w$  is weight,  $v$  is volume, and  $a$  is area of the cover.  $\eta$  is assumed to be a stochastic variable that accounts for measurement errors and the differences in paper quality etc.

- (a) Determine the regression coefficients.
- (b) How much of the variation (percentage) in weight is explained by the model?
- (c) Illustrate how the model fits the data (data points and residual).
- (d) Remove the two observations for which the model fits the worst, and recompute the quantities above.

The data are from the book Maindonald & Braun, Data Analysis and Graphics using R.