SeSE: Matrices and Statistics with Applications

Computer Exercise no. 1: Least Squares and Orthogonal Transformations

September, 2014

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

t	1	2	3	4
$f(t)$	1	2	2	4

   (b) Fit a third degree polynomial to the data

t	1	2	3	4	5
$f(t)$	1	2	2	4	5

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

2. Fit the third degree polynomial $P(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$ to the following data, using normal equations.

t	501	502	503	504
$f(t)$	1	2	2	4

   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

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

for the values $\epsilon = 1, 10^{-1}, \ldots, 10^{-8}, 10^{-9}$ 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 ϵ.

4. At the URL
`http://www.mai.liu.se/~laeld/kurser/Sese-comp-stat/data/`
there are weight data for 15 books. It is assumed that the following model is valid:

$$w = b_0 + b_1 v + b_2 a + \eta,$$

where w is weight, v is volume, and a is area of the cover. η 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 ('Goodnes of fit', R^2)?

(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.

5. Write a function `myqr` that computes the QR decomposition of a matrix. Use the functions `househ` and `apphouse` from the lecture handouts. Compare your decomposition to that given by Matlab or R.