MAT4110 Assignment 1

Sigbjørn L. Foss

September 23, 2019

Polynomial fitting with QR decomposition

Given a data set $(\boldsymbol{x}, \boldsymbol{y})$, assumed to be on the form $y_i = y(x_i) + \varepsilon_i$, a polynomial approximation of the function $y(x)$ of order $< m$ can be found by minimizing

$$F(\boldsymbol{\beta}) = ||\boldsymbol{A\beta} - \boldsymbol{y}||_2^2, \tag{1}$$

where $\boldsymbol{A}$ is the Vandermonde matrix,

$$\boldsymbol{A} = \begin{bmatrix} (\boldsymbol{x}^T)^0 & (\boldsymbol{x}^T)^1 & \dots & (\boldsymbol{x}^T)^{m-1}, \end{bmatrix}$$

with exponentials taken element wise in each column vector, and $\boldsymbol{\beta}$ contains the polynomial coefficients. In the lecture notes, it is shown that the solution

$$\boldsymbol{R}_1\boldsymbol{\beta} = \boldsymbol{c}_1 \tag{2}$$

minimizes (1) where $\boldsymbol{A} = \boldsymbol{QR}$, $\boldsymbol{R}_1$ is the upper $m \times m$ partition of $\boldsymbol{R}$ and $\boldsymbol{c}_1$ is the upper m partition of $\boldsymbol{c} = \boldsymbol{Q}^T\boldsymbol{y}$. Equation (2) is solved by back-substitution as $\boldsymbol{R_1}$ is upper triangular.

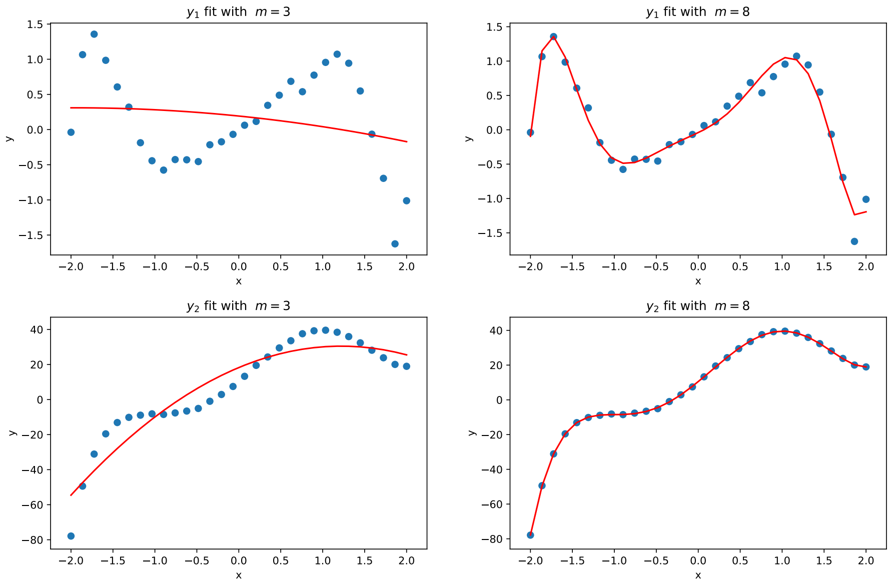

Figure 1: Fitted lines using the QR algorithm

Two data sets are generated by

$$\begin{aligned} y_1 &= x(\cos(\varepsilon + \frac{1}{2}x^3) + \sin(\frac{1}{2}x^3)), \\ y_2 &= 4x^5 - 5x^4 - 20x^3 + 10x^2 + 40x + 10 + \varepsilon, \end{aligned}$$

and fitted by the QR-algorithm for $m = 3$ and $m = 8$ to produce the curves shown in figure 1.

Polynomial fitting with Cholesky decomposition

Minimization of (1) can also be done by calculating $\nabla F(\boldsymbol{\beta}) = 0$. Taking this gradient results in the normal equation

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{A}^T\boldsymbol{y}. \tag{3}$$

The matrix $\boldsymbol{A}^T\boldsymbol{A}$ is positive definite, and its Cholesky decomposition is

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T,$$

where $\boldsymbol{L}$ is lower triangular with ones on the diagonal and $\boldsymbol{D}$ is diagonal. For a positive definite matrix the entries of $\boldsymbol{D}$ are positive and we define

$$\boldsymbol{R} = \boldsymbol{L}\sqrt{\boldsymbol{D}},$$

and the decomposition is

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{R}\boldsymbol{R}^T.$$

The least squares solution of (3) is found by letting $\boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{y}$, solving

$$\boldsymbol{R}\boldsymbol{w} = \boldsymbol{c}$$

by forward substitution and solving

$$\boldsymbol{R}^T\boldsymbol{\beta} = \boldsymbol{w}$$

by backward substitution. Applying the algorithm to y_1 and y_2 results in the curves shown in figure 2.

Comparison of the methods

Consider a perturbation in the data which translates to the least square solution,

$$\boldsymbol{A}(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b},$$
$$\boldsymbol{A}\delta\boldsymbol{x} = \delta\boldsymbol{b}.$$

Assuiming $\boldsymbol{A}$ is invertible, the following holds,

$$||\boldsymbol{A}\boldsymbol{x}|| = ||\boldsymbol{b}|| \leq ||\boldsymbol{A}|| \cdot ||\boldsymbol{x}||$$

$$\Rightarrow \frac{1}{||\boldsymbol{x}||} \leq \frac{||\boldsymbol{A}||}{||\boldsymbol{b}||}, \tag{4}$$

and

$$||\delta\boldsymbol{x}|| = ||\boldsymbol{A}^{-1}\delta\boldsymbol{b}|| \leq ||\boldsymbol{A}^{-1}|| \cdot ||\delta\boldsymbol{b}||. \tag{5}$$

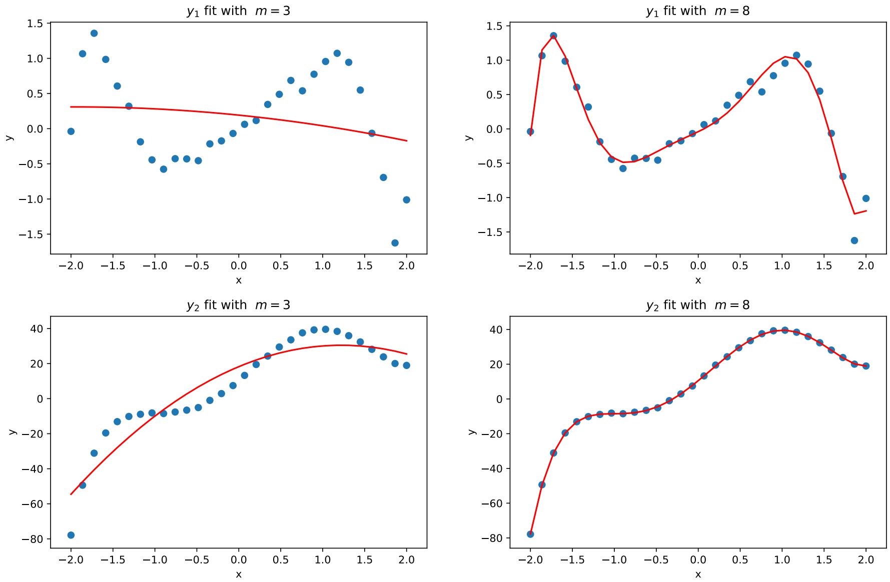

Figure 2: Fitted lines using the Cholesky algorithm

Multiplying (5) by (4),

$$\frac{||\delta \boldsymbol{x}||}{||\boldsymbol{x}||} \leq ||\boldsymbol{A}^{-1}|| \cdot ||\boldsymbol{A}|| \frac{||\boldsymbol{\delta b}||}{||\boldsymbol{b}||} = \mathrm{cond}(\boldsymbol{A}) \frac{||\boldsymbol{\delta b}||}{||\boldsymbol{b}||}. \tag{6}$$

In the QR algorithm, the least-squares problem is reduced to (2), and as $\boldsymbol{Q}$ is orthonormal ,

$$\mathrm{cond}(\boldsymbol{A}) = \mathrm{cond}(\boldsymbol{QR}) = \mathrm{cond}(\boldsymbol{R}),$$

and

$$||\boldsymbol{c}|| = ||\boldsymbol{Q}^T \boldsymbol{y}|| = ||\boldsymbol{y}||.$$

The upper bound of the relative error in the LS solution is scaled by the condition of $\boldsymbol{A}$. The relative error upper bound in the Cholesky method is scaled by

$$\mathrm{cond}(\boldsymbol{X}^T \boldsymbol{X}) = \mathrm{cond}^2(\boldsymbol{X}),$$

meaning that the Cholesky method is potentially much less robust than the QR method. The QR method is, however, generally slower than the Cholesky method as their complexities for large m, n are given by

$$C_{\text{Cholesky}} = \mathcal{O}(nm^2 + \frac{1}{3}m^3)$$
$$C_{\text{QR}} = \mathcal{O}(2nm^2).$$

The QR algorithm is much more stable at a relatively low increase in complexity for most problems, and should generally be preferred over the Cholesky algortihm.

Appendix

The full notebook with the python implementations, figures and a pdf print-out of the code can be found at this link