

MAT3110, Autumn 2019, Compulsory assignment 1

Johan Nereng

Department of Physics, University of Oslo, Norway

Sep 25, 2019

1 Least Squares of a polynomial fit

This assignment focuses on minimizing the sum of squares, S , of errors from a polynomial fit of degree m to a data set $\{x_i, y_i\}_{i=1}^n$. From this data set, a Vandermonde matrix

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{m-1} \\ 2 & x_2 & x_2^2 & \dots & x_2^{m-1} \\ \dots & \dots & \dots & \dots & \dots \\ n & x_n & x_n^2 & \dots & x_n^{m-1} \end{bmatrix} \quad (1)$$

corresponding to the choice of m may be constructed, along with a vector $\mathbf{x} = [c_1, c_2, \dots, c_m]$, containing the coefficients of the polynomial. If $\mathbf{b} = [y_1, y_2, \dots, y_n]$, then $S = \sum_{i=1}^n (p(x) - y_i)^2$ may be expressed as $S = \|\mathbf{Ax} - \mathbf{b}\|_2^2$. In order to minimize S , this assignment utilizes two methods: a QR factorization of A , and solving the normal equations $A^T A \mathbf{x} = A^T \mathbf{b}$ using Cholesky factorization and forward substitution.

1.1 Solving Least Squares with QR factorization

Assuming that the Vandermonde matrix A described earlier has linearly independent columns, it may be factorized as $QR = A$, where Q is an orthogonal \mathbb{R}^n matrix and R an upper \mathbb{R}^n triangular matrix [2]. Additionally, and generally, the vector $\mathbf{x} \in \mathbb{R}^m$ minimizes $\|\mathbf{Ax} - \mathbf{b}\|$ IFF it minimizes $\|\Omega \mathbf{Ax} - \mathbf{b}\|$ for any orthogonal matrix Ω [3]. Using this, the minimization

$$\|\mathbf{Ax} - \mathbf{b}\|^2 = \|Q^T \mathbf{Ax} - Q^T \mathbf{b}\|^2 = \|R\mathbf{x} - Q^T \mathbf{b}\|^2 = \left\| \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{x} - \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right\|^2 \quad (2)$$

Thus solving the minimization corresponds to solving $R_1 \mathbf{x} = c_1$, which, since R_1 is a non-singular, upper triangular matrix, may be accomplished by backward substitution.

1.2 Solving Least Squares with Cholesky factorization and forward substitution

An alternative method for obtaining the coefficients in \mathbf{x} is solving the normal equations using a Cholesky factorization. The 2-norm is a convex function, with a unique minimum. This minimum is found by taking the gradient of the norm, which leads to the normal equation:

$$A^T A \mathbf{x} = A^T \mathbf{b} \quad (3)$$

The matrix $B = A^T A$ is a symmetric $n \times n$ matrix, which allows the application of the Cholesky factorization [2]: $B = R^T R$, where $R \in \mathbb{R}^n$ is a lower triangular matrix. Thus (3) may be written as $RR^T \mathbf{x} = A^T \mathbf{b}$. In order to solve for \mathbf{x} , define $R^T \mathbf{x} = \mathbf{w}$, and solve $R\mathbf{w} = A^T \mathbf{b}$ for \mathbf{w} by forward substitution. $R^T \mathbf{x} = \mathbf{w}$ is then solved by backward substitution.

2 Method implementation

Algorithms [2] [3] for solving the Least Squares fit is implemented using both QR factorization and forward substitution, and Cholesky factorization with forward and backward substitution. For the purpose of this assignment, Python 3.7 using the Numpy library was used. Following the guidelines provided by the professor in class, the code is not included in this assignment.

2.1 Functions for testing the implementation

In order to test the implementation, two function were employed:

$$f_1(x) = x(\cos(r + 0.5x^3) + \sin(0.5x^3)) \quad (4)$$

$$f_2(x) = 4x^5 - 5x^4 - 20x^3 + 10x^2 + 40x + 10 + r \quad (5)$$

where $r = U(0, 1)$.

2.2 Application of methods to test functions

Fitting of polynomials of degree $m = 3$ and $m = 8$, using both $f_1(x)$ and $f_2(x)$ for $x \in [-2, 2]$, are displayed in figure 1. In the plots, the two methods are indistinguishable, indicating similar fitting. At the same time, it is evident that they polynomials do not match the data completely. This is to be expected, as the original functions are trigonometric, which cannot be approximated by m degree polynomials. By-eye evaluation of the polynomials do however suggest that the methods were successfully implemented. More in-depth evaluation could be done using MSE or R^2 , but such evaluation measurements were not used for the purpose of this assignment.

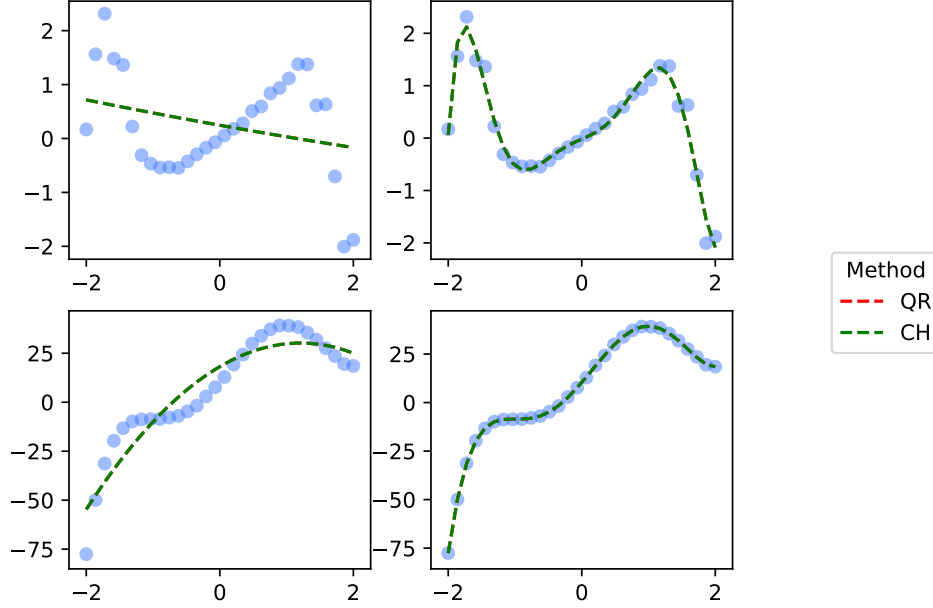


Figure 1: Plot of polynomial fitting of degree $m = 3$ (left column), and $m = 8$ (right column), to $f_1(x)$ (upper row), and $f_2(x)$ (bottom row) using both QR and Cholesky (CH)

3 Method comparison

3.1 Computational Cost

The two applied fitting methods are dissimilar in computational cost, measured in floating point operations (flops) - especially in the case where $m \ll n$. The normal equation solution uses $mn^2 + (1/3)n^3$ flops [1][p.146], while the QR decomposition requires $2mn^2 - (2/3)n^3$ flops [1][p.155]. Which means that the normal equations method is significantly faster than the QR approach.

3.2 Conditioning

The conditioning of the methods, or stability with respect to input perturbation, are measured in conditioning number of the matrix representing the problem, $K(A)$. Higher $K(A)$ means less stable.

$$K(A) = \|A^{-1}\| \times \|A\| \quad (6)$$

In order to acquire the normal equations, the product of A and its transpose is used. Hence, its conditioning number goes as $K(A^T A) = K(A)^2$. This squaring of the conditioning number is not required for the QR approach, which means

that the QR method has a lower conditioning number, and is thus the more stable method.

4 Conclusions

When handling well-conditioned Vandermonde matrices, the normal equations solution with Cholesky is faster than the QR method, and not necessarily particularly unstable. However, the larger the Vandermonde matrix is, the less well-conditioned it becomes. For higher order fittings, one may therefore expect the normal equations approach to be less well suited, as the solution will tend towards instability. On the other hand, the QR method is stable for a much larger variety of problems, and should therefore be employed when the polynomial degree (or number of predictors) is large.

References

- [1] Uri M. Ascher and Chen Greif. Linear least squares problems. In *A First Course on Numerical Methods*, pages 141–166. Society for Industrial and Applied Mathematics, 2011.
- [2] Michael S. Floater. Lecture 3: Cholesky and qr. Department of Mathematics, University of Oslo, September 19 2017.
- [3] Michael S. Floater. Lecture 4, gram-schmidt, qr solution to least squares. Department of Mathematics, University of Oslo, September 4 2017.