Interpolation and approximation with polynomials

Polynomial interpolation

Values of some real function $y(x)$ at a finite set of *distinct* points are prescribed and we want to interpolate them by a polynomial $p(x)$, so that we can estimate intermediate values of the function $y(x)$. Let us denote by $x_i,\ i = 0, 1, 2, \ldots n$ the values of independent variable x and by y_i the prescribed values of the function $y(x)$ at x_i and summarize all given values in a table:

x	x_0	x_1	x_2	$\ldots$	x_n
y	y_0	y_1	y_2	$\ldots$	y_n

Generally, if we have $n + 1$ data points $[x_i, y_i]$, there is exactly one polynomial

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

of degree at most n going through all the data points. Its coefficients $a_0, a_1, \ldots a_n$ are determined by $n + 1$ linear equations $p(x_i) = y_i,\ i = 0, 1, \ldots n$:

$$\begin{aligned} a_0 + a_1 x_0 + a_2 x_0^2 + \cdots + a_n x_0^n &= y_0 \\ a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_n x_1^n &= y_1 \\ &\ldots \\ a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_n x_n^n &= y_n \end{aligned}$$

This can be expressed in matrix form as $\mathbf{Q}\,\mathbf{a} = \mathbf{y}$:

$$\mathbf{Q}\,\mathbf{a} \equiv \begin{bmatrix} 1 & x_0 & x_0^2 & \ldots & x_0^n \\ 1 & x_1 & x_1^2 & \ldots & x_1^n \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_n & x_n^2 & \ldots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \ldots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \ldots \\ y_n \end{bmatrix} \equiv \mathbf{y} \tag{1}$$

The matrix is called Vandermonde matrix and for distinct values of x_i it is invertible, so the system (1) has unique solution for unknown vector of coefficients $\mathbf{a}$.

Note: The system (1) can be ill conditioned and solving it is not an efficient way how to compute the coefficients of the interpolation polynomial. Using *Lagrange interpolation polynomials* can be recommended instead.

Example 1

Consider the following table of data points

x	-1	1	2
y	8	4	5

Find the interpolation polynomial and use it for estimating the value at $x = 0.5$.

Solution

There are 3 data points, it follows that we are seeking the polynomial of the second order $p_2(x) = a_0 + a_1 x + a_2 x^2$ which has 3 unknown coefficients. Let us assemble the equations (1):

$$\begin{aligned} a_0 + a_1 \cdot (-1) + a_2 \cdot (-1)^2 &= 8 \\ a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 &= 4 \\ a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 &= 5 \end{aligned}$$

or, in a matrix form

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix} \quad .$$

The solution is $a_0 = 5$, $a_1 = -2$ and $a_2 = 1$, so the interpolating polynomial is
$p_2(x) = 5 - 2x + x^2$.

Value at 0.5 can be estimated as $p_2(0.5) = 5 - 2 \cdot 0.5 + 0.5^2 = 4.25$

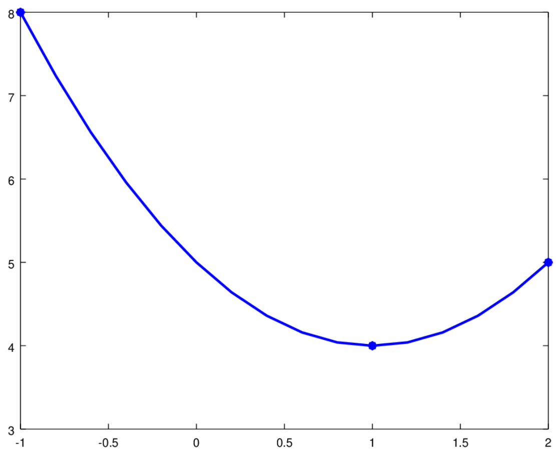

Figure 1: Solution of Example 1: given data (represented by dots) interpolated by the polynomial of second degree.

Polynomial approximation

As before, we are given $n+1$ discrete values, for example results of some measurements of dependency of value y on value of x. Now we want to *approximate* them by a polynomial $p(x)$, so that we can estimate the underlying function. We do not require the polynomial going exactly *through* the data points, we are satisfied if it is just *close* to them. However, we usually require the order m of the polynomial to be rather small, often m is less than 3.

Note: The requirement of x_i to be distinct values is not needed here.

Method of least squares

In this approximation method, the term *close* to the data points means that we want to minimize the sum of squares of vertical distances $r_i = p(x_i) - y_i$ of the given points from the graph of the polynomial, or, by other words, the euclidean norm $\|\mathbf{r}\|_2$ of the vector of residuals $\mathbf{r} = (r_1, r_2, \ldots, r_n)$.

Let us start again with the requirement that the polynomial goes through all the given data points: $p(x_i) = y_i,\ i = 0, 1, \ldots n$. The result will be a system of equations similar to (1):

$$\mathbf{Q}\,\mathbf{a} \equiv \begin{bmatrix} 1 & x_0 & x_0^2 & \ldots & x_0^m \\ 1 & x_1 & x_1^2 & \ldots & x_1^m \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_n & x_n^2 & \ldots & x_n^m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \ldots \\ a_m \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \ldots \\ \ldots \\ \ldots \\ y_n \end{bmatrix} \equiv \mathbf{y} \quad . \tag{2}$$

Now the system matrix $\mathbf{Q}$ is not square any more, however. It has m columns and $n+1$ rows, with m being much smaller than $n+1$. The system is over-determined and existence of solution generally cannot be expected, which means that the residual $\mathbf{r} = \mathbf{Q}\,\mathbf{a} - \mathbf{y}$ will be nonzero for any choice of $\mathbf{a}$. The goal is to get the norm $\|\mathbf{r}\|_2$ of the residual as small as possible. This can be achieved by requirement that the vector $\mathbf{r}$ is orthogonal to column space of the matrix $\mathbf{Q}$: $\mathbf{Q}^T \cdot \mathbf{r} = \mathbf{Q}^T \cdot (\mathbf{Q}\,\mathbf{a} - \mathbf{y}) = 0$. This leads to the system of *normal equations* $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a} = \mathbf{Q}^{\mathrm{T}}\mathbf{y}$:

$$\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a} \equiv \begin{bmatrix} n+1 & \sum_{i=0}^n x_i & \ldots & \sum_{i=0}^n x_i^m \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 & \ldots & \sum_{i=0}^n x_i^{m+1} \\ \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i^3 & \ldots & \sum_{i=0}^n x_i^{m+2} \\ \ldots & \ldots & \ldots & \ldots \\ \sum_{i=0}^n x_i^m & \sum_{i=0}^n x_i^{m+1} & \ldots & \sum_{i=0}^n x_i^{2m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \ldots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n y_i \\ \sum_{i=0}^n x_i y_i \\ \sum_{i=0}^n x_i^2 y_i \\ \ldots \\ \sum_{i=0}^n x_i^m y_i \end{bmatrix} \equiv \mathbf{Q}^{\mathrm{T}}\mathbf{y} \tag{3}$$

The matrix $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ is symmetric positive semidefinite, and if the columns of $\mathbf{Q}$ are linearly independent, then it is positive definite – and consequently invertible.

Note: $n+1$ at the upper left corner of $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ represents the number of the given data points.

Example 1 - continued

Suppose there are 3 more data points added to the table:

x	-1	1	2	-1	0	2
y	8	4	5	7	4	6

Find the approximation polynomial of second degree.

Solution

There are more equations than unknown variables and the system (2) generally has no solution:

$$\mathbf{Q}\,\mathbf{a} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \\ 5 \\ 7 \\ 4 \\ 6 \end{bmatrix} = \mathbf{y} \quad .$$

System of normal equations (3):

$$\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a} = \begin{bmatrix} 6 & 3 & 11 \\ 3 & 11 & 15 \\ 11 & 15 & 35 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 34 \\ 11 \\ 63 \end{bmatrix} = \mathbf{Q}^{\mathrm{T}}\mathbf{y} \quad .$$

The solution is $a_0 = 4.3158$, $a_1 = -1.8816$ and $a_2 = 1.25$, so the approximating polynomial is $p_2(x) = 4.3158 - 1.8816x + 1.25x^2$. The result is depicted in Figure 2 by magenta line. For comparison, the blue line is the polynomial interpolating the original three blue dots.

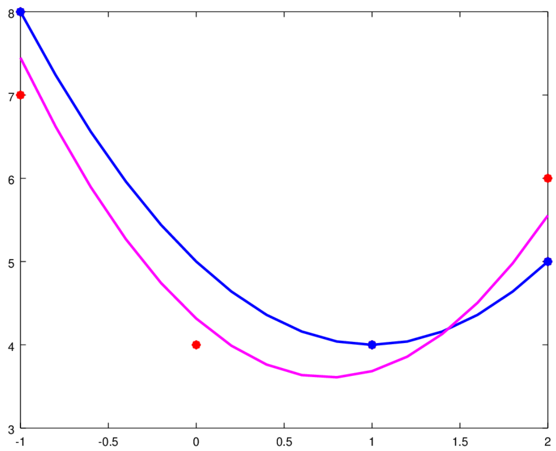

Figure 2: Solution of Example 1: original data (represented by blue dots) and added data (represented by red dots), all together approximated by the polynomial of second degree (magenta line). The original interpolating polynomial is blue.

Theorem 1: approximation by a polynomial using the least squares method
Suppose data points $x_i, y_i,\ i = 0, \dots n$ are given so that the matrix $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ of the system (3) is invertible. Then the solution of (3) represents coefficients of the polynomial $p_m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$ which minimizes the norm $\|\mathbf{r}\|_2$ of the vector of residuals $\mathbf{r} = (r_1, r_2, \dots, r_n)$, where $r_i = p_m(x_i) - y_i$, among all polynomials of degree at most m. The norm $\|\mathbf{r}\|_2$ is called *quadratic deviation* δ.

For better insight, two different ways of proving this theorem follow.

Proof 1 (using Calculus)
We want to minimize the quadratic deviation δ, or δ^2:

$$\delta^2 \equiv \|\mathbf{r}\|_2^2 = \sum_{i=0}^{n} r_i^2 = \sum_{i=0}^{n} (p_m(x_i) - y_i)^2 \quad \rightarrow \quad min\ .$$

The minimum of this sum, with respect to the unknown coefficients $a_0, a_1, \dots a_m$ of the polynomial $p_m(x)$, is found by setting the gradient of the function $\delta^2 \equiv S(a_0, a_1, \dots a_m)$ to zero. Since it has $m+1$ variables, there are $m+1$ gradient equations

$$\begin{aligned}
0 = \frac{\partial S}{\partial a_0} &= \frac{\partial}{\partial a_0} \sum_{i=0}^{n} r_i^2 = \sum_{i=0}^{n} 2r_i \frac{\partial r_i}{\partial a_0} = \sum_{i=0}^{n} 2r_i \\
0 = \frac{\partial S}{\partial a_1} &= \frac{\partial}{\partial a_1} \sum_{i=0}^{n} r_i^2 = \sum_{i=0}^{n} 2r_i \frac{\partial r_i}{\partial a_1} = \sum_{i=0}^{n} 2r_i x_i \\
0 = \frac{\partial S}{\partial a_2} &= \frac{\partial}{\partial a_2} \sum_{i=0}^{n} r_i^2 = \sum_{i=0}^{n} 2r_i \frac{\partial r_i}{\partial a_2} = \sum_{i=0}^{n} 2r_i x_i^2 \\
&\cdots \\
0 = \frac{\partial S}{\partial a_m} &= \frac{\partial}{\partial a_m} \sum_{i=0}^{n} r_i^2 = \sum_{i=0}^{n} 2r_i \frac{\partial r_i}{\partial a_m} = \sum_{i=0}^{n} 2r_i x_i^m
\end{aligned}$$

System (3) is obtained by substitution $r_i = p_m(x_i) - y_i = a_0 + a_1 x_i + \cdots + a_m x_i^m - y_i$ and rearangement. $\square$

Proof 2 is based on linear algebra and on following Theorem 2 (not proved here):
Theorem 2:
Suppose $\mathbf{A}$ is a symmetric and positive definite (*spd*) matrix, $\mathbf{b}$ is a vector and $\mathbf{F}(\mathbf{x})$ is a quadratic functional $\mathbf{F}(\mathbf{x}) = \mathbf{x}^T\,\mathbf{A}\,\mathbf{x} - 2\,\mathbf{x}^T\,\mathbf{b}$. Then $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b} \ \Leftrightarrow\ \mathbf{F}(\bar{\mathbf{x}}) < \mathbf{F}(\mathbf{x}) \quad \forall\ \mathbf{x} \neq \bar{\mathbf{x}}$.

Note: In the Theorem 2, often a functional $\mathbf{J}(x) = \frac{1}{2}\,\mathbf{F}(x)$ instead of $\mathbf{F}(x)$ in is used.

Proof of the Theorem 1 using Theorem 2:
Let $\mathbf{A} = \mathbf{Q}^{\mathrm{T}}\mathbf{Q}$, $\mathbf{b} = \mathbf{Q}^{\mathrm{T}}\mathbf{y}$.
Let us express $\|\mathbf{r}\|_2^2$ as a quadratic functional corresponding to $\mathbf{F}$ in Theorem 2:

$$\begin{aligned}
\|\mathbf{r}\|_2^2 &= \|\mathbf{Q}\,\mathbf{a} - \mathbf{y}\|_2^2 = (\mathbf{Q}\,\mathbf{a} - \mathbf{y})^{\mathrm{T}}(\mathbf{Q}\,\mathbf{a} - \mathbf{y}) = \mathbf{a}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a} - \mathbf{y}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a} - \mathbf{a}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{y} \\
&= \mathbf{a}^{\mathrm{T}}\mathbf{A}\,\mathbf{a} - \mathbf{b}^{\mathrm{T}}\mathbf{a} - \mathbf{a}^{\mathrm{T}}\mathbf{b} + \mathbf{y}^{\mathrm{T}}\mathbf{y} = \mathbf{a}^{\mathrm{T}}\mathbf{A}\,\mathbf{a} - 2\,\mathbf{a}^{\mathrm{T}}\mathbf{b} + \|\mathbf{y}\|_2^2 = \mathbf{F}(\mathbf{a}) + \|\mathbf{y}\|_2^2
\end{aligned}$$

$\|\mathbf{y}\|_2^2$ does not depend on $\mathbf{a}$ and so minimization of $\|\mathbf{r}\|_2^2 = \|\mathbf{Q}\,\mathbf{a} - \mathbf{y}\|_2^2$ with respect to $\mathbf{a}$ is the same as minimization of $\mathbf{F}(\mathbf{a}) = \mathbf{a}^{\mathrm{T}}\mathbf{A}\,\mathbf{a} - 2\,\mathbf{a}^{\mathrm{T}}\mathbf{b}$ with respect to $\mathbf{a}$. From Theorem 2 it follows that this can happen if and only if $\mathbf{a}$ is the solution of the system (3), i.e. $\mathbf{A}\,\mathbf{a} = \mathbf{b}$. $\square$

Illustration of Theorem 2 in R^2 (recapitulation from Calculus):

$$\mathbf{F}(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\,\mathbf{x} - 2\,\mathbf{x}^T\,\mathbf{b} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & c \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2 \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} =$$

$$= a\,x_1^2 + 2c\,x_1x_2 + d\,x_2^2 - 2x_1b_1 - 2x_2b_2$$

$$\text{grad } \mathbf{F}(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial \mathbf{F}}{\partial x_1} \\[2ex] \dfrac{\partial \mathbf{F}}{\partial x_2} \end{bmatrix} = 2 \begin{bmatrix} a\,x_1 + c\,x_2 - b_1 \\[2ex] c\,x_1 + d\,x_2 - b_2 \end{bmatrix} = 2\,(\mathbf{Ax} - \mathbf{b})$$

it follows that $\mathbf{F}$ can have one critical point only – the solution of $\mathbf{Ax} = \mathbf{b}$ (the solution is unique, because the matrix $\mathbf{A}$ is spd and so it is invertible)

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 \mathbf{F}}{\partial x_1^2} & \dfrac{\partial^2 \mathbf{F}}{\partial x_1 x_2} \\[2ex] \dfrac{\partial^2 \mathbf{F}}{\partial x_2 x_1} & \dfrac{\partial^2 \mathbf{F}}{\partial x_2^2} \end{bmatrix} = 2\,\mathbf{A}$$

$\mathbf{A}$ is spd, so the critical point represents minimum and the graph of $\mathbf{F}$ is an elliptic paraboloid oriented upwards.

Example 2

Consider the following table of seven data points

x	-1	0	0	1	1	2	4
y	5	6	5	7	6	8	11

Find the approximation polynomials of the first and the second order and the standard deviations.

Solution

The **first-order** polynomial is a line $p_1(x) = a_0 + a_1 x$, coefficients of which are determined by the linear system (3) with $m = 1$, $n = 6$:

$$\mathbf{A} = \begin{bmatrix} n+1 & \sum_{i=0}^n x_i \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \sum_{i=0}^n y_i \\ \sum_{i=0}^n x_i y_i \end{bmatrix}$$

Let us extend the given table of data in order to compute the elements of the matrix and the right hand side:

								sum
x	-1	0	0	1	1	2	4	7
y	5	6	5	7	6	8	11	48
xy	-5	0	0	7	6	16	44	68
$\mathbf{x^2}$	1	0	0	1	1	4	16	23

which gives two linear equations for unknowns a_0 and a_1:

$$\begin{bmatrix} 7 & 7 \\ 7 & 23 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 48 \\ 68 \end{bmatrix}.$$

The solution is $a_0 = 5.6071$, $a_1 = 1.25$ and the linear approximation of the given data is $p_1(x) = 5.6071 + 1.25\,x$.

$\delta^2 = \sum_{i=0}^{7} (p_1(x_i) - y_i)^2 = (4.3571 - 5)^2 + (5.6071 - 6)^2 + (5.6071 - 5)^2 + (6.8571 - 7)^2 + (6.8571 - 6)^2 + (8.1071 - 8)^2 + (10.6071 - 11)^2 = (-0.6429)^2 + (-0.3929)^2 + 0.6071^2 + (-0.1429)^2 + 0.8571^2 + 0.1071^2 + (-0.3929)^2 = 0.4133 + 0.1543 + 0.3685 + 0.0204 + 0.7346 + 0.0115 + 0.1544 = 1.8571$

$\delta = \sqrt{\delta^2} = 1.3628$

The **second-order** polynomial is a parabola $p_2(x) = a_0 + a_1x + a_2x^2$, coefficients of which are determined by the linear system (3) with $m = 2$, $n = 6$:

$$\mathbf{A} = \begin{bmatrix} n+1 & \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 \\ \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \\ \sum_{i=0}^{n} x_i^2 y_i \end{bmatrix}$$

Let us extend the table by three more rows: which gives three linear equations for unknowns

								sum
x	-1	0	0	1	1	2	4	7
y	5	6	5	7	6	8	11	48
xy	-5	0	0	7	6	16	44	68
$\mathbf{x^2}$	1	0	0	1	1	4	16	23
$\mathbf{x^2y}$	5	0	0	7	6	32	176	226
$\mathbf{x^3}$	-1	0	0	1	1	8	64	73
$\mathbf{x^4}$	1	0	0	1	1	16	256	275

a_0, a_1 and a_2:

$$\begin{bmatrix} 7 & 7 & 23 \\ 7 & 23 & 73 \\ 23 & 73 & 275 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 48 \\ 68 \\ 226 \end{bmatrix}.$$

The solution is $a_0 = 5.5856$, $a_1 = 0.8313$, $a_2 = 0.1340$ and the quadratic approximation of the given data is $p_1(x) = 5.5856 + 0.8313\,x + 0.1340\,x^2$.

$\delta^2 = (4.8883 - 5)^2 + (5.5856 - 6)^2 + (5.5856 - 5)^2 + (6.5509 - 7)^2 + (6.5509 - 6)^2 + (7.7842 - 8)^2 + +(11.0548 - 11)^2 = 1.0819$

$\delta = \sqrt{\delta^2} = 1.0401$

Another way how to obtain the linear system above:

$$\mathbf{Q} = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ \dots & \dots & \dots \\ 1 & x_n & x_n^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix}, \ \mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \begin{bmatrix} 7 & 7 & 23 \\ 7 & 23 & 73 \\ 23 & 73 & 275 \end{bmatrix}, \ \mathbf{Q}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} 48 \\ 68 \\ 226 \end{bmatrix}.$$