Discrete Least Squares Approximations

As stated previously, one of the most fundamental problems in science and engineering is data fitting—constructing a function that, in some sense, conforms to given data points. So far, we have discussed two data-fitting techniques, polynomial interpolation and piecewise polynomial interpolation. Interpolation techniques, of any kind, construct functions that agree exactly with the data. That is, given points \((x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\), interpolation yields a function \(f(x)\) such that \(f(x_i) = y_i\) for \(i = 1, 2, \ldots, m\).

However, fitting the data exactly may not be the best approach to describing the data with a function. We have seen that high-degree polynomial interpolation can yield oscillatory functions that behave very differently than a smooth function from which the data is obtained. Also, it may be pointless to try to fit data exactly, for if it is obtained by previous measurements or other computations, it may be erroneous. Therefore, we consider revising our notion of what constitutes a “best fit” of given data by a function.

One alternative approach to data fitting is to solve the minimax problem, which is the problem of finding a function \(f(x)\) of a given form for which

\[
\max_{1 \leq i \leq n} |f(x_i) - y_i|
\]

is minimized. However, this is a very difficult problem to solve.

Another approach is to minimize the total absolute deviation of \(f(x)\) from the data. That is, we seek a function \(f(x)\) of a given form for which

\[
\sum_{i=1}^{m} |f(x_i) - y_i|
\]

is minimized. However, we cannot apply standard minimization techniques to this function, because, like the absolute value function that it employs, it is not differentiable.

This defect is overcome by considering the problem of finding \(f(x)\) of a given form for which

\[
\sum_{i=1}^{m} [f(x_i) - y_i]^2
\]
is minimized. This is known as the least squares problem. We will first show how this problem is solved for the case where \( f(x) \) is a linear function of the form \( f(x) = a_1 x + a_0 \), and then generalize this solution to other types of functions.

When \( f(x) \) is linear, the least squares problem is the problem of finding constants \( a_0 \) and \( a_1 \) such that the function

\[
E(a_0, a_1) = \sum_{i=1}^{m} (a_1 x_i + a_0 - y_i)^2
\]

is minimized. In order to minimize this function of \( a_0 \) and \( a_1 \), we must compute its partial derivatives with respect to \( a_0 \) and \( a_1 \). This yields

\[
\frac{\partial E}{\partial a_0} = \sum_{i=1}^{m} 2(a_1 x_i + a_0 - y_i), \quad \frac{\partial E}{\partial a_1} = \sum_{i=1}^{m} 2(a_1 x_i + a_0 - y_i)x_i.
\]

At a minimum, both of these partial derivatives must be equal to zero. This yields the system of linear equations

\[
na_0 + \left( \sum_{i=1}^{m} x_i \right) a_1 = \sum_{i=1}^{m} y_i, \\
\left( \sum_{i=1}^{m} x_i \right) a_0 + \left( \sum_{i=1}^{m} x_i^2 \right) a_1 = \sum_{i=1}^{m} x_i y_i.
\]

Using the formula for the inverse of a 2 \( \times \) 2 matrix,

\[
\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},
\]

we obtain the solutions

\[
a_0 = \frac{(\sum_{i=1}^{m} x_i^2)(\sum_{i=1}^{m} y_i) - (\sum_{i=1}^{m} x_i)(\sum_{i=1}^{m} x_i y_i)}{n \sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2},
\]

\[
a_1 = \frac{n \sum_{i=1}^{m} x_i y_i - (\sum_{i=1}^{m} x_i)(\sum_{i=1}^{m} y_i)}{n \sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2}.
\]

**Example** We wish to find the linear function \( y = a_1 x + a_0 \) that best approximates the data shown in Table 1, in the least-squares sense. Using the summations

\[
\sum_{i=1}^{m} x_i = 56.2933, \quad \sum_{i=1}^{m} x_i^2 = 380.5426, \quad \sum_{i=1}^{m} y_i = 73.8373, \quad \sum_{i=1}^{m} x_i y_i = 485.9487,
\]

2
Table 1: Data points \((x_i, y_i)\), for \(i = 1, 2, \ldots, 10\), to be fit by a linear function

<table>
<thead>
<tr>
<th>(i)</th>
<th>(x_i)</th>
<th>(y_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0774</td>
<td>3.3123</td>
</tr>
<tr>
<td>2</td>
<td>2.3049</td>
<td>3.8982</td>
</tr>
<tr>
<td>3</td>
<td>3.0125</td>
<td>4.6500</td>
</tr>
<tr>
<td>4</td>
<td>4.7092</td>
<td>6.5576</td>
</tr>
<tr>
<td>5</td>
<td>5.5016</td>
<td>7.5173</td>
</tr>
<tr>
<td>6</td>
<td>5.8704</td>
<td>7.0415</td>
</tr>
<tr>
<td>7</td>
<td>6.2248</td>
<td>7.7497</td>
</tr>
<tr>
<td>8</td>
<td>8.4431</td>
<td>11.0451</td>
</tr>
<tr>
<td>9</td>
<td>8.7594</td>
<td>9.8179</td>
</tr>
<tr>
<td>10</td>
<td>9.3900</td>
<td>12.2477</td>
</tr>
</tbody>
</table>

we obtain

\[
a_0 = \frac{380.5426 \cdot 73.8373 - 56.2933 \cdot 485.9487}{10 \cdot 380.5426 - 56.2933^2} = \frac{742.5703}{636.4906} = 1.1667,
\]

\[
a_1 = \frac{10 \cdot 485.9487 - 56.2933 \cdot 73.8373}{10 \cdot 380.5426 - 56.2933^2} = \frac{702.9438}{636.4906} = 1.1044.
\]

We conclude that the linear function that best fits this data in the least-squares sense is

\[y = 1.1044x + 1.1667.\]

The data, and this function, are shown in Figure 1. □

It is interesting to note that if we define the \(m \times 2\) matrix \(A\), the 2-vector \(a\), and the \(m\)-vector \(y\) by

\[A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad a = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},\]

then \(a\) is the solution to the system of equations

\[A^T A a = A^T y.\]

These equations are called the normal equations. They arise from the problem of finding the vector \(a\) such that

\[|Aa - y|.\]
is minimized, where, for any vector $\mathbf{u}$, $|\mathbf{u}|$ is the magnitude, or length, of $\mathbf{u}$.

In this case, this expression is equivalent to the square root of the expression we originally intended to minimize,

$$\sum_{i=1}^{m} (a_1x_i + a_0 - y_i)^2,$$

but the normal equations also characterize the solution $\mathbf{a}$, an $n$-vector, to the more general linear least squares problem of minimizing $|A\mathbf{a} - \mathbf{y}|$ for any matrix $A$ that is $m \times n$, where $m \geq n$, whose columns are linearly independent.