Optimization

master in fundamental principles of data science

Optimization Problem 4

Quadratic method minimization problem

Author
Vladislav Nikolov Vasilev

Faculty of Mathematics and Computer Science

Academic year 2021-2022

1 Problem description

Let f be a real function on $\mathbb{R}^n$. Also let $x_0 \in \mathbb{R}^n$, $z \in \mathbb{R}^n$, and $\theta \in \mathbb{R}$. Define

$$F(\theta) = f(x_0 + \theta z)$$

and suppose that we are looking for the minimum of F (that is, for the minimum of f in the direction z through the point x_0). Let $x_0 + \theta_1 z$, $x_0 + \theta_2 z$ and $x_0 + \theta_3 z$ be three points where f is evaluated. Show that the minimum predicted by applying the quadratic approximation method is $x_0 + \theta^* z$, where

$$\theta^* = \frac{[\theta_2^2 - \theta_3^2]F(\theta_1) + [\theta_3^2 - \theta_1^2]F(\theta_2) + [\theta_1^2 - \theta_2^2]F(\theta_3)}{2[(\theta_2 - \theta_3)F(\theta_1) + (\theta_3 - \theta_1)F(\theta_2) + (\theta_1 - \theta_2)F(\theta_3)]}$$

and it is indeed the minimum of the parabola passing through the above three points if

$$\frac{(\theta_2 - \theta_3)F(\theta_1) + (\theta_3 - \theta_1)F(\theta_2) + (\theta_1 - \theta_2)F(\theta_3)}{(\theta_2 - \theta_3)(\theta_3 - \theta_1)(\theta_1 - \theta_2)} < 0$$

2 Solution

The quadratic method allows us to interpolate a given function by using a second-degree polynomial. In this case, we are going to approximate F by

$$\phi(\theta) = a + b\theta + c\theta^2$$

To determine the values of the coefficients of ϕ we can use three points, θ_1, θ_2 and θ_3. These points will be evaluated using the function F. After that, we can easily solve a system of linear equations to find the coefficients.

Notice that, since ϕ approximates F, a local minimum of the function F is also going to be a local minimum of ϕ. To find a first approximation of a local minimum, we can solve for $\phi(\theta) = 0$, which gives us:

$$\theta^* = -\frac{b}{2c} \tag{1}$$

However, for the equation seen in (1) to yield a minimum, it is necessary that $c > 0$, which means that the parabola described by ϕ is convex.

First, let's find out the values of the coefficients. The system of linear equations that we must solve is the following:

$$\begin{pmatrix} \phi(\theta_1) \\ \phi(\theta_2) \\ \phi(\theta_3) \end{pmatrix} = \begin{pmatrix} 1 & \theta_1 & \theta_1^2 \\ 1 & \theta_2 & \theta_2^2 \\ 1 & \theta_3 & \theta_3^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} F(\theta_1) \\ F(\theta_2) \\ F(\theta_3) \end{pmatrix} \tag{2}$$

where the matrix

$$V = \begin{pmatrix} 1 & \theta_1 & \theta_1^2 \\ 1 & \theta_2 & \theta_2^2 \\ 1 & \theta_3 & \theta_3^2 \end{pmatrix}$$

is a Vandermonde matrix, whose determinant can be computed as

$$|V| = \prod_{1 \leq i < j \leq n} (a_j - a_i)$$

which in this case is $|V| = (\theta_2 - \theta_3)(\theta_3 - \theta_1)(\theta_1 - \theta_2)$.

With this information in hand, we can attempt to solve the system seen in (2). Since in equation (1) only appear b and c, we can compute them using using Cramer's rule.

Let's start with c:

$$\begin{aligned} c = \frac{\begin{vmatrix} 1 & \theta_1 & F(\theta_1) \\ 1 & \theta_2 & F(\theta_2) \\ 1 & \theta_3 & F(\theta_3) \end{vmatrix}}{|V|} &= \\ = \frac{\theta_2 F(\theta_3) - \theta_3 F(\theta_2) - \theta_1 F(\theta_3) + \theta_3 F(\theta_1) + \theta_1 F(\theta_2) - \theta_2 F(\theta_1)}{(\theta_2 - \theta_3)(\theta_3 - \theta_1)(\theta_1 - \theta_2)} &= \\ = \frac{(\theta_3 - \theta_2)F(\theta_1) + (\theta_1 - \theta_3)F(\theta_2) + (\theta_2 - \theta_1)F(\theta_3)}{(\theta_2 - \theta_3)(\theta_3 - \theta_1)(\theta_1 - \theta_2)} &> 0 \end{aligned} \tag{3}$$

Following the result of (3), it's easy to note that if $c > 0$, then $-c < 0$, which leads us to:

$$c = \frac{(\theta_2 - \theta_3)F(\theta_1) + (\theta_3 - \theta_1)F(\theta_2) + (\theta_1 - \theta_2)F(\theta_3)}{(\theta_2 - \theta_3)(\theta_3 - \theta_1)(\theta_1 - \theta_2)} < 0$$

Hence, we can see that the parabola is, indeed, convex. Thus, it contains a minimum.

Next, if we compute the value of b, we get the following:

$$\begin{aligned}
b &= \frac{\begin{vmatrix} 1 & F(\theta_1) & \theta_1^2 \\ 1 & F(\theta_2) & \theta_2^2 \\ 1 & F(\theta_3) & \theta_3^2 \end{vmatrix}}{|V|} = \\
&= \frac{\theta_3^2 F(\theta_2) - \theta_2^2 F(\theta_3) - \theta_3^2 F(\theta_1) + \theta_1^2 F(\theta_3) + \theta_2^2 F(\theta_1) - \theta_1^2 F(\theta_2)}{(\theta_2 - \theta_3)(\theta_3 - \theta_1)(\theta_1 - \theta_2)} = \\
&= \frac{(\theta_2^2 - \theta_3^2)F(\theta_1) + (\theta_3^2 - \theta_1^2)F(\theta_2) + (\theta_1^2 - \theta_2^2)F(\theta_3)}{(\theta_2 - \theta_3)(\theta_3 - \theta_1)(\theta_1 - \theta_2)}
\end{aligned}$$

Now, if we try to solve equation (1) using the values of b and c that we have just found out in order to get the minimum θ^*, we obtain the following result:

$$\begin{aligned}
\theta^* &= -\frac{[\theta_2^2 - \theta_3^2]F(\theta_1) + [\theta_3^2 - \theta_1^2]F(\theta_2) + [\theta_1^2 - \theta_2^2]F(\theta_3)}{2[(\theta_3 - \theta_2)F(\theta_1) + (\theta_1 - \theta_3)F(\theta_2) + (\theta_2 - \theta_1)F(\theta_3)]} \\
&= \frac{[\theta_2^2 - \theta_3^2]F(\theta_1) + [\theta_3^2 - \theta_1^2]F(\theta_2) + [\theta_1^2 - \theta_2^2]F(\theta_3)}{2[(\theta_2 - \theta_3)F(\theta_1) + (\theta_3 - \theta_1)F(\theta_2) + (\theta_1 - \theta_2)F(\theta_3)]}
\end{aligned}$$

Thus, we can conclude that the function ϕ that approximates F is convex, and hence, it has a minimum, which is θ^*. Because ϕ interpolates the function F, θ^* is also going to be a minimum of F, which guarantees us that if we minimize ϕ, we are also going to minimize F.