Exercise 4

Lorenzo Vigo

Optimization: 10^{th} October 2021

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

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

The quadratic approximation method finds a minimum of a given function by computing an estimation of it in the following form: $\phi(x) = a + bx + cx^2$. Then, under some conditions that must be satisfied, the minimum will be predicted to be at $\hat{x} = -\frac{b}{2c}$.

We will try to find the minimum of f by computing the estimation ϕ of our function f. In order to do that, we need to discover the values of b and c. Defining $t_i = x_0 + \sigma_i z$ for $i = 1, 2, 3$ and evaluating f on the three given points, we know that:

$$\begin{aligned}
f(x_0 + \sigma_1 z) = f(t_1) = a + bt_1 + ct_1^2 = F(\sigma_1) \\
f(x_0 + \sigma_2 z) = f(t_2) = a + bt_2 + ct_2^2 = F(\sigma_2) \\
f(x_0 + \sigma_3 z) = f(t_3) = a + bt_3 + ct_3^2 = F(\sigma_3)
\end{aligned}$$

This is a set of 3 equations with 3 unset values (a, b, c). We will use the Cramer Solving Equation Rule to learn the values of b and c. First of all, let's define:

$$|D| \coloneqq \begin{vmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{vmatrix} \tag{1}$$

Now, following the Cramer's Rule:

$$b = \frac{\begin{vmatrix} 1 & F(\sigma_1) & t_1^2 \\ 1 & F(\sigma_2) & t_2^2 \\ 1 & F(\sigma_3) & t_3^2 \end{vmatrix}}{|D|} = \frac{F(\sigma_1)(t_2^2 - t_3^2) + F(\sigma_2)(t_3^2 - t_1^2) + F(\sigma_3)(t_1^2 - t_2^2)}{|D|}$$

$$c = \frac{\begin{vmatrix} 1 & t_1 & F(\sigma_1) \\ 1 & t_2 & F(\sigma_2) \\ 1 & t_3 & F(\sigma_3) \end{vmatrix}}{|D|} = \frac{F(\sigma_1)(t_3 - t_2) + F(\sigma_2)(t_1 - t_3) + F(\sigma_3)(t_2 - t_1)}{|D|} \tag{2}$$

Therefore, according to the used method, the minimum $\hat{x}$ has the following value:

$$\hat{x} = -\frac{b}{2c} = -\frac{F(\sigma_1)(t_2^2 - t_3^2) + F(\sigma_2)(t_3^2 - t_1^2) + F(\sigma_3)(t_1^2 - t_2^2)}{2(F(\sigma_1)(t_3 - t_2) + F(\sigma_2)(t_1 - t_3) + F(\sigma_3)(t_2 - t_1))} \tag{3}$$

These following preliminary calculations will be useful in order to simplify the expression of $\hat{x}$:

$$\begin{aligned} t_3 - t_2 &= x_0 + \sigma_3 z - x_0 - \sigma_2 z && = (\sigma_3 - \sigma_2)z \\ t_1 - t_3 &= && = (\sigma_1 - \sigma_3)z \\ t_2 - t_1 &= && = (\sigma_2 - \sigma_1)z \end{aligned}$$

$$\begin{aligned} t_2^2 - t_3^2 &= x_0^2 + \sigma_2^2 z^2 + 2x_0\sigma_2 - x_0^2 - \sigma_3^2 z^2 - 2x_0\sigma_3 z && = (\sigma_2^2 - \sigma_3^2)z^2 + 2x_0(\sigma_2 - \sigma_3)z \\ t_3^2 - t_1^2 &= && = (\sigma_3^2 - \sigma_1^2)z^2 + 2x_0(\sigma_3 - \sigma_1)z \\ t_1^2 - t_2^2 &= && = (\sigma_1^2 - \sigma_2^2)z^2 + 2x_0(\sigma_1 - \sigma_2)z \end{aligned}$$

Following from (3):

$$\hat{x} = -\frac{z((\sigma_2^2 - \sigma_3^2)F(\sigma_1)z + 2x_0(\sigma_2 - \sigma_3)F(\sigma_1) + (\sigma_3^2 - \sigma_1^2)F(\sigma_2)z + 2x_0(\sigma_3 - \sigma_1)F(\sigma_2) + (\sigma_1^2 - \sigma_2^2)F(\sigma_3)z + 2x_0(\sigma_1 - \sigma_2)F(\sigma_2))}{2z((\sigma_3 - \sigma_2)F(\sigma_1) + (\sigma_1 - \sigma_3)F(\sigma_2) + (\sigma_2 - \sigma_1)F(\sigma_3))}$$

$$\hat{x} = \frac{2x_0(\sigma_2 - \sigma_3)F(\sigma_1) + 2x_0(\sigma_3 - \sigma_1)F(\sigma_2) + 2x_0(\sigma_1 - \sigma_2)F(\sigma_3)}{2((\sigma_2 - \sigma_3)F(\sigma_1) + (\sigma_3 - \sigma_1)F(\sigma_2) + (\sigma_1 - \sigma_2)F(\sigma_3))} + \frac{(\sigma_2^2 - \sigma_3^2)F(\sigma_1)z + (\sigma_3^2 - \sigma_1^2)F(\sigma_2)z + (\sigma_1^2 - \sigma_2^2)F(\sigma_3)z}{2((\sigma_2 - \sigma_3)F(\sigma_1) + (\sigma_3 - \sigma_1)F(\sigma_2) + (\sigma_1 - \sigma_2)F(\sigma_3))}$$

$$\hat{x} = \frac{x_0((\sigma_2 - \sigma_3)F(\sigma_1) + (\sigma_3 - \sigma_1)F(\sigma_2) + (\sigma_1 - \sigma_2)F(\sigma_3))}{(\sigma_2 - \sigma_3)F(\sigma_1) + (\sigma_3 - \sigma_1)F(\sigma_2) + (\sigma_1 - \sigma_2)F(\sigma_3)} + \frac{(\sigma_2^2 - \sigma_3^2)F(\sigma_1) + (\sigma_3^2 - \sigma_1^2)F(\sigma_2) + (\sigma_1^2 - \sigma_2^2)F(\sigma_3)}{2((\sigma_2 - \sigma_3)F(\sigma_1) + (\sigma_3 - \sigma_1)F(\sigma_2) + (\sigma_1 - \sigma_2)F(\sigma_3))} z$$

$$\hat{x} = x_0 + \sigma^* z$$

We just proved that the candidate to be the minimum given by exercise is actually the minimum according to the quadratic approximation method. However, this is only true if the simple condition $c > 0$ is satisfied. We will check that this condition is equivalent to the one specified in the exercise.

In order to do that, we will go further with expression (1). We will skip some of the steps:

$$\begin{aligned}
|D| &= t_1t_2^2 + t_2t_3^2 + t_3t_1^2 - t_1t_3^2 - t_2t_1^2 - t_3t_2^2 \\
|D| &= t_1^2(t_3 - t_2) + t_2^2(t_1 - t_3) + t_3^2(t_2 - t_1) \\
|D| &= z^3(\sigma_1^2\sigma_3 - \sigma_1^2\sigma_2 + \sigma_2^2\sigma_1 - \sigma_2^2\sigma_3 + \sigma_3^2\sigma_2 - \sigma_3^2\sigma_1 \\
|D| &= z^3(\sigma_2 - \sigma_3)(\sigma_3 - \sigma_1)(\sigma_1 - \sigma_2)
\end{aligned}$$

Following from the expression of c in (2), $c > 0$ is equivalent to:

$$\begin{aligned}
\frac{(\sigma_3 - \sigma_2)F(\sigma_1)z + (\sigma_1 - \sigma_3)F(\sigma_2)z + (\sigma_2 - \sigma_1)F(\sigma_3)z}{|D|} &> 0 \\
\frac{z((\sigma_3 - \sigma_2)F(\sigma_1) + (\sigma_1 - \sigma_3)F(\sigma_2) + (\sigma_2 - \sigma_1)F(\sigma_3))}{zz^2(\sigma_2 - \sigma_3)(\sigma_3 - \sigma_1)(\sigma_1 - \sigma_2)} &> 0 \\
\frac{(\sigma_3 - \sigma_2)F(\sigma_1) + (\sigma_1 - \sigma_3)F(\sigma_2) + (\sigma_2 - \sigma_1)F(\sigma_3)}{z^2(\sigma_2 - \sigma_3)(\sigma_3 - \sigma_1)(\sigma_1 - \sigma_2)} &> 0
\end{aligned}$$

As z^2 is always positive, it is redundant in the inequality. By multiplying the by -1 on both sides, the condition given in the exercise is obtained. Therefore, it is equivalent to $c > 0$, as we wanted to show:

$$\frac{(\sigma_2 - \sigma_3)F(\sigma_1) + (\sigma_3 - \sigma_1)F(\sigma_2) + (\sigma_1 - \sigma_2)F(\sigma_3)}{(\sigma_2 - \sigma_3)(\sigma_3 - \sigma_1)(\sigma_1 - \sigma_2)} < 0$$