Optimal Step Gradient: Determining the Step Size Using Newton's Method

Consider the function f defined for all $v = (v_1, v_2) \in \mathbb{R}^2$ by

$$f(v) = (v_1 - 4)^2 + 2(v_2 - 3)^2 + v_1 v_2$$

Recall that in the optimal step gradient method, the step size ρ_n changes at each iteration:

$$v_{n+1} = v_n - \rho_n \nabla f(v_n) \quad \text{where} \quad \rho_n = \arg\min_{\rho > 0} f(v_n - \rho \nabla f(v_n))$$

Let $g_n(\rho) = f(v_n - \rho \nabla f(v_n))$, defined from $\mathbb{R}$ to $\mathbb{R}$. At each iteration, to determine the optimal step size ρ_n, we seek the solution of the following (possibly nonlinear) problem:

$$g_n'(\rho) = 0$$

We show that, for all $n \geq 0$, the derivative of g_n is

$$\forall \rho \in \mathbb{R}, \quad g_n'(\rho) = -\nabla f(v_n) \cdot \nabla f(v_n - \rho \nabla f(v_n))$$

To determine an approximate value of the solution $g_n'(\rho) = 0$, we will consider a variant of Newton's method: in Newton's formula given by

$$\rho_{k+1} = \rho_k - \frac{g_n'(\rho_k)}{g_n''(\rho_k)},$$

to avoid calculating g_n'', we replace $g_n''(\rho_k)$ with the rate of change:

$$\frac{g_n'(\rho_k + \delta) - g_n'(\rho_k)}{\delta},$$

with $\delta = 10^{-8}$.

Thus, at each step n of the gradient algorithm, we construct the sequence

$$\begin{cases} \rho_0 \in \mathbb{R} \text{ given} \\ \rho_{k+1} = \rho_k - \delta \frac{g_n'(\rho_k)}{g_n'(\rho_k + \delta) - g_n'(\rho_k)} \end{cases}$$

which converges to ρ_n, the solution of the problem $\rho_n = \arg\min_{\rho > 0} f(v_n - \rho \nabla f(v_n))$.