Gradient methods

Recall that a gradient method applied to a function f uses an initial point $\mathbf{x}_0$ and the update rule

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla h(\mathbf{x}_k).$$

The quantity α_k is the step size. A fixed-step-size algorithm uses a constant $\alpha_k = \alpha$ at each step, whereas other methods will allow α_k to vary. For example, Newton's method uses $\alpha_k = [D^2 f(\mathbf{x}_n)]^{-1}$, assuming the Hessian matrix is invertible, and the steepest descent method optimises for α_k at each iteration.

1. Consider the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(\mathbf{x}) = 3(x_1^2 + x_2^2) + 4x_1x_2 + 5x_1 + 6x_2.$$

   (a) Write $g(\mathbf{x})$ in the form $\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}$ for suitable matrices $\mathbf{Q}$ and $\mathbf{b}$.

   (b) Determine ∇g, and then implement g and ∇g as MATLAB functions.

   (c) By referring to the results referenced at the end of Section 3.5 of the reading materials, determine the largest range of values of α for which the fixed-step-size algorithm is globally convergent.

   (d) Implement the fixed-step-size algorithm, using $\alpha = 1/10$, with the following stopping condition:

$$|g(\mathbf{x}_{k+1}) - g(\mathbf{x}_k)| < 10^{-6}.$$

   Starting with $\mathbf{x}_0 = \mathbf{0}$, how many iterations are needed before the algorithm terminates?

   (e) Implement the steepest descent method using the explicit formula for α_k given at the end of Section 3.5 of the reading materials. Use the same stopping condition used in (d). Starting with $\mathbf{x}_0 = \mathbf{0}$, how many iterations are needed before the algorithm terminates?

2. Consider the Rosenbrock function $h\colon \mathbb{R}^2 \to \mathbb{R}$,

$$h(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

   (a) Show that $\begin{pmatrix}1 & 1\end{pmatrix}^T$ is the unique global minimizer of h.

   (b) Determine ∇h and $D^2 h$, and then implement h, ∇h and $D^2 h$ as MATLAB functions.

   (c) Implement Newton's method for h using the following stopping condition:

$$\left|\frac{h(\mathbf{x}_{k+1}) - h(\mathbf{x}_k)}{h(\mathbf{x}_k)}\right| < 10^{-10}.$$

   Starting with $\mathbf{x}_0 = \mathbf{0}$, how many iterations are needed before the algorithm terminates?

   (d) Try a fixed-step-size gradient method, with $\alpha = 1/100$, starting at $\mathbf{0}$. What happens?

   (e) Implement the downhill simplex method for h, with an initial 2-simplex of $\begin{pmatrix}0 & 0\end{pmatrix}^T$, $\begin{pmatrix}0 & 1\end{pmatrix}^T$ and $\begin{pmatrix}1 & 0\end{pmatrix}^T$. **Do not** use the code at the end of Section 3.7; instead, start with the list of steps given in Section 3.6, and perform simplified calculations that are applicable for the case $n = 2$. For example, sorting the vertices can be done by directly comparing their function values and swapping if they are not in the right order, and the centre of mass calculation takes the form $(\mathbf{x}_1 + \mathbf{x}_2)/2$. Perform 100 iterations, and use the typical values $\alpha = 1$, $\gamma = 2$, $\rho = \sigma = \frac{1}{2}$.

Regression revisited

3. Consider the following data (from Workshop 4):

i	1	2	3	4	5	6	7
x_i	0.28	0.76	0.93	1.88	3.03	4.73	4.90
y_i	1.62	1.22	1.80	1.03	1.17	0.70	0.22

   We wish to fit to a power model $y = ae^{bx}$ using least squares regression. Implement the objective function that needs minimisation, and then use the downhill simplex method to find the coefficients. You will be able to re-use your code from Question 2(e).