C477: Computational Optimisation
Tutorial 4: First-Order Methods

Exercise 1. Given a continuously differentiable, i.e., $f \in \mathcal{C}^1$, function $f : \mathbb{R}^n \to \mathbb{R}$ consider the general iterative algorithm,

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)},$$

where $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \ldots, \boldsymbol{d}^{(k)}, \ldots$ are given vectors in $\mathbb{R}^n$ and α_k is chosen to minimise $f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)})$; that is,

$$\alpha_k = \arg\min_{\alpha} \ \phi_k(\alpha) = \arg\min_{\alpha} \ f\left(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}\right).$$

Assume the gradient exists and show that for each k, the vector $\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$ is orthogonal to $\nabla f\left(\boldsymbol{x}^{(k+1)}\right)$.

Recall: Using the chain rule: $\phi_k'(\alpha) = \nabla f\left(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}\right)^\top \boldsymbol{d}^{(k)}$.

Exercise 2. Write a simple Matlab program that implements the steepest descent algorithm using an exact line search. You may use `fmincon` to compute the step size optimisation subproblem. For the stopping criterion, use the condition:

$$\left\|\nabla f(\boldsymbol{x}^{(k)})\right\|_2 \leq \epsilon,$$

where $\epsilon = 10^{-6}$. Test your program by solving the following problem,

$$\min\ (x_1 - 4)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$$

Use the initial condition $[-4,\ 5,\ 1]^\top$.

Exercise 3. Apply the Matlab program developed above to the following problem,

$$\min \ 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 .$$

Use the initial condition $[-2,\, 2]^\top$. Terminate the algorithm when the norm of the gradient of f is less than 10^{-4}.

Note: This function is called the Rosenbrock function[1] and it poses great difficulty to many algorithms.

Exercise 4. Consider the minimisation problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x},$$

where $\boldsymbol{Q} \succ \mathbf{0}$ is a positive definite 2×2 matrix. Suppose we use the diagonal scaling matrix:

$$\boldsymbol{D} = \begin{pmatrix} Q_{11}^{-1} & 0 \\ 0 & Q_{22}^{-1} \end{pmatrix}.$$

Show that this scaling matrix $\boldsymbol{D}$ improves the condition number of $\boldsymbol{Q}$ in the sense that:

$$\chi\left(\boldsymbol{D}^{1/2}\boldsymbol{Q}\boldsymbol{D}^{1/2}\right) \leq \chi\left(\boldsymbol{Q}\right).$$

Exercise 5. Suppose that we are given a Lipschitz continuous function $f : [x^L, x^U] \mapsto \mathbb{R}$ and told that the Lipschitz constant is $L \in \mathbb{R}$. How many function evaluations do we have to perform (in the worst case) to come within ϵ of the global solution?

[1] `http://en.wikipedia.org/wiki/Rosenbrock_function`