C6.2/B2. Continuous Optimization

Problem Sheet 1

Please hand-in for marking Problems 1 (i, iii–v), 2 and 3; please note that Problem 1 can be found with proof in various optimization textbooks and you are welcome to have a look. The other problems are optional/for revision.

1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We say that f is convex if and only if

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y), \quad \text{for all } x \text{ and } y \text{ in } \mathbb{R}^n, \text{ and any } \lambda \in [0,1]. \tag{1}$$

   Prove the following statements:

   (i) If f is convex, then $x^* \in \mathbb{R}^n$ is a local minimizer of f if and only if it is a global minimizer.

   (ii) (optional) Assume that $f \in \mathcal{C}^1(\mathbb{R}^n)$. Then f is convex if and only if for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, we have

$$f(y) \ge f(x) + \nabla f(x)^\top (y - x). \tag{2}$$

   (*Comment: this property means that geometrically, the graph of the first order approximation of f at x lies below the graph of f.*)

   (iii) If $f \in \mathcal{C}^1(\mathbb{R}^n)$ is convex and x^* is a stationary point of f (i.e., $\nabla f(x^*) = 0$), then x^* is a global minimizer of f.

   (iv) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$, $x \in \mathbb{R}^n$ and $0 \ne s \in \mathbb{R}^n$. Write down the second-order Taylor expansion or second-order mean-value theorem of the (univariate) function $\alpha \to f(x + \alpha s)$ around $\alpha = 0$.

   (v) Using (iv), show that $f \in \mathcal{C}^2(\mathbb{R}^n)$ is convex if and only if $\nabla^2 f(x)$ is positive semi-definite for all $x \in \mathbb{R}^n$ (i.e., $s^T \nabla^2 f(x) s \ge 0$ for all $s \in \mathbb{R}^n$.)

2. Consider the function

$$f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2, \quad x = (x_1 \;\; x_2)^T \in \mathbb{R}^2.$$

   (a) Compute the gradient vector and the Hessian matrix of f at (any) $x \in \mathbb{R}^2$. Find all stationary points of f. Show that $x^* = (1 \;\; 1)^T$ is the unique global minimizer of f and that the Hessian of f at x^* is positive definite.

   (b) Show that the Hessian matrix $\nabla^2 f(x)$ of f is singular if and only if x satisfies the condition

$$x_2 - x_1^2 = 0.05.$$

   Hence show that $\nabla^2 f(x)$ is positive definite for all x such that $f(x) < 0.025$.

   (c) Show that f is not a convex function.

3. Show that the function

$$f(x) = (x_2 - x_1^2)^2 + x_1^5$$

   has only one stationary point which is neither a local maximum nor a local minimum.

4. Suppose that $g \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n \times n}$ are constant, H is a symmetric matrix and that the quadratic function $q : \mathbb{R}^n \mapsto \mathbb{R}$ is defined by $q(x) = g^T x + \frac{1}{2} x^T H x$. By writing q in terms of the entries in g and H, show that $\nabla q(x) = g + Hx$ and $\nabla^2 q(x) = H$. Then show that if H is positive semidefinite, then $q(x)$ is a convex function; if H is negative semidefinite, then $q(x)$ is a concave function.

Consider minimizing $q(x)$ by applying a generic linesearch method with search directions s^k and exact linesearch. Show that if $(s^k)^T H s^k > 0$, the exact linesearch is well-defined and has the following explicit expression for the stepsize α_k,

$$\alpha_k = -\frac{\nabla q(x^k)^T s^k}{(s^k)^T H s^k}.$$

(*Comment: The solution to the second part of this problem can be found in the lecture slides.*)

5. Let $\Phi : \mathbb{R} \to \mathbb{R}$ be a univariate (i.e., one variable) nonlinear function $\Phi = \Phi(\alpha)$. Consider approximating Φ by a quadratic function $q(\alpha) = a\alpha^2 + b\alpha + c$, for some $a, b, c \in \mathbb{R}$, such that

$$q(0) = \Phi(0), \quad q'(0) = \Phi'(0) \quad \text{and} \quad q(\alpha_0) = \Phi(\alpha_0), \tag{3}$$

for some $\alpha_0 > 0$; we say that q *interpolates* Φ at these points. Find the values of a, b and c (in terms of the known quantities $\Phi(0)$, $\Phi'(0)$ and $\Phi(\alpha_0)$) such that the conditions (3) are satisfied. Then find a condition that $\Phi(0)$, $\Phi'(0)$ and $\Phi(\alpha_0)$ need to satisfy to ensure that q has a (global) minimizer. Also, find conditions (on the same quantities) such that $\Phi(\alpha)$ and $q(\alpha)$ are guaranteed to have a minimizer in the interval $(0, \alpha_0)$.
(*Comment: The interpolation approach above is used to numerically approximate the exact linesearch stepsize for nonlinear, nonquadratic functions. In particular, in a generic linesearch method applied to minimizing some function f (see GLM in the handouts), let $\Phi(\alpha) := \Phi_k(\alpha) = f(x^k + \alpha s^k)$, and set the stepsize α^k to the minimizer of $q(\alpha)$ above (which approximates the minimizer of $\Phi(\alpha)$); alternatively, replace one of the interpolation points, 0 or α_0, by the minimizer of $q(\alpha)$, and repeat the interpolation process. Note that cubic polynomials may also be used for interpolation as long as the interpolation points are carefully chosen so that Φ, and the interpolating polynomial, has a minimizer in the interval determined by these points.*)

6. Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Consider applying the generic linesearch method (GLM) to minimizing f starting from $x^0 = 2$.

   (i) Let the directions in GLM be $s^k := (-1)^{k+1}$ and the stepsizes $\alpha^k := 2 + 3/2^{k+1}$. Write down the expression of the iterates x^k generated by the GLM and plot the pairs $(x^k, f(x^k))$ on the graph of f. What do you observe? Show that the sequence $\{x^k\}$ has two limit points: $+1$ and -1. Is any of these points a stationary point of f?

   (ii) Similarly, let now $s^k := -1$ and $\alpha^k = 1/2^{k+1}$. Again, write down the expression of the iterates x^k generated by the GLM and plot the pairs $(x^k, f(x^k))$ on the graph of f. What do you observe? Show that $\{x^k\}$ converges to 1.

(*Comment: For plots, see the lecture slides on GLM-inexact linesearch. This problem illustrates that even when the search directions are descent in a generic linesearch method (GLM) and the stepsize ensures the function values at the iterates decrease, the GLM may not be convergent to a minimizer or stationary point of our objective f; the amount of decrease the stepsize gives in f in relation to its length is crucial. In particular, case i) above illustrates that the stepsize cannot be "too long" when it yields little decrease in f; case ii) exemplifies that that stepsize cannot be "too short" as it cuts the direction too much and yields little progress. Recall lecture slides.*)