Exercise 2

January 21, 2025

Problem 1.

Define the function $f\colon \mathbb{R}^3 \to \mathbb{R}$,

$$f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9.$$

a) Assess whether this function is (strictly) convex.

b) Find all local and global minimisers of f.

Solution.

The first order necessary condition for the optimisation problem given in the task, implies

$$\nabla f(x, y, z) = (4x + y - 6, x + 2y + z - 7, y + 2z - 8)^T = 0.$$

Now we have the following system of three equation

$$\begin{aligned} 4x + y &= 6, \\ x + 2y + z &= 7, \\ y + 2z &= 8. \end{aligned} \tag{1}$$

By solving (1), we obtain the critical point $(x, y, z) = \left(\frac{6}{5}, \frac{6}{5}, \frac{17}{5}\right)$. Now, we find the Hessian matrix

$$H_f = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

Note here that the Hessian is a constant matrix, independent of (x, y, z). Since all leading principal minors of this matrix are positive (alternatively, one has that the matrix is irreducibly diagonally dominant), it is positive definite, and thus the function f is strictly convex. We can therefore conclude that the optimisation problem has the unique global minimiser $\left(\frac{6}{5}, \frac{6}{5}, \frac{17}{5}\right)$.

Problem 2.

Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a convex function.

a) Show that the set of minimisers of f is a convex set.

b) Assume that f is strictly convex. Show that the problem $\min_{x\in\mathbb{R}^d} f(x)$ has at most one global solution. In addition, find a strictly convex function f that has no global minimiser at all.

Solution.

a) Let x, y be minimisers of f. Denote

$$v := f(x) = f(y) = \min_{z\in\mathbb{R}^d} f(z).$$

Then we have for all $0 < \lambda < 1$ that

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) = z = \min_{z\in\mathbb{R}^d} f(z),$$

which shows that $\lambda x + (1-\lambda)y$ is a minimiser of f as well. This shows that the set of minimisers of f is convex.

b) We assume to the contrary that this problem has two distinct minimisers, say, $x_1 \neq x_2 \in \mathbb{R}^d$, and we denote

$$v := f(x_1) = f(x_2) = \min_{x\in\mathbb{R}^d} f(x). \tag{2}$$

Since, the objective function f is strictly convex, we have

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2),\ \forall\lambda \in (0,1). \tag{3}$$

By combining inequalities (2) and (3), we obtain that

$$f(\lambda x_1 + (1-\lambda)x_2) < v,$$

which is not possible because we cannot have a function value of f smaller than the minimum. Thus the optimisation problem has a unique solution.

As an example of a strictly convex function that has no minimum, we consider the function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$. This function is obviously strictly convex (its second derivative is always positive), but its infimum 0 is not attained. □

Problem 3.

Show that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$,

$$f(x,y) = \log\big(e^x + e^y\big)$$

is convex.

Solution.

We have that

$$\nabla f(x,y) = \left(\frac{e^x}{e^x+e^y}, \frac{e^y}{e^x+e^y}\right)^T$$

and

$$H_f(x,y) = \frac{e^{x+y}}{(e^x+e^y)^2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Evidently, $\frac{e^{x+y}}{(e^x+e^y)^2} > 0$ for all $(x,y) \in \mathbb{R}^2$, and the matrix $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ is symmetric and positive semi-definite (with eigenvalues 0 and 2). Thus $H_f(x,y)$ is positive semi-definite for all $(x,y) \in \mathbb{R}^2$, which shows that f is convex.

Problem 4.

Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is a function satisfying $f(x) > 0$ for all $x \in \mathbb{R}^d$ and define $g\colon \mathbb{R}^d \to \mathbb{R}$, $g(x) := \log(f(x))$. Assume that g is convex. Show that f is then convex as well.

Solution.

Define the function $g\colon \mathbb{R}^n \to \mathbb{R}$, $g(x) = \log(f(x))$, so that $f(x) = \exp(g(x))$. By assumption, the function g is convex, and we need to prove that the function f is convex.

Convexity of g implies that, for all x, $y \in \mathbb{R}^d$ and $\lambda \in [0,1]$, we have the inequality

$$g\big(\lambda x + (1-\lambda)y\big) \le \lambda g(x) + (1-\lambda)g(y).$$

This inequality with the fact that the exponential function is monotonically increasing yields that

$$\exp\big(g(\lambda x + (1-\lambda)y)\big) \le \exp\big(\lambda g(x) + (1-\lambda)g(y)\big).$$

This inequality can be rewritten as

$$f\big(\lambda x + (1-\lambda)y\big) \le \exp\big(\lambda g(x) + (1-\lambda)g(y)\big).$$

Now, we use the convexity of exp in the above inequality and obtain that

$$f(\lambda x + (1-\lambda)y) \le \lambda \exp(g(x)) + (1-\lambda)\exp(g(y)) = \lambda f(x) + (1-\lambda)f(y).$$

Therefore, the function f is convex. □

Problem 5.

a) Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ (see Exercise 1, Problem 3b),

$$f(x, y) = 2x^2 - 4xy + y^4 + 5y^2 - 10y.$$

Perform one step of the gradient descent method with backtracking (Armijo) line search starting from the point $x_0 = (0, 0)$. Start with an initial step length $\alpha = 1$ and use the parameters $c = 0.1$ (sufficient decrease parameter) and $\rho = 0.1$ (contraction factor).

b) Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$,

$$f(x, y) = x^4y^2 + x^4 - 2x^3y - 2x^2y - x^2 + 2x + 2.$$

Perform one step of the gradient descent method with backtracking (Armijo) line search starting from the point $x_0 = (0, 0)$. Start with an initial step length $\alpha = \frac{1}{2}$ and use the parameters $c = \frac{1}{2}$ (sufficient decrease parameter) and $\rho = 0.1$ (contraction factor).

Solution.

a) First we find the search direction p_0 from the starting point $x_0 = (0, 0)$, which is $p_0 = -\nabla f(x_0)^T = (0, 10)^T$. Now the Armijo condition at x_0 and p_0 with parameter $c = 0.1$ gives

$$10^4\alpha^4 + 500\alpha^2 \le 90\alpha. \tag{4}$$

Since $\alpha = 1$ does not satisfy the inequality (4), we cannot take step length $\alpha = 1$. Thus we try $\alpha = 0.1$, which satisfies (4). Therefore, we choose the step length $\alpha = 0.1$. Thus the next iterate in the gradient descent method is $x_1 = x_0 + \alpha p_0 = (0, 1)$.

b) First we find the search direction p_0 from the starting point $x_0 = (0, 0)$, which is $p_0 = -\nabla f(x_0)^T = (-2, 0)^T$. Now the Armijo condition at x_0 and p_0 with parameter $c = \frac{1}{2}$, $f(x_0 + \alpha p_0) \le f(x_0) + c\alpha\nabla f(x_0)^T p_0$ gives

$$16\alpha^4 - 4\alpha^2 - 4\alpha + 2 \le 2 - 2\alpha.$$

With the initial step length $\alpha = \frac{1}{2}$, this inequality is satisfied. Thus we can choose the step length $\alpha = \frac{1}{2}$ and the next iterate of gradient descent method is $x_1 = x_0 + \alpha p_0 = (-1, 0)$.

Problem 6.

Assume that the sequence $\{x_k\}_{k\in\mathbb{N}}$ is generated by the gradient descent method with backtracking (Armijo) line search for the minimisation of a continuously differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}$, and that $\nabla f(x_k) \neq 0$ for all k. Moreover, assume that $\bar{x}$ is an accumulation point[1] of the sequence $\{x_k\}_{k\in\mathbb{N}}$. Show that $\bar{x}$ is not a local maximum of f.

Remark: It is in theory possible, if highly unlikely, that one of the iterates in the gradient descent method turns out to be a critical point, in which case the iteration terminates. If this is not the case, then this result shows that the gradient descent method at least will not converge to a maximum of f.

Solution.

Since the sequence x_k is generated by using a backtracking line search method, it satisfies the Armijo condition

$$f(x_{k+1}) = f(x_k + \alpha_k p_k) \le f(x_k) + c\alpha_k\langle\nabla f(x_k), p_k\rangle,$$

with $p_k = -\nabla f(x_k) \neq 0$ and $\alpha_k > 0$, which implies that

$$f(x_{k+1}) \le f(x_k) - c\alpha_k\|\nabla f(x_k)\|^2 < f(x_k),$$

and thus $f(x_{k+1}) < f(x_k)$. Therefore, the sequence $\{f(x_k)\}_{k\in\mathbb{N}}$ is strictly decreasing. Now assume that $\overline{x}$ is an accumulation point of the sequence $\{x_k\}_{k\in\mathbb{N}}$. Then there exists a subsequence $\{x_{k'}\}$ converging to $\overline{x}$. Moreover, f is a continuous function, and therefore $f(x_{k'}) \to f(\overline{x})$ too. Since $\{f(x_{k'}\}_{k\in\mathbb{N}}$ is strictly decreasing, it follows that $f(x_{k'}) > f(\overline{x})$ for all k'. Thus we have found a sequence converging to $\overline{x}$, where all the function values are larger than $f(\overline{x})$. This shows that $\overline{x}$ cannot be a local maximum of f. $\square$

Problem 7.

We say that a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *quasi-convex*, if for every $\alpha \in \mathbb{R}$, the level set $L_f(\alpha)$ is convex.

a) Show that every convex function is quasi-convex.

b) Find a quasi-convex function that is not convex.

c) Show that a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is quasi-convex, if and only if for all $x, y \in \mathbb{R}^d$ and $0 < \lambda < 1$ we have that

$$f\big(\lambda x + (1-\lambda)y\big) \le \max\big\{f(x), f(y)\big\}.$$

d) Show by means of an example that a local minimum of a quasi-convex function need not be a global minimum.

   Hint: Consider a function on $\mathbb{R}$ that is locally constant.

[1] Recall that $\bar{x}$ is an accumulation point of the sequence $\{x_k\}_{k\in\mathbb{N}}$, if there exists a subsequence $\{x_{k'}\}_{k'}$ with $\bar{x} = \lim_{k'} x_{k'}$.

Solution.

a) Suppose that $f\colon \mathbb{R}^d \to \mathbb{R}$ is a convex function, and let α be some arbitrary real number. Let $x, y \in L_f(\alpha)$, and $0 < \lambda < 1$. Then

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

This shows that $\lambda x + (1-\lambda)y$ is in $L_f(\alpha)$, showing the convexity of $L_f(\alpha)$.

b) We use the function $f(x) = x^3$. We see easily that this is not convex (look at the second derivative, which is negative for $x < 0$ to see this). However, the level set of any α is given by $L_f(\alpha) = (-\infty, (\alpha)^{\frac{1}{3}}]$, which is a convex set, and so $f(x)$ is quasi-convex.

c) We start by showing that if f is quasi-convex, then for any $x, y \in \mathbb{R}^d$ and $0 < \lambda < 1$, we have $f\big(\lambda x + (1-\lambda)y\big) \leq \max\big\{f(x), f(y)\big\}$:

Suppose to that end that f is quasi-convex. Let $x, y \in \mathbb{R}^d$ and $0 < \lambda < 1$. Then $L_f(\max\big\{f(x), f(y)\big\})$ is a convex set (by problem a), so we get $\lambda x + (1-\lambda)y \in L_f(\max\big\{f(x), f(y)\big\})$, implying that $f\big(\lambda x + (1-\lambda)y\big) \leq \max\big\{f(x), f(y)\big\}$.

Next, we show that, if $f\big(\lambda x + (1-\lambda)y\big) \leq \max\big\{f(x), f(y)\big\}$ for every $x, y \in \mathbb{R}^d$ and $0 < \lambda < 1$, then f is quasi-convex.

Assume that $f\big(\lambda x + (1-\lambda)y\big) \leq \max\big\{f(x), f(y)\big\}$ for any $x, y \in \mathbb{R}^d$ and $0 < \lambda < 1$. Let $\alpha \in \mathbb{R}$, and let $x, y \in L_f(\alpha)$, and $0 < \lambda < 1$. Then $f(x) \leq \alpha$, and $f(y) \leq \alpha$, meaning $\max\{f(x), f(y)\} \leq \alpha$. By assumption, $f\big(\lambda x + (1-\lambda)y\big) \leq \max\big\{f(x), f(y)\big\} \leq \alpha$, implying that $\lambda x + (1-\lambda)y \in L_f(\alpha)$. So the level set $L_f(\alpha)$ is convex, and we are done.

d) We define the function

$$f(x) = \begin{cases} x \quad \text{for } 0 < x \\ 0 \quad \text{for } -1 < x \leq 0 \\ x+1 \quad \text{for } -2 < x \leq -1 \\ -x-3 \quad \text{for } x \leq -2. \end{cases}$$

Then $f(x)$ is quasi-convex, and every $x \in (-1, 0]$ is a local minimiser of f. However, the global minimum is at $x = -2$.