Exercise #2

January 24, 2023

Problem 1.

a) Show that a (not necessarily differentiable) function $f\colon \mathbb{R}^n \mapsto \mathbb{R}_{>0}$ is convex, if the function $x \mapsto \log(f(x))$ is convex.

b) Show that an optimization problem $\min_{x\in\mathbb{R}^n} f(x)$ has at most one global minimizer if the objective function $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ is strictly convex. In addition, find a strictly convex objective function f that has no global minimizer at all.

Solution.

a) Define a function $g\colon \mathbb{R}^n \mapsto \mathbb{R}$ such that $g(x) = \log(f(x))$, it follows that $f(x) = \exp(g(x))$. Now, we have the function g is convex, and need to prove that the function f is convex.
Convexity of the function g implies that for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$, we have the following inequality

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

The above inequality with the fact that exp (exponential function) is monotonic increasing yields that

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

The above inequality can be rewritten as

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

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

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

Therefore, the function f is convex. □

b) We assume to the contrary that this problem has two distinct minimizers, say, $x_1, x_2 \in \mathbb{R}^n$, such that

$$f(x_1) = f(x_2) = \min f. \tag{1}$$

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

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2), \ \forall \lambda \in (0,1). \tag{2}$$

By combining inequalities (1) and (2), we obtain

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

which is not possible because we cannot have a objective value of the function f smaller than the minimal value of the function f (i.e., $\min f$). Therefore, the mentioned optimization problem has unique solution.

Further, we define a function $f: \mathbb{R} \mapsto \mathbb{R}$ as $f(x) = e^x$. We have the following

$$f'(x) = e^x,$$

$$f''(x) = e^x > 0, \ \forall x \in \mathbb{R}.$$

Therefore, f is strictly convex function. However, $f'(x) = e^x \neq 0$ for any $x \in \mathbb{R}$. Therefore, the optimization problem $\min_{x \in \mathbb{R}} f(x)$ has no global minimizer, though the objective function f is strictly convex. □

Problem 2.

Show that the function $f: \mathbb{R}^2 \mapsto \mathbb{R}$,

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

is convex.

Solution.

Further, for the given objective function $f: \mathbb{R}^2 \mapsto \mathbb{R}$ as $f(x, y) = \log(e^x + e^y)$, we get

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

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

Evidently, $\frac{e^{x+y}}{(e^x+e^y)^2} > 0$ for all $(x, y) \in \mathbb{R}^2$, and let's say $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ which is symmetric matrix and has eigen values 0 and 2. Therefore Hessian matrix is positive semi-definite for all $(x, y) \in \mathbb{R}^2$. Consequently, the given function f is convex.

Problem 3.

Consider the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

where the objective function $f: \mathbb{R}^3 \mapsto \mathbb{R}$ is defined as

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

Prove that this optimization problem has a unique global minimizer and find it.

Solution.

The first order necessary condition for the optimization problem (??) 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

$$\begin{aligned} 4x + y &= 6, \\ x + 2y + z &= 7, \\ y + 2z &= 8. \end{aligned} \tag{3}$$

By solving (3), 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

$$\nabla^2 f = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

The approximate eigen values of the Hessian matrix are $4.48,\ 2.69,$ and 0.83. It is evident that the Hessian matrix is symmetric and has non-zero positive eigenvalues. Therefore, Hessian matrix is positive definite and consequently the objective function f is strictly convex. Eventually, we can conclude that the optimization problem has unique global minimizer $\left(\frac{6}{5}, \frac{6}{5}, \frac{17}{5}\right)$.

Problem 4.

a) Consider the function $f\colon \mathbb{R}^2 \mapsto \mathbb{R}$ (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 lenght $\alpha = 1$ and use the parameters $c = 0.1$ (sufficient decrease parameter) and $\rho = 0.1$ (contraction factor).

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

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

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 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

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

Since $\alpha = 1$ does not satisfy the inequality (4), we cannot take step length $\alpha = 1$. Then 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 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 $c = \frac{1}{2}$, $f(x_0 + \alpha p_0) \leq f(x_0) + c\alpha \nabla f(x_0)^T p_0$ gives

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

The initial step length $\alpha = \frac{1}{2}$ in the above inequality implies

$$0 \leq 1.$$

Therefore, $\alpha = \frac{1}{2}$ satisfies the Armijo condition. Now, we can choose the step length $\alpha = \frac{1}{2}$. Thus the next iterate of gradient descent method is $x_1 = x_0 + \alpha p_0 = (-1, 0)$.

Problem 5.

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

b) We consider a line search method of the form $x_{k+1} = x_k + \alpha_k p_k$ for the minimization of the function $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ with the search direction p_k given as

$$p_k = -\operatorname{sgn}\big((\nabla f(x_k))_i\big) e_i,$$

where the index i is chosen such that $|(\nabla f(x_k))_i|$ is maximal. Here e_i with $1 \le i \le n$ denotes i^{th} standard basis vector in $\mathbb{R}^n$. Show that the direction p_k is a descent direction whenever x_k is not a stationary point of f (that is, $\nabla f(x_k) \neq 0$).

Solution.

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

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

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

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

which implies that $f(x_{k+1}) < f(x_k)$. Therefore, the sequence $\{f(x_k)\}_{k\in\mathbb{N}}$ is strictly decreasing. Now, we have $\overline{x}$ is an accumulation point of the sequence $\{x_k\}_{k\in\mathbb{N}}$. Thus there exists a subsequence $\{x_{k'}\}$ converging to $\overline{x}$. Moreover, f is continuous function, therefore, $f(x_{k'}) \longrightarrow f(\overline{x})$ too. Since $f(x_k)$ is strictly decreasing, $f(x_{k'})$ is also strictly decreasing sequence, and we have $f(x_{k'}) \longrightarrow f(\overline{x})$, implying that $f(x_{k'}) > f(\overline{x})$ for every k' (because bounded decreasing sequence converges to its infimum (greatest lower bound), and the sequence $f(x_{k'})$ is convergent and hence bounded too), which in turn shows that $\overline{x}$ is not a local maximum of f. □

b) We recall that p_k is a descent direction for f at x_k if and only if $p_k^T \nabla f(x_k) < 0$, which we have to prove. To this end, assume that x_k is not a stationary point of f, that is, $\nabla f(x_k) \neq 0$. Since the index i in the direction of p_k is chosen in such a way that $|(\nabla f(x_k))_i|$ is maximal, we obtain in particular that $|(\nabla f(x_k))_i| > 0$. Thus

$$\begin{aligned}
p_k^T \nabla f(x_k) &= -\operatorname{sgn}((\nabla f(x_k))_i) e_i^T \nabla f(x_k) \\
&= -\operatorname{sgn}((\nabla f(x_k))_i)(\nabla f(x_k))_i \\
&= -\frac{|(\nabla f(x_k))_i|}{(\nabla f(x_k))_i}(\nabla f(x_k))_i \\
&= -|(\nabla f(x_k))_i| < 0.
\end{aligned}$$

Therefore, p_k is the descent direction. □