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) Have already discussed in the class!
For your future convenience, some hints are given below:
Define a function $g\colon \mathbb{R}^n \mapsto \mathbb{R}$ such that $g(x) = \log(f(x))$, it follows that $f(x) = \exp(g(x))$, then use the monotonic increasing and convexity property of exponential (exp) function.

b) Have already discussed in the class!
For your future convenience, some hints are given below:
Start proving by assuming to the contrary that this problem has two distinct minimizers, say, $x_1, x_2 \in \mathbb{R}^n$, such that

$$f(x_1) = f(x_2) = \min f,$$

and use the strict convexity property of f.

Problem 2.

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

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

is convex.

Solution.

Have already discussed in the class!
For your future convenience, some hints are given below:
Prove this by proving positive semi-definiteness of the Hessian matrix.

Problem 3.

Consider the optimization problem

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

where the objective function $f\colon \mathbb{R}^3 \mapsto \mathbb{R}$ 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

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

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

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

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.

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

$$f(x,y) = \frac{x^2}{2} + x\cos y.$$

We want to perform one step of a line search method with initial value $x_0 = (1, \frac{\pi}{4})$ and search direction $p_0 = (-1, 0)$.

a) Confirm that p_0 is a descent direction from the initial point x_0.

b) State the Armijo condition. What is the range of admissible values for the step length α, if a parameter $c = 0.1$ is used?

c) Perform one step of the line search method using the optimal value of α as step length.

Solution.

a) The search direction p_0 is a descent direction from the initial point x_0, if $\nabla f(x_0)^T p_0 < 0$. We have

$$\nabla f(x) = (x + \cos y, -x \sin y)^T,$$

which implies

$$\nabla f(x_0) = \nabla f\left(1, \frac{\pi}{4}\right) = \left(1 + \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)^T.$$

Thus, $\nabla f(x_0)^T p_0 = -1 - \frac{1}{\sqrt{2}} < 0$.

b) The Armijo condition is given as

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c\alpha_k \nabla f(x_k)^T p_k.$$

For $x_0 = \left(1, \frac{\pi}{4}\right)$ and $p_0 = (-1, 0)$, the Armijo condition $f(x_0 + \alpha p_0) \leq f(x_0) + c\alpha \nabla f(x_0)^T p_0$ gives

$$\frac{(1-\alpha)^2}{2} + \frac{(1-\alpha)}{\sqrt{2}} \leq \frac{1}{2} + \frac{1}{\sqrt{2}} - c\alpha\left(1 + \frac{1}{\sqrt{2}}\right)$$

$$\begin{aligned}
&\Rightarrow \frac{(1-\alpha)(1-\alpha+\sqrt{2})}{2} \leq \frac{1+\sqrt{2}}{2} - c\alpha\frac{(2+\sqrt{2})}{2} \\
&\Rightarrow \alpha(\alpha - 2 - \sqrt{2} + c(2+\sqrt{2})) \leq 0 \text{ (since } \alpha > 0) \\
&\Rightarrow \alpha - (2+\sqrt{2}) + c(2+\sqrt{2}) \leq 0 \\
&\Rightarrow \alpha \leq (1-c)(2+\sqrt{2}).
\end{aligned}$$

By putting the value $c = 0.1$, we obtain the following admissible value range of α

$$0 < \alpha \leq 3.0727.$$

c) One step with the line search method means to solve the following one-dimensional optimization problem

$$\min_{\alpha>0} f(x_0 + \alpha p_0) = \min_{\alpha>0}\left\{\frac{(1-\alpha)^2}{2} + \frac{1-\alpha}{\sqrt{2}}\right\}.$$

Say, $\phi(\alpha) = \frac{(1-\alpha)^2}{2} + \frac{1-\alpha}{\sqrt{2}}$. Then, $\nabla\phi(\alpha) = 0$ gives the solution $\alpha = 1 + \frac{1}{\sqrt{2}}$. The new iterate is then $x_1 = x_0 + \alpha_0 p_0$, which gives $x_1 = \left(-\frac{\sqrt{2}}{2}, \frac{\pi}{4}\right)$.

Problem 5.

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) Have already discussed in the class!
A hint is $\alpha = 1$ does not satify the Armijo condition so you have to check for another α, let's say for $\alpha = 0.1$.

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 $c = \frac{1}{2}$, $f(x_0 + \alpha p_0) \le f(x_0) + c\alpha \nabla f(x_0)^T p_0$ gives

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

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

$$0 \le 1.$$

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 6.

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 $\bar{x}$ is an accumulation point of the sequence $\{x_k\}_{k\in\mathbb{N}}$. Thus there exists a subsequence $\{x_{k'}\}$ converging to $\bar{x}$. Moreover, f is continuous function, therefore, $f(x_{k'}) \longrightarrow f(\bar{x})$ too. Since $f(x_k)$ is strictly decreasing, $f(x_{k'})$ is also strictly decreasing sequence, and we have $f(x_{k'}) \longrightarrow f(\bar{x})$, implying that $f(x_{k'}) > f(\bar{x})$ 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 $\bar{x}$ 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

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

$$= -\operatorname{sgn}((\nabla f(x_k))_i)(\nabla f(x_k))_i$$
$$= -\frac{|(\nabla f(x_k))_i|}{(\nabla f(x_k))_i}(\nabla f(x_k))_i$$

$$= -|(\nabla f(x_k))_i| < 0.$$

Therefore, p_k is the descent direction. □