Chair of Decision Sciences and Systems
TUM School of Computation, Information and Technology
Technical University of Munich

Business Analytics & Machine Learning
Tutorial sheet 11: Convex Optimization – Solution

Prof. Dr. Martin Bichler, Prof. Dr. Jalal Etesami
Julius Durmann, Markus Ewert, Johannes Knörr, Yutong Chao
January 3, 2024

Exercise T11.1 *Convex function*

You are given the following function:

$$f(x, y) = a \exp(3x) + \frac{b}{2}xy + y^2$$

Determine all parameters $a, b \in \mathbb{R}$ such that f is convex.

Solution

We determine the gradient and the Hessian of f.

$$\nabla f(x, y) = \begin{pmatrix} 3a \exp(3x) + \frac{b}{2}y \\ \frac{b}{2}x + 2y \end{pmatrix}$$

$$\nabla^2 f(x, y) = \begin{pmatrix} 9a \exp(3x) & \frac{b}{2} \\ \frac{b}{2} & 2 \end{pmatrix}$$

f is convex iff $\nabla^2 f(x, y)$ is positive semidefinite for all $x, y \in \mathbb{R}$. This is the case iff all principal minors are non-negative.

- The first principal minor is $H_1(x, y) = 9a \exp(3x)$. $H_1 \geq 0$ is equivalent to $a \geq 0$.
- The second principal minor is $H_2(x, y) = 2 \geq 0$.
- The third principal minor is $H_3(x, y) = 18a \exp(3x) - \frac{1}{4}b^2$. $H_3(x, y) \geq 0$ for $a \geq 0$ and for all x, y iff $b = 0$. For $b \neq 0$, we can always find a sufficiently small $x < 0$ such that $18a \exp(3x) < \frac{1}{4}b^2$ and thus $H_3(x, y) < 0$.

Thus, f is convex for $a \geq 0, b = 0$.

Exercise T11.2 *Operations preserve convexity*

You are given the following convex functions $g_1(x), g_2(x)$. Prove that the following functions are also convex functions:

- $h_1(x) = g_1(Ax + b)$ where A is a matrix and b is a vector.
- $h_2(x) = C_1 g_1(x) + C_2 g_2(x)$, where C_1 and C_2 are nonnegative constant.
- $h_3(x) = \max\{g_1(x), g_2(x)\}$.

Solution

We prove the convexity from its definition:

$$\begin{aligned} h_1(\lambda x + (1-\lambda)y) &= g_1(A(\lambda x + (1-\lambda)y) + b), \\ &= g_1(\lambda(Ax+b) + (1-\lambda)(Ay+b)), \\ &\leq \lambda g_1(Ax+b) + (1-\lambda)g_1(Ay+b), \\ &= \lambda h_1(x) + (1-\lambda)h_1(y). \end{aligned}$$

$$\begin{aligned} h_2(\lambda x + (1-\lambda)y) &= C_1 g_1(\lambda x + (1-\lambda)y) + C_2 g_2(\lambda x + (1-\lambda)y), \\ &\leq C_1(\lambda g_1(x) + (1-\lambda)g_1(y)) + C_2(\lambda g_2(x) + (1-\lambda)g_2(y)), \\ &= \lambda(C_1 g_1(x) + C_2 g_2(x)) + (1-\lambda)(C_1 g_1(y) + C_2 g_2(y)), \\ &= \lambda h_2(x) + (1-\lambda)h_2(y). \end{aligned}$$

$$\begin{aligned} h_3(\lambda x + (1-\lambda)y) &= \max\{g_1(\lambda x + (1-\lambda)y), g_2(\lambda x + (1-\lambda)y)\}, \\ &\leq \max\{\lambda g_1(x) + (1-\lambda)g_1(y), \lambda g_2(x) + (1-\lambda)g_2(y)\}, \\ &\leq \max\{\lambda h_3(x) + (1-\lambda)h_3(y), \lambda h_3(x) + (1-\lambda)h_3(y)\}, \\ &= \lambda h_3(x) + (1-\lambda)h_3(y). \end{aligned}$$

Exercise T11.3 *Gradient descent*

You are given the following function:

$$f(x,y) = 2x^2 + 0.5y^2 - 3x - y - 2xy + 5$$

With starting point $z^{(1)} = (0,0)$, conduct two steps of the gradient descent algorithm. Choose the step size α using line search.

Solution

We determine the gradient of f.

$$\nabla f(x,y) = \begin{pmatrix} 4x - 3 - 2y \\ y - 1 - 2x \end{pmatrix}$$

Step 1:

We start at $z^{(1)} = (0,0)$ with $f(z^{(1)}) = 5$ and $\nabla f(z^{(1)}) = \begin{pmatrix} -3 \\ -1 \end{pmatrix}$.

Using line search, we select α to be $\alpha = \arg\min f(z^{(1)} - \alpha \nabla f(z^{(1)})) = \arg\min f(3\alpha, \alpha) = \arg\min 12.5\alpha^2 - 10\alpha + 5$. Using the first-order condition $25\alpha - 10 = 0$, we obtain $\alpha^* = 0.4$ (which is a minimum by the second-order condition).

We conduct the first step of gradient descent:

$$z^{(2)} = z^{(1)} - \alpha^* \nabla f(z^{(1)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0.4 \begin{pmatrix} -3 \\ -1 \end{pmatrix} = \begin{pmatrix} 1.2 \\ 0.4 \end{pmatrix}$$

Step 2:

Now $f(z^{(2)}) = 3$ and $\nabla f(z^{(2)}) = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$.

Using line search, we select α to be $\alpha = \arg\min f(z^{(2)} - \alpha \nabla f(z^{(2)})) = \arg\min f(1.2 - \alpha, 0.4 + 3\alpha) = \arg\min 12.5\alpha^2 - 10\alpha + 3$. Using the first-order condition $25\alpha - 10 = 0$, we obtain $\alpha^* = 0.4$ (which is a minimum by the second-order condition).

We conduct the second step of gradient descent:

$$z^{(3)} = z^{(2)} - \alpha^* \nabla f(z^{(2)}) = \begin{pmatrix} 1.2 \\ 0.4 \end{pmatrix} - 0.4 \begin{pmatrix} 1 \\ -3 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 1.6 \end{pmatrix}$$