Optimization

Fall semester 2022/23

Kristóf Bérczi,
Lydia Mendoza

Eötvös Loránd University
Institute of Mathematics
Department of Operations Research

Outline

Set 5

Question 1

Let $F : K \to \mathbb{R}$ be a convex, differentiable function. Prove that $D_F(x, y) \geqslant 0$.

Recall. $D_F(x, y) = F(x) - F(y) - \langle \nabla F(y), x - y \rangle$.

Since F is convex, we know by the first order condition that for $x, y \in K$

$$F(x) \geqslant F(y) + \langle \nabla F(y), x - y \rangle$$

$$\Rightarrow F(x) - F(y) - \langle \nabla F(y), x - y \rangle \geqslant 0$$

Question 2

Define a function F for which $D_F(x, y) = ||x - y||_2^2$. (1pt)

Observe that we need a function F so that
$D_F(x, y) = ||x - y||_2^2 = \sum_i (x_i - y_i)^2 = \sum_i x_i^2 + y_i^2 - 2x_i y_i$

Let $F(x) = ||x||_2^2 = \sum_i x_i^2$.

$\Rightarrow \nabla F(x) = (2x_1, \ldots, 2x_n) = 2x$.

Then,

$$D_F(x, y) = F(x) - F(y) - \langle \nabla F(y), x - y \rangle \tag{1}$$

$$= \sum_i x_i^2 - \sum_i y_i^2 - \langle 2y, x - y \rangle \tag{2}$$

$$= \sum_i x_i^2 - \sum_i y_i^2 - \langle 2y, x \rangle + \langle 2y, y \rangle \tag{3}$$

$$= \sum_i x_i^2 + y_i^2 - 2x_i y_i, \tag{4}$$

as required

Question 3

Let $F : K \to \mathbb{R}$ be a convex, differentiable function, and let $x, y, z \in K$. Prove that $\langle \nabla F(y) - \nabla F(z), y - x \rangle = D_F(x, y) + D_F(y, z) - D_F(x, z)$

$$
\begin{aligned}
& D_F(x, y) + D_F(y, z) - D_F(x, z) && (5)\\
& = F(x) - F(y) - \langle \nabla F(y), x - y \rangle + F(y) - F(z) - \langle \nabla F(z), y - z \rangle + && (6)\\
& \quad - (F(x) - F(z) - \langle \nabla F(z), x - z \rangle) && (7)\\
& = -\langle \nabla F(y), x - y \rangle - \langle \nabla F(z), y - z \rangle - \langle \nabla F(z), -x + z \rangle && (8)\\
& = \langle \nabla F(y), y - x \rangle - \langle \nabla F(z), y \rangle + \langle \nabla F(z), z \rangle - \langle \nabla F(z), -x \rangle - \langle \nabla F(z), z \rangle && (9)\\
& = \langle \nabla F(y), y - x \rangle - \langle \nabla F(z), y - x \rangle && (10)\\
& = \langle \nabla F(y) - \nabla F(z), y - x \rangle && (11)
\end{aligned}
$$

Question 4

Let $f(x) = x^2 - a$. Show that Newton's method leads to the recurrence $x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$

Since $f'(x) = 2x$, we obtain

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{12}$$

$$= x_n - \frac{x_n^2 - a}{2x_n} \tag{13}$$

$$= x_n - \frac{x_n}{2} + \frac{a}{2x_n} \tag{14}$$

$$= \frac{1}{2}\left(x_n + \frac{a}{x_n}\right) \tag{15}$$

Question 5

Let $f : \mathbb{R}^n \to \mathbb{R}$ a function and define $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ as $\tilde{f}(x) = f(Ax + b)$ where $A \in \mathbb{R}^{n \times n}$ is an invertible matrix and $b \in \mathbb{R}^n$. Verify that if x_0 moves to x_1 by applying one step of Newton's method with respect to $\tilde{f}$, then $y_0 = Ax_0 + b$ moves to $y_1 = Ax_1 + b$ by applying one step of Newton's method with respect to f. (2pts)

Recall. $\nabla \tilde{f}(x) = A^T \nabla f(Ax + b) \Rightarrow \nabla^2 \tilde{f}(x) = A^T \nabla^2 f(Ax + b) A$

From x_0, we apply one step of Newton's method with respect to $\tilde{f}$:

$$x_1 = x_0 - \left(\nabla^2 \tilde{f}(x_0)\right)^{-1} \cdot \nabla \tilde{f}(x_0) \tag{16}$$

$$= x_0 - A^{-1} \left(\nabla^2 f(Ax_0 + b)\right)^{-1} A^{T^{-1}} \cdot A^T \nabla f(Ax_0 + b) \tag{17}$$

$$= x_0 - A^{-1} \left(\nabla^2 f(Ax_0 + b)\right)^{-1} \cdot \nabla f(Ax_0 + b) \tag{18}$$

$$\Rightarrow Ax_1 + b = Ax_0 + b - \left(\nabla^2 f(Ax_0 + b)\right)^{-1} \cdot \nabla f(Ax_0 + b) \tag{19}$$

From y_1, we apply one step of Newton's method with respect to f:

$$y_1 = y_0 - \left(\nabla^2 f(y_0)\right)^{-1} \cdot \nabla f(y_0) \tag{20}$$

$$= Ax_0 + b - \left(\nabla^2 f(Ax_0 + b)\right)^{-1} \cdot \nabla f(Ax_0 + b) \tag{21}$$

Question 6

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Prove that the function $x \mapsto D_f(x, y)$ for a fixed $y \in \mathbb{R}^n$ is strictly convex. (2pts)

Let $y \in \mathbb{R}^n$ be fixed, $x, z \in \mathbb{R}^n$ and $\theta \in [0, 1]$.

We want to show that

$$
\begin{aligned}
&D_f(\theta x + (1-\theta)z,\ y) \\
&\quad < \theta D_f(x, y) + (1-\theta) D_f(z, y) \\
&\quad = \theta \left[f(x) - f(y) - \langle \nabla f(y), x - y \rangle \right] + (1-\theta) \left[f(z) - f(y) - \langle \nabla f(y), z - y \rangle \right] \\
&\quad = \theta \left[f(x) - \langle \nabla f(y), x \rangle \right] + (1-\theta) \left[f(z) - \langle \nabla f(y), z \rangle \right] - f(y) - \langle \nabla f(y), -y \rangle
\end{aligned}
$$

By definition of D_f and strongly convexity of f,

$$
\begin{aligned}
&D_f(\theta x + (1-\theta)z,\ y) \\
&= f(\theta x + (1-\theta)z) - f(y) - \langle \nabla f(y), \theta x + (1-\theta)z - y \rangle \\
&< \theta f(x) + (1-\theta) f(z) - f(y) - \langle \nabla f(y), \theta x + (1-\theta)z - y \rangle \\
&= \theta f(x) + (1-\theta) f(z) - f(y) - \theta \langle \nabla f(y), x \rangle - (1-\theta) \langle \nabla f(y), z \rangle - \langle \nabla f(y), -y \rangle,
\end{aligned}
$$

as desired

Question 6, second option

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Prove that the function $x \mapsto D_f(x, y)$ for a fixed $y \in \mathbb{R}^n$ is strictly convex. (2pts)

Suppose that the function is twice differentiable.

Then,

$$\begin{aligned} D_f(x, y) &= f(x) - f(y) - \langle \nabla f(y), x - y \rangle \\ \nabla_x D_f(x,\, y) &= \nabla f(x) - \nabla f(y) \\ \nabla^2_{x,x} D_f(x,\, y) &= \nabla^2 f(x) \succ 0 \end{aligned}$$

Question 7

Prove that for all $p \in \Delta_n$, $D_{KL}(p, p^1) \leqslant \log n$. Here p^1 is the uniform probability distribution with $p_i^1 = \frac{1}{n}$ for $i = 1, \ldots, n$. (2pts)

Recall. $D_{KL}(p, q) = -\sum_{i=1}^n p_i \log \frac{q_i}{p_i}$.

$$D_{KL}(p, p^1) = -\sum_{i=1}^{n} p_i \log \frac{p_i^1}{p_i} \tag{22}$$

$$= -\sum_{i=1}^{n} p_i \log \frac{1}{np_i} \tag{23}$$

$$= \sum_{i=1}^{n} p_i \log np_i \tag{24}$$

$$\leqslant \sum_{i=1}^{n} p_i \log n \tag{25}$$

$$= \log n \cdot \sum_{i=1}^{n} p_i \tag{26}$$

$$= \log n \tag{27}$$