Continuous Optimization: Assignment 12

Due on July 16, 2024

Honglu Ma **Hiroyasu Akada** **Mathivathana Ayyappan**

Exercise 1

1

(i) Not convex. Consider $p = 3$ we have $f(x) = x^3$ and $f''(x) = 6x \leq 0$ when $x \leq 0$.

(ii) Convex. $f''(x) = x^{-2} \geq 0$

(iii) Convex. $f''(x) = \alpha^2 e^{\alpha x} \geq 0$

(iv) Convex. $f''(x) = \frac{1}{(1-x)} \geq 0$ when $x \in (0, 1)$

2

(i) Convex.
Take $x, y \in C$.
Consider $||(1-\lambda)x + \lambda y||_2 \overset{Cauchy-Schwarz}{\leq} ||(1-\lambda)x||_2 + ||\lambda y||_2 = (1-\lambda)||x||_2 + \lambda||y||_2 \leq 1$.

(ii) Not convex. Consider $x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\lambda = \frac{1}{2}$.
Clearly $||(1-\lambda)x + \lambda y||_2 = \frac{\sqrt{2}}{2} \neq 1$.

(iii) Not convex. Consider $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $y = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ and $\lambda = \frac{1}{2}$.
Clearly $||(1-\lambda)x + \lambda y||_\infty = || \begin{pmatrix} 0 \\ 0 \end{pmatrix} ||_\infty = 0 \neq 1$.

(iv) Convex. Similar to (i).

3

See appendix

4

$$T_C(\begin{pmatrix} 1 & 1 \end{pmatrix}^\top) = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 \leq 1, x_2 \leq 1 \right\}$$
$$N_C(\begin{pmatrix} 1 & 1 \end{pmatrix}^\top) = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 \geq 1, x_2 \geq 1 \right\}$$

Exercise 2

(A)

In steepest descent method, we have $x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)})$. We can approximate $f(x^{(k+1)})$ with its second order Taylor expansion around $x^{(k)}$.

$$\begin{aligned} f(x^{(k+1)}) &\approx f(x^{(k)}) + \nabla f(x^{(k)})^\top (x^{(k+1)} - x^{(k)}) + \frac{1}{2}(x^{(k+1)} - x^{(k)})^\top \nabla^2 f(x^{(k)})(x^{(k+1)} - x^{(k)}) \\ &= f(x^{(k)}) - \alpha \nabla f(x^{(k)})^\top \nabla f(x^{(k)}) + \frac{\alpha^2}{2} \nabla f(x^{(k)})^\top \nabla^2 f(x^{(k)}) \nabla f(x^{(k)}) =: g(\alpha) \end{aligned}$$

We can find the optimal α by solving $\frac{dg}{d\alpha} = 0$.

$$\begin{aligned}
\frac{dg}{d\alpha} &= -\nabla f(x^{(k)})^\top \nabla f(x^{(k)}) + \alpha \nabla f(x^{(k)})^\top \nabla^2 f(x^{(k)}) \nabla f(x^{(k)}) = 0 \\
\nabla f(x^{(k)})^\top \nabla f(x^{(k)}) &= \nabla f(x^{(k)})^\top (\alpha \nabla^2 f(x^{(k)})) \nabla f(x^{(k)})
\end{aligned}$$

By choosing $\alpha = (\nabla^2 f(x^{(k)}))^{-1}$, we can cancel the affect of rescaling of the Hessian matrix. Thus we arrive at the Newton's method: $x^{(k+1)} = x^{(k)} - (\nabla^2 f(x^{(k)}))^{-1} \nabla f(x^{(k)})$.

(B)

Step 0 Let $x^{(0)} := \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\begin{aligned}
d^{(0)} = r^{(0)} &= -\nabla f(x^{(0)}) \\
&= -b - Qx^{(0)} \\
&= -\begin{pmatrix} 1 \\ -1 \end{pmatrix} - \begin{pmatrix} 4 & 3 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\
\tau_0 &= \frac{\langle r^{(0)}, r^{(0)} \rangle}{\langle d^{(0)}, Qd^{(0)} \rangle} \\
&= \frac{1}{2}
\end{aligned}$$

Step 1

$$\begin{aligned}
x^{(1)} &= x^{(0)} + \tau_0 d^{(0)} \\
&= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \\
r^{(1)} &= r^{(0)} + \tau_0 Q d^{(0)} \\
&= \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 4 & 3 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{3}{2} \\ -\frac{5}{2} \end{pmatrix} \\
\beta_1 &= \frac{\langle r^{(1)}, r^{(1)} \rangle}{\langle r^{(0)}, r^{(0)} \rangle} \\
&= \frac{17}{4} \\
d^{(1)} &= -r^{(1)} + \beta_1 d^{(0)} \\
&= \begin{pmatrix} \frac{11}{4} \\ -\frac{7}{4} \end{pmatrix}
\end{aligned}$$

Exercise 3

See Appendix A: Handwritten Solution for Exercise 3

Exercise 4

(A)

The optimality condition is

$$\begin{aligned} \nabla f(x^\star) + \lambda^\star \nabla g(x^\star) + \langle \mu^\star, \nabla h(x^\star) \rangle &= 0 \\ \Rightarrow -(\alpha + x^\star)^{-1} + \lambda^\star x^\star - \mu^\star &= 0 \\ x^\star &\geq 0 \\ \mu^\star &\geq 0 \\ \langle \mu^\star, x^\star \rangle &= 0 \end{aligned}$$

where $f(x) = \sum_{i=1}^n -\log(\alpha + x_i)$, $g(x) = \sum_{i=1}^n x_i - 1$ and $h(x) = -x$. Note $(v^{-1})_i = \frac{1}{v_i}$ for $v \in \mathbb{R}^n$.

(B)

In each time step k, we first compute a point $\tilde{x}^{(k)} = x^{(k)} - \alpha \nabla f(x^{(k)})$. Then we project $\tilde{x}^{(k)}$ onto the feasible set $C = \{x \in \mathbb{R}^n \mid \alpha^\top x = \beta\}$. We know that C is a hyperplane and the projection has closed form:

$$\hat{x}^{(k)} = \text{proj}_C(\tilde{x}^{(k)}) = \tilde{x}^{(k)} - \frac{\alpha^\top \tilde{x}^{(k)} - \beta}{\|\alpha\|_2^2} \alpha$$

Then we can update compute the next time step by

$$x^{(k+1)} = x^{(k)} + \tau_k (\hat{x}^{(k)} - x^{(k)})$$

with τ_k that satisfies the Armijo condition.

Appendix A: Handwritten Solution for Exercise 3

$f(x_1, x_2, x_3) = x_1 \sin(x_2) + x_3 e^{x_1 x_2}$

$x_1 = x_1$

$x_2 = x_2$

$x_3 = x_3$

$x_4 = \sin(x_2)$

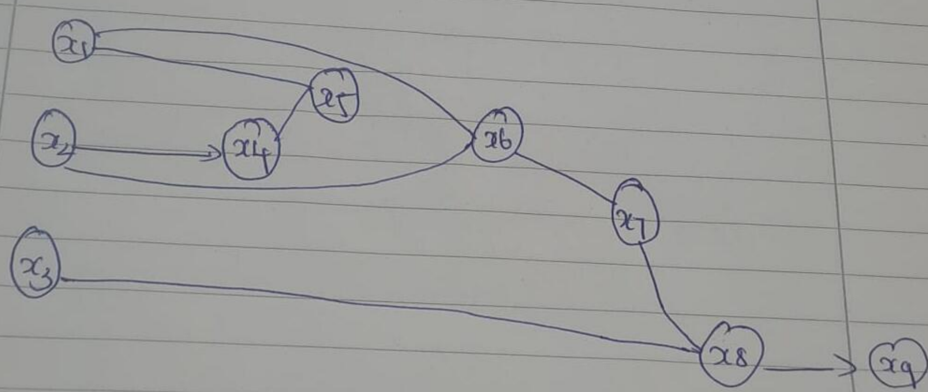
$x_5 = x_1 x_4$

$x_6 = x_1 x_2$

$x_7 = e^{x_6}$

$x_8 = x_3 x_7$

$x_9 = x_8 + x_5$



derivates

$\frac{\delta x_1}{\delta x_1} = 1$

$\frac{\delta x_2}{\delta x_2} = 1$

$\frac{\delta x_3}{\delta x_3} = 1$

$\frac{\delta x_4}{\delta x_2} = \cos(x_2)$

$\frac{\delta x_5}{\delta x_1} = x_4$

$\frac{\delta x_5}{\delta x_4} = x_1$

$\frac{\delta x_6}{\delta x_1} = x_2$

$\frac{\delta x_6}{\delta x_2} = x_1$

$\frac{\delta x_7}{\delta x_6} = e^{x_6}$

$\frac{\delta x_8}{\delta x_7} = x_3$

$\frac{\delta x_8}{\delta x_3} = x_7$

$\frac{\delta x_9}{\delta x_8} = 1$

$\frac{\delta x_9}{\delta x_5} = 1.$

Forum mode

$x_1 = \tilde{x}_1$

$x_2 = \tilde{x}_2$

$x_3 = \tilde{x}_3$

$x_4 = \sin(\tilde{x}_2)$

$x_5 = \tilde{x}_1 \sin(\tilde{x}_2)$

$x_6 = \tilde{x}_1 \tilde{x}_2$

$x_7 \Rightarrow e^{\tilde{x}_1 \tilde{x}_2}$

$x_8 \Rightarrow \tilde{x}_3 e^{\tilde{x}_1 \tilde{x}_2}$

$x_9 \Rightarrow \tilde{x}_1 \sin(\tilde{x}_2) + \tilde{x}_3 e^{\tilde{x}_1 \tilde{x}_2}$

$\dot{x}_1 \Rightarrow 1$

$\dot{x}_2 = 1$

$\dot{x}_3 = 1$

$\dot{x}_4 = \cos(x_2)(\dot{x}_2)$

$\dot{x}_5 \Rightarrow \dot{x}_1 \sin(x_2) + \sin(\dot{x}_2) x_1$

$\dot{x}_6 \Rightarrow \dot{x}_1 x_2 + x_2 \dot{x}_1$

$\dot{x}_7 \Rightarrow e^{(x_1 x_2)} (\dot{x}_1)(\dot{x}_2)$

$\dot{x}_8 = \dot{x}_3 e^{x_1 x_2} + (e^{x_1 x_2})(\dot{x}_1 \dot{x}_2)$

x_3

$x_9 \Rightarrow \dot{x}_8 + \dot{x}_5$

$\Rightarrow$

$$\tilde{x}_9 = \frac{\delta 8}{x_9} = 1.$$

$$\tilde{x}_8 \Rightarrow \frac{\delta 8}{x_9} \cdot \frac{\delta 9}{\delta 8} \Rightarrow 1 \cdot 1 = 1$$

$$\tilde{x}_7 \Rightarrow \frac{\delta 9}{\delta 8} \cdot \frac{\delta 8}{\delta 7} = 1 \cdot \tilde{x}_3 = x_3$$

$$\tilde{x}_6 \Rightarrow \frac{\delta 8}{\delta 7} \cdot \frac{\delta 7}{\delta 6} \Rightarrow \tilde{x}_3 e^{\tilde{x}_6} \Rightarrow \tilde{x}_3 e^{\tilde{x}_1 \tilde{x}_2}$$

$$\tilde{x}_1 \Rightarrow \frac{\delta 7}{\delta 6} \cdot \frac{\delta 6}{\delta x_1} = \tilde{x}_3 e^{\tilde{x}_6} \tilde{x}_2 \Rightarrow \tilde{x}_3 x_2 e^{x_6} \Rightarrow x_3 x_2 e^{x_1 x_2}$$