ESE415 OPTIMIZATION

Tutorial 4

Xiaojian Xu

02/22/2019

Guideline

HW3 RECITATION

HW2 RECAP

LECTURE RECAP

HW3 Recitation(P2)

Consider the problem of minimizing $f(x)$ and $g(x)$ by using the Newton's method with $\gamma = 1$.

a) $f(x) = \log(e^x + e^{-x})$, where $\log$ is the natural logarithm.

(i) What is the minimizer x^* of f? Is it unique?

(ii) Does the Newton's method converge to x^* for $x^0 = 1$?

(iii) Does the Newton's method converge to x^* for $x^0 = 1.1$?

(iv) Find an interval $I \subseteq \mathbb{R}$ such that for any $x^0 \in I$, the Newton's method converges.

b) $g(x) = \sqrt[3]{x^4}$.

(i) What is the minimizer x^* of g? Is it unique?

(ii) Does the Newton's method converge to x^* for $x^0 = 1$?

(iii) Does the Newton's method converge to x^* for $x^0 = 0.1$?

(iv) Find an interval $I \subseteq \mathbb{R}$ such that for any $x^0 \in I$, the Newton's method converges.

Solution (P2)

a) We compute the first two derivatives of the function

$$f(x) = \log(\mathrm{e}^{x} + \mathrm{e}^{-x}) \quad \Rightarrow \quad f'(x) = \frac{\mathrm{e}^{2x} - 1}{\mathrm{e}^{2x} + 1} \quad \Rightarrow \quad f''(x) = \frac{4\mathrm{e}^{2x}}{(\mathrm{e}^{2x}+1)^2} > 0.$$

The function is strictly convex. By solving $f'(x) = 0$, we get the unique solution $x^* = 0$.

The Newton's step for $\gamma = 1$ is given by

$$x^+ = x - [f''(x)]^{-1}(f'(x)) = x - \frac{1}{4}\left(\mathrm{e}^{2x} - \mathrm{e}^{-2x}\right) = x - \sinh(x)\cosh(x).$$

Consider

$$\delta(x) \triangleq \frac{|x^+ - x^*|}{|x - x^*|} = \left|\frac{x^+}{x}\right| = \left|1 - \frac{\sinh(x)\cosh(x)}{x}\right|.$$

Note that for any $\delta(x) \in (0,1)$, the method will bring the next iterate closer to the solution. On the other hand, for any $\delta(x) > 1$, the algorithm increases the distance of the iterate away from x^*. We have $\delta(1) \approx 0.8134$ and $\delta(1.1) \approx 1.0260$. Hence, in the first case the method converges and in the second it diverges. In particular, for any $|x| < 1.0886$, we have $\delta(x) < 1$.

b) Observe that $g(0) = 0$ and $g(x) > 0$ for $x \neq 0$. Hence $x^* = 0$ is the unique global minimizer. We compute the first two derivatives

$$g(x) = x^{4/3} \quad \Rightarrow \quad g'(x) = \frac{4}{3}x^{1/3} \quad \Rightarrow \quad g''(x) = \frac{4}{9}x^{-2/3}.$$

The Newton's step for $\gamma = 1$ is given by

$$x^+ = x - [f''(x)]^{-1}(f'(x)) = x - \left[\frac{9}{4}x^{2/3}\right] \cdot \frac{4}{3}x^{1/3} = -2x.$$

We consider

$$\delta(x) = \left|\frac{x^+}{x}\right| = 2,$$

which means that

$$|x^t| = 2^t|x^0|.$$

Hence, no matter the initial value $x^0 \neq 0$, the Newton's sequence diverges.

HW3 Recitation(P2)

Problem 2

Consider functions $f \in \mathcal{F}^1(\mathbb{R}^n)$ and $g(\boldsymbol{x}) = (L/2)\|\boldsymbol{x}\|^2 - f(\boldsymbol{x})$, as well as the following inequality

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^\mathsf{T}(\boldsymbol{x} - \boldsymbol{y}) \geq \frac{1}{L}\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|^2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n. \tag{$\star$}$$

Show the following

(a) ∇f is Lipschitz continuous $\quad \Rightarrow \quad g$ is convex

(b) g is convex $\quad \Rightarrow \quad (\star)$

(c) $(\star) \quad \Rightarrow \quad \nabla f$ is Lipschitz continuous

(d) What can we conclude from (a)-(c) above?

Hints (P2)

(a) Use the relationship between f(x) and g(x) to substitute f(x) as g(x)

(b) Define two two convex functions and reuse the inequality in Problem 4 of Assignment #1

Problem 4

Consider $f \in \mathcal{C}_L^1(\mathbb{R}^n)$ and its global minimizer $\boldsymbol{x}^* \in \mathbb{R}^n$. Show that

$$\frac{1}{2L}\|\nabla f(\boldsymbol{x})\|^2 \leq (f(\boldsymbol{x}) - f(\boldsymbol{x}^*)) \leq \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

How does this result relate to the *first-order necessary condition* in Lecture 3, Theorem 3?

(c) Cauchy-Schwartz

(b) Define two two convex functions and reuse the inequality in Problem 4 of Assignment #1

Hints (P2.(b))

(b) we define two convex functions for $z \in \mathbb{R}^n$.

$$\boxed{\begin{cases} \varphi_x(z) = f(z) - \nabla f(x)^T z \\ \varphi_y(z) = f(z) - \nabla f(y)^T z \end{cases}}$$

Since g is convex, we know that the corresponding "g" function for $\varphi_x(z)$ and $\varphi_y(z)$ are convex:

$$\boxed{\begin{cases} r_x(z) = \frac{L}{2}\|z\|^2 - \varphi_x(z) \\ r_y(z) = \frac{L}{2}\|z\|^2 - \varphi_y(z) \end{cases}}, \quad \text{note that} \Rightarrow \boxed{\begin{matrix} \varphi_x(z) = 0 \Rightarrow z_x^* = x \\ \varphi_y(z) = 0 \Rightarrow z_y^* = y \end{matrix}}$$

And in (a) we proved that:

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\| \Leftrightarrow g(y) \ge g(x) + \nabla g(x)^T (y-x).$$

which implies that the convexity of "g" function leads to the quadratic upper bound of the "f" function.

Therefore now we are able to use the conclusion in Assignment 1 #4:

$$\frac{1}{2L}\|\nabla f(x)\|^2 \le f(x) - f(x^*) \le \frac{L}{2}\|x - x^*\|$$

and get:

$$\Rightarrow \quad \varphi_x(y) - \varphi_x(x) = f(y) - f(x) - \nabla f(x)^T (y-x)$$

$$\ge \frac{1}{2L}\|\nabla \varphi_x(y)\|^2 = \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2 \quad ①$$

$$\varphi_y(x) - \varphi_y(y) = f(x) - f(y) - \nabla f(y)^T (x-y)$$

$$\ge \frac{1}{2L}\|\nabla \varphi_y(x)\|^2 = \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \quad ②$$

Combine ① and ② will lead us to the conclusion.

Lecture Recap

Properties of convex functions

1. The linear combination of convex functions are also convex functions.

   Theorem 6. If f_1 and f_2 are $\mathcal{F}^1(C)$ and $a, b \geq 0$, then $f = af_1 + bf_2 \in \mathcal{F}^1(C)$.

2. The composite function of a convex function and a increasing convex function are still convex function.

   Theorem 7. If $f(\boldsymbol{x})$ is a convex (respectively, strictly convex) function on a convex set $C \subseteq \mathbb{R}^n$ and if $\varphi(t)$ is an increasing (respectively, strictly increasing) convex function defined on the range of f in $\mathbb{R}$, then the composite function $g(\boldsymbol{x}) = \varphi(f(\boldsymbol{x}))$ is convex (respectively, strictly convex) on C.

3. The gradient of the convex function is monotonically increasing:

$$\mathrm{f}(x) \geq f(\mathrm{y}) + \nabla f(\mathrm{y})^T(x - y)$$
$$\mathrm{f}(y) \geq f(\mathrm{x}) + \nabla f(x)^T(y - x)$$

   Therefore we get $(\nabla f(y) \ - \nabla f x)^T (y - x) \geq 0$;

HW2 Recap(P4)

Problem 4

(a) Comment on the convexity of

$$f(\boldsymbol{x}) = \boldsymbol{x}^\mathsf{T} \boldsymbol{A}\boldsymbol{x} - \mathsf{log}(x_1 \cdot x_2 \cdots x_n), \quad \boldsymbol{A} \succ 0,$$

over the set $C = \{\boldsymbol{x} \in \mathbb{R}^n : x_1 > 0, \ldots, x_n > 0\}$. The function $\mathsf{log}(\cdot)$ is the natural logarithm.

(b) Comment on the convexity of

$$g(\boldsymbol{x}) = \beta_1 \exp(\boldsymbol{w}_1^\mathsf{T}\boldsymbol{x}) + \beta_2 \exp(\boldsymbol{w}_2^\mathsf{T}\boldsymbol{x}), \quad \beta_1, \beta_2 \geq 0$$

over $\mathbb{R}^n$.

(c) Consider a convex set $C \subseteq \mathbb{R}^n$ and the function

$$h(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} \{\|\boldsymbol{x} - \boldsymbol{y}\|\},$$

which measures the distance to the *farthest point* in C. Comment on the convexity of h over $\mathbb{R}^n$.

Solution (P4)

(c) Consider $\theta \in [0,1]$ and $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$, we then have

$$\begin{aligned} h(\theta\boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2) &= \sup_{\boldsymbol{y}\in C} \{\|\theta\boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 - \boldsymbol{y}\|\} \\ &\le \sup_{\boldsymbol{y}\in C} \{\theta\|\boldsymbol{x}_1 - \boldsymbol{y}\| + (1-\theta)\|\boldsymbol{x}_2 - \boldsymbol{y}\|\} \\ &\le \theta \sup_{\boldsymbol{y}\in C} \{\|\boldsymbol{x}_1 - \boldsymbol{y}\|\} + (1-\theta) \sup_{\boldsymbol{y}\in C} \{\|\boldsymbol{x}_2 - \boldsymbol{y}\|\} \\ &= \theta h(\boldsymbol{x}_1) + (1-\theta) h(\boldsymbol{x}_2), \end{aligned}$$

where we used the fact that the norm is convex and that supremum gets only bigger if it is taken separately. As a side note, this relates to the *pointwise maximum property* of convex functions.

Thanks