OP202 HW2

Yanyu ZHOU

February 8, 2023

1 Convec function

Theorem 1 (19)**.** *Function $f : X \subset \mathbf{R}^n \to \mathbf{R}$ is convex and belongs to $\mathcal{C}^1(X)$ iff $\forall x, y \in X$, $f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$.*

Proof: We denote $x_\lambda = \lambda x + (1-\lambda)y$. For one direction, if f is convex and $f \in \mathcal{C}^1$, for some $\lambda \in [0, 1)$, we have

$$
\begin{aligned}
f(x_\lambda) &\leq \lambda f(x) + (1-\lambda) f(y) \\
f(x) &\geq \frac{f(x_\lambda)}{\lambda} - \frac{1-\lambda}{\lambda} f(y) = \frac{f(\lambda x + (1-\lambda)y) - f(y)}{\lambda} + f(y).
\end{aligned}
$$

Takeing the limit of λ to 0, by the definition of derivative, we obtain:

$$
f(x) \geq f(y) + \lim_{\lambda \to 0} \frac{f(\lambda x + (1-\lambda)y) - f(y)}{\lambda} = f(y) + \langle \nabla f(y), x - y \rangle
$$

For the other direction, if $\forall x, y \in X,\ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$, subtituding y with x_λ, we have

$$
\begin{aligned}
f(x) &\geq f(x_\lambda) + \langle \nabla f(x_\lambda), x - x_\lambda \rangle \\
f(y) &\geq f(x_\lambda) + \langle \nabla f(x_\lambda), y - x_\lambda \rangle
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
f(x_\lambda) &\leq f(x) + \langle \nabla f(x_\lambda), x_\lambda - x \rangle = f(x) + (\lambda - 1) \langle \nabla f(x_\lambda), x - y \rangle \\
f(x_\lambda) &\leq f(y) + \langle \nabla f(x_\lambda), x_\lambda - y \rangle = f(y) + \lambda \langle \nabla f(x_\lambda), x - y \rangle
\end{aligned}
$$

Mutilplying the first inquality by λ and the second one by $1 - \lambda$, we obtain

$$
\begin{aligned}
\lambda f(x_\lambda) &\leq \lambda f(x) + \lambda(\lambda - 1) \langle \nabla f(x_\lambda), x - y \rangle \\
(1-\lambda) f(x_\lambda) &\leq (1-\lambda) f(y) + \lambda(1-\lambda) \langle \nabla f(x_\lambda), x - y \rangle
\end{aligned}
$$

Addind the two inqualities together, we get $f(x_\lambda) \leq \lambda f(x) + (1-\lambda) f(y)$.

Theorem 2 (20)**.** *Function $f : X \subset \mathbf{R}^n \to \mathbf{R}$ is convex and belongs to $\mathcal{C}^2(X)$ iff $\forall x, y \in X$, $\nabla^2 f(x) \succeq 0$.*

Proof: For one direction, if $f \in \mathcal{C}^2(X)$ is convex, by1, we have

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \quad f(y) \geq f(x) + \langle \nabla f(x), x - y \rangle.$$

Adding them together, we get

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0. \tag{1}$$

We denote $x_\tau = x + \tau s$, $\tau > 0$, $s \in \mathbf{R}^n$, then obtain

$$0 \leq \frac{1}{\tau} \langle \nabla f(x_\tau) - \nabla f(x), x_\tau - x \rangle = \frac{1}{\tau} \langle \nabla f(x_\tau) - \nabla f(x), s \rangle = \frac{1}{\tau} \int_0^\tau \langle \nabla^2 f(x + \lambda s) s, s \rangle d\lambda$$

Takeing the limit of τ to 0, by the definition of derivative, we obtain:

$$0 \leq \lim_{\tau \to 0} \frac{1}{\tau} \int_0^\tau \langle \nabla^2 f(x + \lambda s) s, s \rangle d\lambda = \nabla^2 f(x)$$

For the other direction, if $\forall x, y \in X$, $\nabla^2 f(x) \succeq 0$, we fix y, by Taylor's Theorem with integral remainder and the Fundamental Theorem of Calculus, then obtain:

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \int_0^\tau \langle \nabla^2 f(x + \lambda(y - x))(y - x), y - x \rangle \\ &\geq f(x) + \langle \nabla f(x), y - x \rangle. \end{aligned}$$

2 Strongly Convex function

Theorem 3 (21)**.** *Function $f : X \subset \mathbf{R}^n \to \mathbf{R}$ is m-strongly convex and belongs to $\mathcal{C}^1(X)$ iff $\forall x, y \in X$, $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq m||x - y||^2$.*

Proof: For one direction, let f be m-strongly convex, then by definition $g(x) = f(x) - \frac{m}{2}||x||^2$ is convex and $\nabla g(x) = \nabla f(x) - mx$. By monotinicty condition for convexity we have:

$$\begin{aligned} 0 &\leq \langle \nabla g(x) - \nabla g(y), x - y \rangle \\ 0 &\leq \langle \nabla f(x) - mx - (\nabla f(y) - my), x - y \rangle \\ 0 &\leq \langle \nabla f(x) - \nabla f(y), x - y \rangle - \langle mx - my, x - y \rangle \\ m||x - y||^2 &\leq \langle \nabla f(x) - \nabla f(y), x - y \rangle. \end{aligned}$$

For the another direction, suppose f satisfies that $\forall x, y \in X$, $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq m||x - y||^2$. Since $f, g \in \mathcal{C}^1(X)$, by 1 it's equivalent to prove that $g(x) \geq g(y) + \langle \nabla g(y), x - y \rangle$. Using Lemma 2.1.3 of Nesterov's book, we can prove it.

Theorem 4 (22)**.** *Function $f : X \subset \mathbf{R}^n \to \mathbf{R}$ is m-strongly convex and belongs to $\mathcal{C}^2(X)$ iff $\forall x, y \in X$, $\nabla^2 f(x) \succeq mI_n$.*

Proof: For one direction, let f be m-strongly convex, then by definition $g(x) = f(x) - \frac{m}{2}||x||^2$ is convex and $\nabla g(x) = \nabla f(x) - mx$. Since $f, g \in \mathcal{C}^2(X)$, by 2, $\nabla^2 g(x) \succeq 0$, i.e,. $\nabla^2 f(x) - mI_n \succeq 0$.

For the another direction, suppose$\nabla^2 g(x) \succeq 0$, by Lemma 2.1.4 of Nesterov's book, we can prove $g \in \mathcal{C}^2(X)$, so does f, and g is convex.

3 Smooth function

Theorem 5 (23,24)**.** *For a convex function $f : X \subset \mathbf{R}^n \to \mathbf{R}$, the following statements are equivalent:*

1. *f is L-smooth.*
2. *$\forall x, y \in X$, $||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$*
3. *Lower bound: $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2L}||\nabla f(x) - \nabla f(y)||^2$*
4. *Co-coercivity: $\forall x, y \in X$, $\frac{1}{L}||\nabla f(x) - \nabla f(y)||^2 \leq \langle \nabla f(x) - \nabla f(y), x - y\rangle$*

Proof:

Let f be L-smooth, then by definition $g(x) = \frac{L}{2}||x||^2 - f(x)$ is convex and $\nabla g(x) = Lx - \nabla f(x)$. By 1 we have:

$$\begin{aligned} g(x) &\geq g(y) + \langle \nabla g(y), x - y\rangle \\ \frac{L}{2}||x||^2 - f(x) &\geq \frac{L}{2}||y||^2 - f(y) + \langle \nabla(\frac{L}{2}||y||^2 - f(y)), x - y\rangle \\ &= \frac{L}{2}||y||^2 - f(y) + \langle Ly - \nabla f(y), x - y\rangle \\ &= Lxy - \frac{L}{2}||y||^2 - f(y) - \langle \nabla f(y), x - y\rangle \\ f(x) &\leq \frac{L}{2}(||x||^2 + ||y||^2 - 2Lxy) + f(y) + \langle \nabla f(y), x - y\rangle \\ &\leq \frac{L}{2}(||x - y||^2) + f(y) + \langle \nabla f(y), x - y\rangle \end{aligned} \tag{2}$$

$1 \Rightarrow 3$: We define: $h(x) = f(x) - \langle x, \nabla f(\xi)\rangle$ which is convex with a global minimum at ξ. Then:

$$h(\xi) \leq h(x - \frac{1}{L}\nabla h(x)) \leq h(x) + \langle \nabla h(x), -\frac{1}{L}\nabla h(x)\rangle + \frac{L}{2}||-\frac{1}{L}\nabla h(x)||^2 \leq h(x) - \frac{1}{2L}||\nabla h(x)||^2$$

$$\Downarrow \nabla h(x) = \nabla f(x) - \nabla f(\xi)$$

$$f(\xi) - \langle \xi, \nabla f(\xi)\rangle \leq f(x) - \langle x, \nabla f(\xi)\rangle - \frac{1}{2L}||\nabla f(x) - \nabla f(\xi)||^2$$

$3 \Rightarrow 4$: Applying 3 twice for (x, y) and (y, x), and sum to get

$$0 \geq \frac{1}{L}||\nabla f(x) - \nabla f(y)||^2 + \langle \nabla f(x) - \nabla f(y), y - x\rangle$$

$4 \Rightarrow 2$: By generalized Cauchy-Schwart, we have

$$||\nabla f(x) - \nabla f(y)||^2 \le L\langle \nabla f(x) - \nabla f(y), x - y\rangle \le L^2||x-y||^2,$$

thus,

$$||\nabla f(x) - \nabla f(y)|| \le L||x-y||.$$

Theorem 6 (22)**.** *A convex function $f : X \subset \mathbf{R}^n \to \mathbf{R}$ is L-smooth and belongs to $\mathcal{C}^2(X)$ iff $\forall x, y \in X$, $0 \preceq \nabla^2 f(x) \preceq LI_n$.*

Proof: Since f is convex, we have $0 \preceq \nabla^2 f(x)$. Since $f \in \mathcal{C}_L^{2,1}$, by Lemma 3 in first class, $||\nabla^2 f(x)|| \le L$, therefore, $\nabla^2 f(x) \preceq LI_n$.

4 Nesterov's gradient

Question: For Nesterov's accelerated gradient method, when $\alpha \le \frac{1}{L}$, do we have the following statements?

- the global convergence is of rate $O(\frac{1}{t^2})$. Yes, because this method provides an optimal convergence rate for smooth convex functions.
- better than in the non-convex case. Yes, because for non-convex cases, the convergence rate is $O(\frac{1}{\sqrt{t}})$.

Remark: The Homeowrk 2 is too much. I spent at least 10 hours on it.