MATH 6222 LECTURE NOTE 1: SMOOTH CONVEX FUNCTIONS

JINGRONG WEI

ABSTRACT. This lecture note mainly introduces the function class considered in smooth convex optimization and is prepared based on [1, Chapter 2].

CONTENTS

1. CONVEX FUNCTIONS

Definition 1.1 (Convex sets)**.** *A set $V \subseteq \mathbb{R}^n$ is called convex if for any $x, y \in V$ and α from $[0,1]$ we have $\alpha x + (1-\alpha)y \in V$.*

Thus, a convex set contains the whole segment $[x, y]$ provided that the end points x and y belong to the set.

Definition 1.2 (Convex function)**.** *A continuously differentiable function $f(\cdot)$ is called convex on a convex set V (notation $f \in \mathcal{S}^1(V)$) if for any $x, y \in V$ we have*

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

If $-f(\cdot)$ is convex, we call $f(\cdot)$ concave.

For convex functions, the following properties hold.

Lemma 1.3 (Closed under linear combination)**.** *If f_1 and f_2 are convex and $\alpha, \beta \geq 0$, then the function $f = \alpha f_1 + \beta f_2$ is convex.*

Lemma 1.4 (Closed under linear transformation)**.** *If $f \in \mathcal{S}^1(V)$, then for $A : \mathbb{R}^m \to \mathbb{R}^n$ and $b \in \mathbb{R}^n$ such that $Ax + b \in V$, $\phi(x) = f(Ax + b)$ is convex.*

Theorem 1.5 (Definition of convex function without differentiability)**.** *A continuously differentiable function f belongs to the class $\mathcal{S}^1(V)$ if and only if for any $x, y \in V$ and $\alpha \in [0,1]$ we have*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y).$$

Proof. ($\Rightarrow$) This direction is straightforward using the definition of convex function at $x_\alpha = \alpha x + (1-\alpha)y$.

Date: January 22, 2025.

($\Leftarrow$) Choose some $\alpha \in [0,1)$. Then

$$\begin{aligned} f(y) &\geq \frac{1}{1-\alpha}\left[f\left(x_\alpha\right)-\alpha f(x)\right]=f(x)+\frac{1}{1-\alpha}\left[f\left(x_\alpha\right)-f(x)\right] \\ &=f(x)+\frac{1}{1-\alpha}[f(x+(1-\alpha)(y-x))-f(x)]. \end{aligned}$$

Let α tend to 1 , we get the desired result. $\square$

The *epigraph* of a function $f: V \rightarrow [-\infty,+\infty]$ is defined as

$$\operatorname{epi} f := \{(x,w) | x \in V, w \in \mathbb{R}, f(x) \leq w\}.$$

Proposition 1.6. *A function $f: V \rightarrow \mathbb{R}$ is convex if and only if $\operatorname{epi} f \subset \mathbb{R}^{n+1}$ is convex.*

Theorem 1.7. *A continuously differentiable function f belongs to the class $\mathcal{S}^1(V)$ if and only if for any $x, y \in V$ we have*

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq 0. \tag{2}$$

Proof. ($\Rightarrow$) This direction is straightforward.
($\Leftarrow$)Define $x_\tau = x + \tau(y-x) \in V$. Then

$$\begin{aligned} f(y) &= f(x)+\int_0^1 \langle \nabla f(x+\tau(y-x)), y-x \rangle d\tau \\ &= f(x)+\langle \nabla f(x), y-x\rangle+\int_0^1\left\langle\nabla f\left(x_\tau\right)-\nabla f(x), y-x\right\rangle d\tau \\ &= f(x)+\langle \nabla f(x), y-x\rangle+\int_0^1 \frac{1}{\tau}\left\langle\nabla f\left(x_\tau\right)-\nabla f(x), x_\tau-x\right\rangle d\tau \\ &\geq f(x)+\langle \nabla f(x), y-x\rangle. \end{aligned}$$

$\square$

For any $f \in \mathcal{C}^1$, the linear Taylor expansion at x is

$$f_l(y;x) := f(x) + \langle \nabla f(x), y-x \rangle.$$

For $f \in \mathcal{S}^1(V)$, the Bregman divergence is defined as

$$D_f(y,x) := f(y) - f_l(y;x) = f(y) - f(x) - \langle \nabla f(x), y-x \rangle$$

For fixed $x \in V, D_f(\cdot, x)$ is also convex. Bregman divergence is in general asymmetric, i.e.,

$$D_f(y,x) \neq D_f(x,y), \quad \text{ if } x \neq y$$

We then introduce its symmetrization

$$M_{\nabla f}(x,y) := D_f(y,x) + D_f(x,y) = \langle \nabla f(x) - \nabla f(y), x-y \rangle$$

Theorem 1.8 (Definition using twice differentiablility)**.** *Let V be an open set. A twice continuously differentiable function f belongs to the class $\mathcal{S}^2(V)$ if and only if for any $x \in V$ we have*

$$\nabla^2 f(x) \succeq 0$$

Proof. ($\Rightarrow$) Let a function f from $\mathcal{C}^2(V)$ be convex and $s \in \mathbb{R}^n$. Let $x_\tau = x + \tau s \in V$ for $\tau > 0$ small enough. Then, in view of (2), we have

$$\begin{aligned} 0 &\le \frac{1}{\tau^2} \langle \nabla f(x_\tau) - \nabla f(x), x_\tau - x \rangle = \frac{1}{\tau} \langle \nabla f(x_\tau) - \nabla f(x), s \rangle \\ &= \frac{1}{\tau} \int_0^\tau \left\langle \nabla^2 f(x + \lambda s) s, s \right\rangle d\lambda \end{aligned}$$

and we get the desired result by letting τ tend to zero.

($\Leftarrow$) For all $x, y \in V$ we have

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

$\square$

Let us look at some examples of differentiable convex functions on $\mathbb{R}^n$.

- Every linear function $f(x) = \alpha + \langle a, x \rangle$ is convex.
- Let matrix A be symmetric and positive semidefinite. Then the quadratic function
$$f(x) = \alpha + \langle a, x \rangle + \frac{1}{2} \langle Ax, x \rangle$$
is convex (since $\nabla^2 f(x) = A \succeq 0$).
- The following functions of one variable belong to $\mathcal{S}^1(\mathbb{R})$:
$$\begin{aligned} f(x) &= e^x, \\ f(x) &= |x|^p, \quad p > 1, \\ f(x) &= \frac{x^2}{1 - |x|}, \\ f(x) &= |x| - \ln(1 + |x|) \end{aligned}$$
- ℓ_p-norm approximation:
$$f(x) = \sum_{i=1}^m |\langle a_i, x \rangle - b_i|^p$$

2. Smooth Functions

Denote by $\mathcal{C}_L^{1,1}$ the set of all $\mathcal{C}^1$ functions, the gradient for which is Lipschitz continuous with constant $L > 0$:

$$\|\nabla f(x) - \nabla f(y)\|_* \leqslant L \|x - y\| \quad \forall x, y \in V$$

where, for $g \in V^*$, the dual norm is

$$\|g\|_* := \sup_{\substack{v \in V \\ \|v\| = 1}} \langle g, v \rangle = \sup_{v \in V \setminus \{0\}} \frac{\langle g, v \rangle}{\|v\|}.$$

Theorem 2.1 (Bounds using Lipschitz smoothness)**.** *All conditions below, holding for all $x, y \in \mathbb{R}^n$, are equivalent to the inclusion $f \in \mathcal{C}_L^{1,1} \cap \mathcal{S}^1$:*

(1) $0 \le D_f(y, x) \le \frac{L}{2} \|x - y\|^2$,
(2) $\frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|_*^2 \le D_f(y, x)$,
(3) *(co-coercivity)* $\frac{1}{L} \|\nabla f(x) - \nabla f(y)\|_*^2 \le M_{\nabla f}(x, y)$,

(4) $0 \le M_{\nabla f}(x,y) \le L\|x-y\|^2$.

Proof. $f \in \mathcal{C}_L^{1,1} \cap \mathcal{S}^1 \Rightarrow (1)$. Indeed, the first inequality in (1) follows from the definition of convex functions. To prove the second one, note that

$$\begin{aligned} D_f(y,x) &= f(y) - f(x) - \langle \nabla f(x), y-x\rangle \\ &= \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x), y-x\rangle d\tau \\ &\le \int_0^1 L\tau\|y-x\|^2 d\tau = \frac{L}{2}\|y-x\|^2. \end{aligned}$$

$(1) \Rightarrow (2)$. Let us fix $x_0 \in \mathbb{R}^n$. Consider the function $\phi(y) = f(y) - \langle \nabla f(x_0), y\rangle$. Note thats $\phi(y) - \phi(x_0) = D_f(y,x_0) \ge 0$ for any y. Therefore, we have

$$\begin{aligned} \phi(x_0) &= \min_{x\in\mathbb{R}^n} \phi(x) \overset{(1)}{\le} \min_{x\in\mathbb{R}^n}\left\{\phi(y) + \langle \nabla\phi(y), x-y\rangle + \frac{L}{2}\|x-y\|^2\right\} \\ &= \min_{r\ge 0}\left\{\phi(y) - r\|\nabla\phi(y)\|_* + \frac{L}{2}r^2\right\} = \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|_*^2 \end{aligned}$$

and we get (2) since $\nabla\phi(y) = \nabla f(y) - \nabla f(x_0)$.

$(2) \Rightarrow (3)$ is trivial.

$(3) \Rightarrow f \in \mathcal{C}_L^{1,1} \cap \mathcal{S}^1$. Applying the Cauchy-Schwarz inequality to (3), we get $\|\nabla f(x) - \nabla f(y)\|_* \le L\|x-y\|$.

$(1) \Rightarrow (4)$ is trivial. $(4) \Rightarrow (1)$ is similar to Theorem 1.7 using integration. □

Theorem 2.2. *A twice continuously differentiable function f belongs to the class $\mathcal{C}_L^{2,1}$ if and only if for any $x, h \in \mathbb{R}^n$ we have*

$$0 \le \left\langle \nabla^2 f(x)h, h\right\rangle \le L\|h\|^2$$

Proof. Similar to Theorem 1.8. □

3. Strongly Convex Functions

Definition 3.1. *A continuously differentiable function $f(\cdot)$ is called strongly convex on $\mathbb{R}^n$ (notation $f \in \mathcal{S}_\mu^1(V)$) if there exists a constant $\mu > 0$ such that for any $x, y \in V$ we have*

$$f(y) \ge f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{2}\mu\|y-x\|^2.$$

The constant μ is called the convexity parameter of function f. Our notation for convex function $\mathcal{S}^1(V) = \mathcal{S}_0^1(V)$ with $\mu = 0$.

Let us describe the result of addition of two strongly convex functions.

Lemma 3.2. *If $f_1 \in \mathcal{S}_{\mu_1}^1(V_1)$, $f_2 \in \mathcal{S}_{\mu_2}^1(V_2)$ and $\alpha, \beta \ge 0$, then $f = \alpha f_1 + \beta f_2 \in \mathcal{S}_{\alpha\mu_1+\beta\mu_2}^1(V_1 \cap V_2)$.*

Theorem 3.3. *Let f be continuously differentiable. For all $x, y \in V$,*

$$M_{\nabla f}(x,y) \ge \mu\|x-y\|^2$$

is equivalent to inclusion $f \in \mathcal{S}_\mu^1(V)$.

The properties in the next statement is useful.

Theorem 3.4. *If $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$, then for any x and y from $\mathbb{R}^n$ we have*

$$D_f(y,x) \leq \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|_*^2,$$
$$M_{\nabla f}(x,y) \leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_*^2,$$
$$\mu\|x-y\| \leq \|\nabla f(x) - \nabla f(y)\|_*.$$

The proof of this theorem is very similar to the proof of Theorem 2.1 and we leave it as an exercise for the reader.

Let us present a second-order characterization of the class $\mathcal{S}^1_\mu(V)$.

Theorem 3.5 (Definition using twice differentiability)**.** *Let a continuous function f be twice continuously differentiable in int V. It belongs to the class $\mathcal{S}^2_\mu(V)$ if and only if for all $x \in \operatorname{int} V$ and $h \in \mathbb{R}^n$ we have*

$$\left\langle \nabla^2 f(x)h, h \right\rangle \succeq \mu\|h\|^2.$$

Notice that for $f \in \mathcal{S}^1_\mu$, $D_f(y,x)$ is positive, and furthermore if f is twice differentiable

$$D_f(y,x) = \frac{1}{2}\|y-x\|^2_{\nabla^2 f(\xi(x))}$$

by Taylor expansion. Therefore, $D_f(y,x)$ induced a special metric associated to the Hessian of a strongly convex function.

References

[1] Y. Nesterov et al. *Lectures on convex optimization*, volume 137. Springer, 2018. 1