Fenchel Duality between Strong Convexity and Lipschitz Continuous Gradient

Xingyu Zhou

The Ohio State University

zhou.2055@osu.edu

December 5, 2017

Outline

Main Result

Theorem

(i) *If f is closed and strong convex with parameter μ, then f^* has a Lipschitz continuous gradient with parameter $\frac{1}{\mu}$.*

(ii) *If f is convex and has a Lipschitz continuous gradient with parameter L, then f^* is strong convex with parameter $\frac{1}{L}$.*

- We provide a very simple proof for this theorem (two-line).
- To this end, we first present equivalent conditions for strong convexity and Lipschitz continuous gradient.

Outline

Outline

Definition

Definition (Convexity)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ with the extended-value extension. Then, f is said to be a convex function if for any $\alpha \in [0, 1]$,

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

for any x and y.

Note: This definition does not require the differentiability of f.

- We are familiar that this definition is equivalent to the *first-order* condition and *monotonicity* condition assuming f is differentiable.
- In the following, we will first show that this equivalence still holds for a non-smooth function.

Outline

Equivalent Conditions for Convexity

Lemma (Equivalence for Convexity)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ with the extended-value extension. Then, the following statements are equivalent:

(i) *(Jensen's inequality): f is convex.*

(ii) *(First-order): $f(y) \geq f(x) + s_x^T(y - x)$ for all x, y and any $s_x \in \partial f(x)$.*

(iii) *(Monotonicity of subgradient): $(s_y - s_x)^T(y - x) \geq 0$ for all x, y and any $s_x \in \partial f(x)$, $s_y \in \partial f(y)$.*

Note: The lemma is a non-trivial extension as we will see later in the proof. Moreover, this lemma will be extremely useful as we later deal with strong convexity.

Specifically, we will prove this lemma by showing $(i) \implies (ii) \implies (iii) \implies (i)$.

Proof

(i) $\implies$ (ii): Since f is convex, by definition, we have for any $0 < \alpha \leq 1$

$$
\begin{aligned}
& f(x+\alpha(y-x)) \leq \alpha f(y) + (1-\alpha) f(x) \\
\iff & \frac{f(x+\alpha(y-x)) - f(x)}{\alpha} \leq f(y) - f(x) \\
\overset{(a)}{\implies} & f'(x, y-x) \leq f(y) - f(x) \\
\overset{(b)}{\implies} & s_x^T (y-x) \leq f(y) - f(x) \quad \forall s_x \in \partial f(x)
\end{aligned}
$$

- (a) is obtained by letting $\alpha \to 0$ and the definition of *directional derivate*. Note that the limit always exists for a convex f by the monotonicity of difference quotient.
- (b) follows from the fact that $f'(x, v) = \sup_{s \in \partial f(x)} s^T v$ for a convex f.

Proof (Cont'd)

(ii) $\implies$ (iii): From (ii), we have

$$f(y) \geq f(x) + s_x^T (y - x)$$
$$f(x) \geq f(y) + s_y^T (x - y)$$

Adding together directly yields (iii).

Proof (Cont'd)

(iii) $\implies$ (i): To this end, we will show (iii) $\implies$ (ii) and (ii) $\implies$ (i).

First, (iii) $\implies$ (ii): Let $\phi(\alpha) := f(x + \alpha(y - x))$ and $x_\alpha := x + \alpha(y - x)$, then

$$f(y) - f(x) = \phi(1) - \phi(0) = \int_0^1 s_\alpha^T (y - x) d\alpha \tag{1}$$

where $s_\alpha \in \partial f(x_\alpha)$ for $t \in [0, 1]$. Then by (iii), for any $s_x \in \partial f(x)$, we have

$$s_\alpha^T (y - x) \geq s_x^T (y - x)$$

which combined with inequality above directly implies (ii).
Second, (ii) $\implies$ (i): By (ii), we have

$$f(y) \geq f(x_\alpha) + s_t^T (y - x_\alpha) \tag{2}$$

$$f(x) \geq f(x_\alpha) + s_t^T (x - x_\alpha) \tag{3}$$

Multiplying (2) by α and (3) by $1 - \alpha$, and adding together yields (i). □

Outline

Outline

Strongly Convex Function

Definition (Strong Convexity)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ with the extended-value extension. Then, f is said to be strongly convex with parameter μ if

$$g(x) = f(x) - \frac{\mu}{2} x^T x$$

is convex.

Note: As before, this definition does not require the differentiability of f.

In the following, by applying the previous equivalent conditions of convexity to g, we can easily establish equivalence between strong convexity.

Outline

Equivalent Conditions for Strong Convexity

Lemma (Equivalence for Strong Convexity)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ with the extended-value extension. Then, the following statements are equivalent:

(i) *f is strongly convex with parameter μ.*

(ii) $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\mu}{2}\alpha(1-\alpha)\|y-x\|^2$ *for any x, y.*

(iii) $f(y) \geq f(x) + s_x^T(y-x) + \frac{\mu}{2}\|y-x\|^2$ *for all x, y and any $s_x \in \partial f(x)$.*

(iv) $(s_y - s_x)^T(y-x) \geq \mu\|y-x\|^2$ *for all x, y and any $s_x \in \partial f(x)$, $s_y \in \partial f(y)$.*

Proof:

(i) $\iff$ (ii): It follows from the definition of convexity for $g(x) = f(x) - \frac{\mu}{2}x^T x$.

(i) $\iff$ (iii): It follows from the equivalent first-order condition of convexity for g.

(i) $\iff$ (iv): It follows from the equivalent monotonicity of (sub)-gradient of convexity for g. □

Outline

Conditions Implied by Strong Convexity (SC)

Lemma (Implications of SC)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ with the extended-value extension. The following conditions are all implied by strong convexity with parameter μ:

(i) $\frac{1}{2}\|s_x\|^2 \geq \mu(f(x) - f^*)$, $\forall x$ *and* $s_x \in \partial f(x)$.

(ii) $\|s_y - s_x\| \geq \mu\|y - x\|$ $\forall x, y$ *and any* $s_x \in \partial f(x)$, $s_y \in \partial f(y)$.

(iii) $f(y) \leq f(x) + s_x^T(y - x) + \frac{1}{2\mu}\|s_y - s_x\|^2$ $\forall x, y$ *and any* $s_x \in \partial f(x)$, $s_y \in \partial f(y)$.

(iv) $(s_y - s_x)^T(y - x) \leq \frac{1}{\mu}\|s_y - s_x\|^2$ $\forall x, y$ *and any* $s_x \in \partial f(x)$, $s_y \in \partial f(y)$.

Note: In the case of smooth function f, (i) reduced to Polyak-Lojasiewicz (PL) inequality, which is more general than SC and is extremely useful for linear convergence.

Proof

$$\boxed{\text{(i) } \frac{1}{2}\|s_x\|^2 \geq \mu(f(x) - f^*)}$$

(SC) $\implies$ (i): By equivalence of strong convexity in the previous lemma, we have

$$f(y) \geq f(x) + s_x^T(y-x) + \frac{\mu}{2}\|y-x\|^2$$

Taking minimization with respect to y on both sides, yields

$$f^* \geq f(x) - \frac{1}{2\mu}\|s_x\|^2$$

Re-arranging it yields (i).

Proof (Cont'd)

$$\boxed{\text{(ii) } \|s_y - s_x\| \geq \mu \|y - x\|}$$

SC $\implies$ (ii): By equivalence of strong convexity in the previous lemma, we have

$$(s_y - s_x)^T (y - x) \geq \mu \|y - x\|^2$$

Applying Cauchy-Schwartz inequality to the left-hand-side, yields (ii).

Proof (Cont'd)

$$\boxed{\text{(iii) } f(y) \leq f(x) + s_x^T(y-x) + \tfrac{1}{2\mu}\left\|s_y - s_x\right\|^2}$$

SC $\Longrightarrow$ (iii): For any given $s_x \in \partial f(x)$, let $\phi_x(z) := f(z) - s_x^T z$. First, since

$$(s_{z_1}^{\phi} - s_{z_2}^{\phi})(z_1 - z_2) = (s_{z_1}^{f} - s_{z_2}^{f})(z_1 - z_2) \geq \mu \left\|z_1 - z_2\right\|^2$$

which implies that $\phi_x(z)$ is also strong convexity with parameter μ.

Then, applying PL-inequality to $\phi_x(z)$, yields

$$\begin{aligned} \phi_x^* = f(x) - s_x^T x \geq \phi_x(y) - \frac{1}{2\mu}\left\|s_y^{\phi}\right\|^2 \\ = f(y) - s_x^T y - \frac{1}{2\mu}\left\|s_y - s_x\right\|^2. \end{aligned}$$

Re-arranging it yields (iii).

Proof (Cont'd)

$$\boxed{\text{(iv) } (s_y - s_x)^T(y - x) \leq \frac{1}{\mu}\|s_y - s_x\|^2}$$

SC $\implies$ (iv): From previous result, we have

$$f(y) \leq f(x) + s_x^T(y - x) + \frac{1}{2\mu}\|s_y - s_x\|^2$$
$$f(x) \leq f(y) + s_y^T(x - y) + \frac{1}{2\mu}\|s_x - s_y\|^2$$

Adding together yields (iv). □

Outline

Outline

Definition

Definition

A differentiable function f is said to have an L-Lipschitz continuous gradient if for some $L > 0$

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \ \forall x, y.$$

Note: The definition does not require the convexity of f.

In the following, we will see different conditions that are equivalent or implied by Lipschitz continuous gradient condition.

Outline

Relationship

[0] $\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|, \ \forall x, y.$

[1] $g(x) = \frac{L}{2}x^Tx - f(x)$ is convex , $\forall x$

[2] $f(y) \le f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|^2, \ \forall x, y.$

[3] $(\nabla f(x) - \nabla f(y)^T(x-y) \le L\|x-y\|^2, \ \forall x, y.$

[4] $f(\alpha x + (1-\alpha)y) \ge \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)L}{2}\|x-y\|^2, \ \forall x, y$ and $\alpha \in$

[5] $f(y) \ge f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2, \ \forall x, y.$

[6] $(\nabla f(x) - \nabla f(y)^T(x-y) \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2, \ \forall x, y.$

[7] $f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)}{2L}\|\nabla f(x) - \nabla f(y)\|^2, \ \forall x, y$

Relationship (Cont'd)

Lemma

For a function f with a Lipschitz continuous gradient over $\mathbb{R}^n$, the following relations hold:

$$[5] \iff [7] \implies [6] \implies [0] \implies [1] \iff [2] \iff [3] \iff [4] \quad (4)$$

If the function f is convex, then all the conditions [0]-[7] are equivalent.

Note: From (5), we can see that conditions [0]-[7] can be grouped into four classes.

- Class A: [1]-[4]
- Class B: [5],[7]
- Class C: [6]
- Class D: [0]

Proof

Class A

$$[1]\ g(x) = \frac{L}{2}x^T x - f(x) \text{ is convex}$$

$$[2]\ f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|^2$$

$$[3]\ (\nabla f(x) - \nabla f(y))^T (x - y) \leq L\|x - y\|^2$$

$$[4]\ f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)L}{2}\|x - y\|^2$$

[1] $\iff$ [2]: It follows from the first-order equivalence of convexity.
[1] $\iff$ [3]: It follows from the monotonicity of gradient equivalence of convexity.
[1] $\iff$ [4]: It follows from the definition of convexity.

Proof (Cont'd)

Class B

[5] $f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2.$

[7] $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)}{2L}\|\nabla f(x) - \nabla f(y)\|^2.$

[5] $\implies$ [7]: Let $x_\alpha := \alpha x + (1-\alpha)y$, then

$$f(x) \geq f(x_\alpha) + \nabla f(x_\alpha)^T(x - x_\alpha) + \frac{1}{2L}\|\nabla f(x) - \nabla f(x_\alpha)\|^2$$
$$f(y) \geq f(x_\alpha) + \nabla f(z)^T(y - x_\alpha) + \frac{1}{2L}||\nabla f(y) - \nabla f(x_\alpha)||^2$$

Multiplying the first inequality with α and second inequality with $1-\alpha$, and adding them together and using $\alpha\|x\|^2 + (1-\alpha)\|y\|^2 \geq \alpha(1-\alpha)\|x-y\|^2$, yields [7].

Proof (Cont'd)

Class B

$$[5]\ f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2.$$

$$[7]\ f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \frac{\alpha(1-\alpha)}{2L}\|\nabla f(x) - \nabla f(y)\|^2.$$

[7] $\implies$ [5]: Interchanging x and y in [7] and re-writing it as

$$f(y) \geq f(x) + \frac{f(x+\alpha(y-x)) - f(x)}{\alpha} + \frac{1-\alpha}{2L}\|\nabla f(x) - \nabla f(y)\|^2$$

Letting $\alpha \to 0$, yields [5].

Relationship (Cont'd)

Lemma

For a function f with a Lipschitz continuous gradient over $\mathbb{R}^n$, the following relations hold:

$$[5] \iff [7] \implies [6] \implies [0] \implies [1] \iff [2] \iff [3] \iff [4] \tag{5}$$

If the function f is convex, then all the conditions [0]-[7] are equivalent.

Note: From (5), we can see that conditions [0]-[7] can be grouped into four classes.

- Class A: [1]-[4]
- Class B: [5],[7]
- Class C: [6]
- Class D: [0]

Proof (Cont'd)

Class B ([5]) $\Longrightarrow$ Class C ([6]) $\Longrightarrow$ Class D ([0]) $\Longrightarrow$ Class A ([3])

[5] $f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2.$

[6] $(\nabla f(x) - \nabla f(y))^T (x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$

[0] $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$

[3] $(\nabla f(x) - \nabla f(y))^T(x - y) \leq L\|x - y\|^2$

[5] $\Longrightarrow$ [6]: Interchanging x and y and adding together.
[6] $\Longrightarrow$ [0]: It follows directly from Cauchy-Schwartz inequality.
[0] $\Longrightarrow$ [3]: It following directly from Cauchy-Schwartz inequality.

Proof (Cont'd)

Class A ([3]) $\implies$ Class B ([5]) for convex f

$$[3]\ (\nabla f(x) - \nabla f(y))^T(x-y) \leq L\|x-y\|^2.$$

$$[5]\ f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2.$$

[3] $\implies$ [5]: Let $\phi_x(z) := f(z) - \nabla f(x)^T z$. Note that since f is convex, $\phi_x(z)$ attains its minimum at $z^* = x$. By [3], we have

$$(\nabla \phi_x(z_1) - \nabla \phi_x(z_2)^T(z_1 - z_2) \leq L\|z_1 - z_2\|^2$$

which implies that

$$\phi_x(z) \leq \phi_x(y) + \nabla \phi_x(y)^T(z-y) + \frac{L}{2}\|z-y\|^2.$$

Taking minimization over z, yields [5]. □

Outline

Outline

Definition

Definition (Conjugate)

The conjugate of a function f is

$$f^*(s) = \sup_{x \in \mathrm{dom} f} (s^T x - f(x))$$

Note: f^* is convex and closed (even f is not).

Definition (Subgradient)

s is a subgradient of f at x if

$$f(y) \geq f(x) + s^T(y - x) \text{ for all } y$$

Note: f need not be convex.

Outline

Useful Results

Lemma (Fenchel-Young's Equality)

Consider the following conditions for a general f:

$$[1]\ f^*(s) = s^T x - f(x)$$
$$[2]\ s \in \partial f(x)$$
$$[3]\ x \in \partial f^*(s)$$

Then, we have

$$[1] \iff [2] \implies [3]$$

Further, if f is closed and convex, then all these conditions are equivalent.

Lemma (Differentiability)

For a closed and strictly convex f, $\nabla f^*(s) = \text{argmax}_x(s^T x - f(x))$.

Proof

$$[1]\ f^*(s) = s^T x - f(x)$$
$$[2]\ s \in \partial f(x)$$

[1] $\iff$ [2]:

$$\begin{aligned} s \in \partial f(x) &\iff f(y) \geq f(x) + s^T(y-x)\ \forall y \\ &\iff s^T x - f(x) \geq s^T y - f(y)\ \forall y \\ &\iff s^T x - f(x) \geq f^*(s) \end{aligned}$$

Since $f^*(s) \geq s^T x - f(x)$ always holds, [1] $\iff$ [2].

Proof (Cont'd)

$$[2]\ s \in \partial f(x)$$
$$[3]\ x \in \partial f^*(s)$$

[2] $\implies$ [3]: If [2] holds, then by previous result we have $f^*(s) = s^T x - f(x)$. Thus,

$$\begin{aligned} f^*(z) &= \sup_u (z^T u - f(u)) \\ &\geq z^T x - f(x) \\ &= s^T x - f(x) + x^T (z - s) \\ &= f^*(s) + x^T (z - s) \end{aligned}$$

This holds for all z, which implies $x \in \partial f^*(s)$.

For a closed and convex function, we have [3] $\implies$ [2]: This follows from $f^{**} = f$.

Proof (Differentiability)

- Suppose $x_1 \in \partial f^*(s)$ and $x_2 \in \partial f^*(s)$.
- By closeness and convexity, we have [3] $\implies$ [2], which gives

$$s \in \partial f x_1 \text{ and } s \in \partial f x_2$$

- By previous result ([2] $\implies$ [1]), we have

$$x_1 = \text{argmax}(s^T x - f(x)) \text{ and } x_2 = \text{argmax}(s^T x - f(x))$$

- By strictly convexity, $s^T x - f(x)$ has a unique maximizer for every s. Thus, $x := x_1 = x_2$
- Thus $\partial f^*(s) = \{x\}$, which implies f^* is differentiable at s and $\nabla f^*(s) = \text{argmax}_x(s^T x - f(x))$.

Outline

Proof of Main Result

Theorem

(i) *If f is closed and strong convex with parameter μ, then f^* has a Lipschitz continuous gradient with parameter $\frac{1}{\mu}$.*

(ii) *If f is convex and has a Lipschitz continuous gradient with parameter L, then f^* is strong convex with parameter $\frac{1}{L}$.*

Proof: (1): By implication of strong convexity, we have

$$\|s_x - s_y\| \geq \mu \|x - y\| \;\; \forall s_x \in \partial f(x), s_y \in \partial f(y)$$

which implies

$$\|s_x - s_y\| \geq \mu \|\nabla f^*(s_x) - \nabla f^*(s_y)\|$$

Hence, f^* has a Lipschitz continuous gradient with $\frac{1}{\mu}$.

Proof (Cont'd)

(2): By implication of Lipschitz continuous gradient for convex f, we have

$$(\nabla f(x) - \nabla f(y)^T (x - y) \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2$$

which implies

$$(s_x - s_y)^T (x - y) \geq \frac{1}{L}(s_x - s_y) \; \forall x \in \partial f^*(s_x), y \in \partial f^*(s_y)$$

Hence, f^* is strongly convex with parameter $\frac{1}{L}$.

Summary

- A very simple proof of Fenchel duality between strong convexity and Lipschitz continuous gradient.
- A summary of conditions for strong convexity.
- A summary of conditions for Lipschitz continuous gradient
- Common tricks:
  - Taking limit ($\alpha \to 0$).
  - Taking minimization over both sides.
  - Interchanging x and y, and adding together.
  - Applying mean-value theorem.
  - Applying Cauchy-Schwartz inequality.

Thank you!
Q & A