Strong convexity and Lipschitz continuity of gradients

Stephen Becker
Applied Math, U. Colorado Boulder stephen.becker@colorado.edu

February 20, 2023

Lipschitz continuity of derivative and/or strong convexity of f

The definition of Lipschitz continuity of ∇f (with constant L) is

$$\forall x, y \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \tag{1}$$

and the definition of f being μ strongly convex means that the function $x \mapsto f(x) - \frac{\mu}{2}\|x\|^2$ is convex[1]. In the lines below, if L or μ appears, then we are assuming the gradient is Lipschitz with constant L or f is strongly convex with constant μ, respectively. Most references to Nesterov's book are to his first edition [Nes04], not the recent 2018 edition [Nes18].

These two inequalities are very helpful; see, e.g., Thm 2.1.5 and Thm 2.1.10 from [Nes04].

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

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|^2 \tag{3}$$

If we drop convexity but keep Lipschitz continuity of the gradient, then the first equation is still true, but the second equation is not true with $\mu = 0$, but it is true with $\mu = -L$. This is often written as $|f(y) - (f(x) + \langle \nabla f(x), y - x \rangle)| \leq \frac{L}{2}\|x - y\|^2$.

The main inequalities can be summarized by:

$$\left.\begin{array}{rl} L^{-1}\|\nabla f(x) - \nabla f(y)\|^2 & \text{(a)} \\ \mu\|x - y\|^2 & \text{(b)} \\ \dfrac{\mu L}{\mu + L}\|x - y\|^2 + \dfrac{1}{\mu + L}\|\nabla f(x) - \nabla f(y)\|^2 & \text{(c)} \end{array}\right\} \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle \leq \begin{cases} \text{(d)} & L\|x - y\|^2 \\ \text{(e)} & \mu^{-1}\|\nabla f(x) - \nabla f(y)\|^2 \end{cases} \tag{4}$$

The inequality (a) really follows from the co-coercivity of gradients; this result is actually surprisingly strong, since it makes implicit use of the Baillon-Haddad theorem. The result (e) for μ also requires f be continuously differentiable. The (c) inequality assumes both strong convexity and Lipschitz continuity of the gradient; see [Nes04, Thm. 2.1.12] for a derivation.

Restating some of the above

Denote the set of C^1 convex functions with Lipschitz continuous gradients to be $\mathcal{F}^1_L(X) = \mathcal{F}^1(X) \cap C^1_L(X)$. Then $f \in \mathcal{F}^1_L(X)$ implies the following for all $x, y \in X$ (see Thm. 2.1.5 [Nes04]):

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

The proof is elementary but some parts are a bit clever.

[1] See Thm. 5.17 and Remark 5.18 in [Bec17] — this is actually only true if $\|\cdot\|$ is the induced norm from the inner product. However, most other properties hold for a general norm.

Characterization via conjugate functions

We use "L-strongly smooth" to mean the gradient is L-Lipschitz. Let E be an finite-dimensional inner product space. In Euclidean space, both primal and dual norms are the Euclidean norm.

Theorem 1 (Thm. 5.26 in [Bec17], "conjugate correspondence theorem")**.** *Let $\sigma > 0$, then*

1. *If $f : E \to \mathbb{R}$ is a $1/\sigma$-strongly-smooth then f^* is σ-strongly-convex (with respect to the dual norm); and*
2. *If $f : E \to (-\infty, +\infty]$ is a proper, lsc, σ-strongly-convex function, then f^* is $1/\sigma$-strongly-smooth.*

A similar theorem with a slightly different setup is Theorem 18.15 in [BC17].
In a similar flavor,

Theorem 2 (Prop. 18.9 in [BC17])**.** *Let $f \in \Gamma_0(\mathcal{H})$ be such that f^* is strictly convex on every nonempty convex subset of $\operatorname{dom} \partial f^*$. Then f is Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f$.*

For notions of "essentially differentiable" and "locally Lipschitz gradient", see Goebel & Rockafellar 2008 "Local strong convexity and local Lipschitz continuity of the gradient of convex functions."

References

[BC17] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces,* 2nd edition, Springer, 2017.

[Bec17] A. Beck, *First-Order Methods in Optimization*, SIAM, 2017.

[Nes04] Yu. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87 of *Applied Optimization*. Kluwer, Boston, 2004.

[Nes18] Yu. Nesterov. *Lectures on Convex Optimization*. Springer International Publishing, 2018.