Strong convexity and Lipschitz continuity of gradients

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

January 12, 2021

Lipschitz continuity of derivative and/or strong convexity of f

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

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

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].

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

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

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

The main inequalities can be summarized by:

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

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

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.

References

[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.