Analysis Toolbox: Strong convexity and Lipschitz continuity of gradients

Stephen Becker

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

Created Sep 27 2019, last edited September 27, 2019

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. 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:

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

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.

Sub-optimality bounds Added Sept 17 2019. For unconstrained smooth optimization, if $x^\star$ is a minimizer, then $\nabla f(x^\star) = 0$. Note there are 3 equivalent definitions of optimality: x is optimal if

$$\|x - x^\star\| = 0, \quad f(x) - f^\star = 0, \quad \|\nabla f(x)\| = 0 \tag{5}$$

and this would be "iff" if we assume the optimal solution is unique. Now, given a Lipschitz continuous derivative, we can bound

$$\|\nabla f(x)\| = \|\nabla f(x) - \nabla f(x^\star)\| \leq L\|x - x^\star\| \quad \text{by (1)} \tag{6}$$

$$f(x) - f^\star \leq \frac{L}{2}\|x - x^\star\|^2 \quad \text{by (2)} \tag{7}$$

$$\|\nabla f(x)\|^2 \leq 2L\left(f(x) - f^\star\right) \quad \text{by Eq. (9.14) in [BV04]} \tag{8}$$

and given μ strong convexity, we can bound in the other direction:

$$\|x - x^\star\|^2 \leq \frac{1}{\mu^2} \|\nabla f(x)\|^2 \quad \text{by (4) (b) and (e)} \tag{9}$$

$$\|x - x^\star\|^2 \leq \frac{2}{\mu} \left(f(x) - f^\star \right) \quad \text{by (3), with } x = x^\star, y = x \tag{10}$$

$$f(x) - f^\star \leq \frac{1}{2\mu} \|\nabla f(x)\|^2 \quad \text{by Eq. (9.9) in [BV04]. This is PL} \tag{11}$$

Given both L and μ, we can combine the bounds, and bound any one of the 3 error metrics in terms of another, i.e., $\|\nabla f(x)\|^2 \leq \frac{2L^2}{\mu} \left(f(x) - f^\star \right)$ and $f(x) - f^\star \leq \frac{L}{2\mu^2} \|\nabla f(x)\|^2$. But these are not good bounds; the bounds in Eq (8) and (11) are better. Note: (11) is the Polyak-Lojasiewicz (PL) inequality, see Karimi, Nutini, Schmidt for details.

References

[BC11]   H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces, 1st edition*, Springer, 2011.

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

[BV04]   S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[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.