Strong convexity and Lipschitz continuity of gradients

Stephen Becker
LJLL
Paris, France
stephen.becker@upmc.fr

May 30, 2013

Abstract

Collection of known results

Lipschitz continuity of derivative or strong convexity of f Nesterov's book Thm 2.1.5 and Thm 2.1.10. 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.

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

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

$$L^{-1}\|\nabla f(x) - \nabla f(y)\|^2 \;\leq\; \langle \nabla f(x) - \nabla f(y), x - y \rangle \;\leq\; L\|x - y\|^2 \tag{3}$$

$$\mu\|x - y\|^2 \;\leq\; \langle \nabla f(x) - \nabla f(y), x - y \rangle \;\leq\; \mu^{-1}\|\nabla f(x) - \nabla f(y)\|^2. \tag{4}$$

The left-most $\leq$ above in the line for L really follows from the co-coercivity of gradients. The second result for μ also requires f be continuously differentiable. This result is actually surprisingly strong, since it makes implicit use of the Baillon-Haddad theorem.

Of course we also have the definition of Lipschitz continuity:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \tag{5}$$

Lipschitz continuity of derivative *and* strong convexity of f By assuming both properties at once, we get (Nesterov, Thm. 2.1.12)

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\mu L}{\mu + L}\|x - y\|^2 + \frac{1}{\mu + L}\|\nabla f(x) - \nabla f(y)\|^2. \tag{6}$$

Firmly non-expansive operators Let T be an operator (like a gradient) that is necessarily single-valued. Then it is **firmly non-expansive** if (Defn 4.1 Bauschke and Combettes' book)

$$\|Tx - Ty\|^2 + \|(I - T)x - (I - T)y\|^2 \leq \|x - y\|^2 \tag{7}$$

Proposition 1 (Prop 4.2 in Bauschke-Combettes)**.** *TFAE*

1. *T is firmly non-expansive*
2. *$I - T$ is firmly non-expansive*
3. *$2T - I$ is non-expansive*
4. *$\|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty \rangle$ and from paper with Wajs, $\|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(I - T)x - (I - T)y\|^2$*
5. *$0 \leq \langle Tx - Ty, (I - T)x - (I - T)y \rangle$*

We say T is β **cocoercive** (or β-inverse **strongly monotone**) if βT is firmly nonexpansive (Defn 4.4). N.B. This is an unusual use of β: you might think β^{-1} is more natural, so do not confuse it!

Exercise 4.12: if T_1 and T_2 are both firmly nonexpansive, then $T_1 - T_2$ is nonexpansive. N.B. while nonexpansiveness is closed under composition, firm nonexpansiveness is not!

Fact: the projector onto a closed convex set is firmly nonexpansive (Prop 4.8). More generally, all proximity operators are firmly nonexpansive (Prop 12.27). Even more generally, all resolvents $T = (A + I)^{-1}$ are firmly nonexpansive iff A is monotone, and T has full domain iff A is maximally monotone.

Fact (corollary 16.16, Baillon-Haddad theorem). Let f be Frechet differentiable convex. Then ∇f is β-Lipschitz continuous *if and only if* ∇f is β^{-1} cocoercive; in particular, ∇f is nonexpansive iff it is firmly nonexpansive.

Monotone A set-valued operator A is **monotone** if $\langle x - y, Ax - Ay\rangle \geq 0$ (note: I use Ax to denote an arbitrary element of the set Ax), and **strongly monotone** if $A - \beta I$ is monotone, i.e., $\langle x - y, Ax - Ay\rangle \geq \beta\|x - y\|^2$. See defn. 22.1. These notions can be localized to a subset C. Obvious fact: if f is **strongly convex** with constant β, then ∂f is strongly monotone with β. Vandenberghe's notes use "strongly monotone" (with $A = \nabla f$) and "coercive" interchangeable.

Let T be a single-valued operator and $A = T^{-1}$. Then T is β-cocoercive iff A is strongly monotone with constant β.

Defn 11.10, coercive: f is coercive of $\lim_{\|x\|\to+\infty} f(x) = +\infty$ and supercoercive if $f(x)/\|x\| \to +\infty$ as well.