3 Smoothness and strong convexity

3.1 Dual norms

Recall that if $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then the dual norm is defined by

$$\|y\|_* = \sup_{\|x\|=1} \langle y, x\rangle .$$

In particular we have the generalized Cauchy-Schwarz inequality

$$\langle x, y\rangle \leq \|x\|\|y\|_* \qquad \forall x, y \in \mathbb{R}^n.$$

Exercise: Show that the dual norm of the Euclidean norm $\|x\|_2 = \sqrt{\langle x, x\rangle}$ is the Euclidean norm itself. More generally show the dual of the p-norm $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ is the q norm where $1/p + 1/q = 1$.

3.2 L-smoothness

We say that a differentiable function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *L-smooth* with respect to a norm $\|\cdot\|$, if for any $x, y \in \mathbf{int}\,\mathbf{dom}(f)$,

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

where $\|\cdot\|_*$ is the dual norm to $\|\cdot\|$. (We will sometimes omit the reference to the norm, in which case this means we work with the Euclidean norm.) The following lemma will be important in the analysis of optimization algorithms.

Lemma 1 (Descent lemma)**.** *If f is L-smooth, then for any $x \in \mathbf{int}\,\mathbf{dom}(f)$ and $y \in \mathbf{dom}(f)$,*

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

Remark 1. *To appreciate the implication of the inequality above, assume $\|\cdot\| = \|\cdot\|_2$ is the Euclidean norm, and consider taking $y = x - (1/L)\nabla f(x)$, i.e., one step of the gradient method with step size $t = 1/L$. Then we get $f(y) \leq f(x) - \frac{1}{2L}\|\nabla f(x)\|_2^2 < f(x)$, i.e., the function value decreases at each iteration.*

Proof of Lemma 1. Let $h = y - x$ and $\phi(t) = f(x + th) - (f(x) + t\langle \nabla f(x), h\rangle)$. Then ϕ is differentiable and $\phi'(t) = \langle \nabla f(x + th) - \nabla f(x), h\rangle \leq \|\nabla f(x + th) - \nabla f(x)\|_*\|h\| \leq Lt\|h\|^2$ where we used the Lipschitz assumption (1). Thus it follows that $\phi(1) = \phi(0) + \int_0^1 \phi'(t)dt \leq L/2\|h\|^2$ which gives precisely the desired inequality (2). □

A simple way to check L-smoothness, is by analyzing the Hessian matrix. One can show the following proposition:

Proposition 3.1. *Assume $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is such that $\mathbf{dom}(f)$ is open, and f is twice continuously differentiable on its domain. Then f is L-smooth if, and only if,*

$$\forall u, v \in \mathbb{R}^n, \quad \langle \nabla^2 f(x)u, v\rangle \leq L\|u\|\|v\|. \tag{3}$$

Remark 2. *Condition* (3) *can be equivalently written as* $\|\nabla^2 f(x)u\|_* \leq L\|u\|$ *for all* $u \in \mathbb{R}^n$. *Equivalently, this is saying that the* $(\mathbb{R}^n, \|\cdot\|) \to (\mathbb{R}^n, \|\cdot\|_*)$ *induced norm of the linear map* $\nabla^2 f(x)$ *is at most* L.

Proof. $\Leftarrow$ Let $x, y \in \mathbf{dom}(f)$. The fundamental theorem of calculus applied to the function $t \mapsto \nabla f(x + th)$ with $h = y - x$ tells us that $\nabla f(y) - \nabla f(x) = \int_0^1 \nabla^2 f(x + th)h dt$. Thus we can write

$$\|\nabla f(y) - \nabla f(x)\|_* \leq \int_0^1 \|\nabla^2 f(x + th)h\|_* dt \leq \int_0^1 L\|h\| dt = L\|h\|$$

as desired.

$\Rightarrow$ Assume f is L-smooth. Let u, v be arbitrary vectors, and define $\psi(t) = \langle \nabla f(x + tu) - \nabla f(x), v\rangle$. Then by L-smoothness, $\psi(t) \leq Lt\|u\|\|v\|$, and so $\psi'(0) = \lim_{t\to 0}(\psi(t) - \psi(0))/t \leq L\|u\|\|v\|$. But $\psi'(0) = \left\langle \nabla^2 f(x)u, v\right\rangle$. □

Remark 3. *If* $\|\cdot\| = \|\cdot\|_2$ *is the Euclidean norm, then condition* (3) *is equivalent to saying that the eigenvalues of* $\nabla^2 f(x)$ *are all in* $[-L, L]$.

3.3 Strong convexity

We say that f is m-strongly convex (with respect to the norm $\|\cdot\|$) if for any $x, y \in \mathbf{dom}(f)$, and $t \in [0, 1]$

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{m}{2}t(1-t)\|x - y\|^2. \tag{4}$$

- If f is m-strongly convex and differentiable at x, then for any $y \in \mathbf{dom}(f)$ we have

  $$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{m}{2}\|y - x\|^2. \tag{5}$$

  (This can be proved simply by subtracting $f(x)$ from both sides of (4), dividing by t, and letting $t \to 0$.) The converse is also true, i.e., if $\mathbf{dom}(f)$ is open, f is differentiable everywhere on its domain, and (5) holds for all $x, y \in \mathbf{dom}(f)$, then f is m-strongly convex. (Exercise)

- If f is twice continuously differentiable on its domain (assumed open), then strong convexity is equivalent to $\left\langle \nabla^2 f(x)h, h\right\rangle \geq m\|h\|^2$ for all $x \in \mathbf{dom}(f)$ and $h \in \mathbb{R}^n$. (Proof left as an exercise.)

Remark 4. *When considering the Euclidean norm, we see that a convex function* f *is* L*-smooth if, and only if,* $\nabla^2 f(x) \preceq LI$, *i.e.,* $LI - \nabla^2 f(x)$ *is positive semidefinite (where* I *is the identity matrix), i.e., all the eigenvalues of* f *are* $\leq L$. *Similarly, a function* f *is* m*-strongly convex if, and only if,* $\nabla^2 f(x) \succeq mI$, *i.e., all the eigenvalues of* $\nabla^2 f(x)$ *are* $\geq m$.

To summarize, if a function f is L-smooth, and m-strongly convex, then we can find, at any point $x \in \mathbf{int}\,\mathbf{dom}(f)$ *global* quadratic lower and upper bounds on f:

$$\underbrace{f(x) + \langle \nabla f(x), y - x\rangle + \frac{m}{2}\|y - x\|^2}_{\text{strong convexity}} \;\leq\; f(y) \;\leq\; \underbrace{f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2}_{L\text{-smoothness}}. \tag{6}$$

The ratio $\kappa = L/m$ can be interpreted as a *condition number* of f. This quantity will play a prominent role in the convergence analysis of optimization algorithms for strongly convex functions.

The inequalities (6) can be expressed more concisely if we introduce the so-called *Bregman divergence* of f, defined as the gap between f and its linear approximation:

$$D_f(y|x) = f(y) - (f(x) + \langle \nabla f(x), y - x \rangle).$$

The inequalities above can then be written as:

$$\frac{m}{2}\|y - x\|^2 \leq D_f(y|x) \leq \frac{L}{2}\|y - x\|^2.$$