2.1 A few concepts

For a differentiable function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$, the gradient of $f$ at a point $x \in \text{dom}f$ is the vector containing the partial derivatives of the function at that point, namely, $\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_n}(x))$.

For a twice differentiable function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^n$, the Hessian of $f$ at a point $x \in \text{dom}f$ is the matrix containing the second derivatives of the function at that point, namely, $\nabla^2 f(x)$ is the matrix with elements given by

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x), 1 \leq i, j \leq n$$

We say that a function $f$ is $c$-Lipschitz with respect to a norm $\| \cdot \|$ for $c \in \mathbb{R}^+$ if

$$|f(x) - f(y)| \leq c\|x - y\|, \forall x, y \in \text{dom}(f).$$

Claim 2.1 Let $f$ be a real-valued differentiable function. Then, $\|\nabla f(x)\| \leq c$ if and only if $f$ is $c$-Lipschitz.

Proof: the $\Rightarrow$ direction: Assume $\forall x \in \text{dom}(f), \|\nabla f(x)\| \leq c$. Then for $x, y \in \text{dom}(f)$, there exists $t \in [0, 1]$ such that

$$|f(x) - f(y)| = |\nabla f(y + t(x - y))^T(x - y)|$$

By the Schwarz’s inequality, the equation gives the estimate:

$$|f(x) - f(y)| \leq \|\nabla f(y + t(x - y))\| \|x - y\|$$

$$\leq c\|x - y\|$$

the $\Leftarrow$ direction: Assume $\forall x, y \in \text{dom}(f), f(x) - f(y) \leq c\|x - y\|$. Then the directional derivative of $f$ along $u$ is:

$$\nabla f(x)^T u = \lim_{\delta \to 0} \frac{f(x + \delta u) - f(x)}{\delta} \leq \lim_{\delta \to 0} \frac{c\|x + \delta u - x\|}{\delta} = c\|u\|$$

Set $u = \frac{(\nabla f(x))^T}{\|\nabla f(x)\|}$, then we have $\|\nabla f(x)\| \leq c$.

2.2 Convexity

Definition 2.2 (convex set) A set $U \subseteq \mathbb{R}^n$ is convex if for all $x, y \in U$ and all $\alpha$ in the interval $[0, 1]$, the point $\alpha x + (1 - \alpha)y$ also belongs to $U$.

Definition 2.3 (convex function) Let $X$ be a convex set in $\mathbb{R}^n$ and let $f : X \to \mathbb{R}$ be a function. We say that $f$ is convex if $\forall x, y \in X, \forall \alpha \in [0, 1] : f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$.

We say that $f$ is strictly convex if $\forall x \neq y \in X, \forall \alpha \in (0, 1) : f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$.
Here are some alternative characterizations of convexity:

- A function $f$ is convex if and only if it satisfies the Jensen\'s inequality everywhere: $\forall x \in \text{dom}(f), E(f(x)) \geq f(E(x))$.
- A differentiable function $f$ is convex if and only if $f(x + u) \geq f(x) + \nabla f(x)^T u$.
- A twice differentiable function $f$ is convex if and only if $\forall x \in \text{dom}(f), \nabla^2 f(x) \succeq 0$.

Here are some examples of convex functions:

- $f(x) = \|x\|^2$,
- $f(x) = x^T M x$, when $M$ is positive semidefinite,
- If $f(x, y)$ is convex in $x$ (e.g. $f(x, y) = \|x\|^2 - \|y\|^2$), then $g_1(x) = E_y f(x, y)$ and $g_2(x) = \sup_y f(x, y)$ are convex.

Definition 2.4 (strongly convex) A differentiable function $f$ is $c$-strongly convex with respect to a norm $\| \cdot \|$ if for all $x, u$ such that $x, x + u \in \text{dom} f$, the following inequality holds:

$$f(x + u) \geq f(x) + \nabla f(x)^T u + \frac{c}{2} \|u\|^2.$$

Definition 2.5 (strongly smooth) A differentiable function $f$ is $c$-strongly smooth with respect to a norm $\| \cdot \|$ if for all $x, u$ such that $x, x + u \in \text{dom} f$, the following inequality holds:

$$f(x + u) \leq f(x) + \nabla f(x)^T u + \frac{c}{2} \|u\|^2.$$

For example, $f(x) = \frac{1}{2} \|x\|^2$ is both 1-strongly convex and 1-strongly smooth.

Fact 2.6 When $f$ is twice differentiable, $f$ is $c$-strongly convex iff $\nabla^2 f(x) \succeq 0$.

2.3 Bregman divergence

Definition 2.7 (Bregman divergence) The Bregman divergence associated with $f$ is a function $D_f : \text{dom}(f) \times \text{dom}(f) \to \mathbb{R}$ defined by $D_f(x, y) = f(x) - f(y) - \nabla f(y)^T (x - y)$.

Here are some examples:

- $f(x) = \|x\|^2, D_f(x, y) = \|x - y\|^2$,
- $f(p) = \sum_{i=1}^n p_i \log p_i, D_f(p, q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i}$, which is the Kullback-Leibler divergence.

Here are some properties of Bregman Divergence:

- If $f$ is convex, $D_f(x, y) \geq 0$.
- $\forall x \in \text{dom}(f), D_f(x, x) = 0$.
- In general, $D_f(x, y) \neq D_f(y, x)$.

Fact 2.8 If $f$ is $c$-strongly convex, $D_f(x, y) \geq \frac{c}{2} \|x - y\|^2$. 
2.4 convex conjugate

Definition 2.9 (Fenchel conjugate) For a convex function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), its Fenchel conjugate is

\[
f^*(\theta) = \sup_{x \in \mathbb{R}^n} x^T \theta - f(x).
\]

For example, we have

- \( f(x) = \frac{1}{2} \|x\|^2 \), \( f^*(\theta) = \frac{1}{2} \|\theta\|^2 \).
- \( f(x) = \frac{1}{2} x^T M x \) and \( M \) is positive semidefinite, then \( f^*(\theta) = \frac{1}{2} \theta^T M^{-1} \theta \).

Fact 2.10 (biconjugate) Under a weak condition\(^1\), \( f = f^{**} \).

Fact 2.11 If \( f \) is differentiable and strongly convex, \( \forall x \in \text{dom}(f), \theta \in \text{dom}(f^*) \) we have \( \nabla f^*(\nabla f(x)) = x \) and \( \nabla f(\nabla f^*(\theta)) = \theta \).

Fact 2.12 If \( f \) is strictly convex and differentiable, \( D_f(x, y) = D_{f^*}(\nabla f(y), \nabla f(x)) \).

\(^1\) \( f \) is closed convex