Subgradient Methods

Lecturer: Jiaming Liang *January 23, 2024*

1 Convexity

1.1 Convex set

Definition 1. *A set $S \subseteq \mathbb{R}^n$ is called convex if for any $x, y \in S$ and $\lambda \in [0,1]$ it holds that $\lambda x + (1-\lambda)y \in S$.*

1.2 Convex function

Definition 2. *A proper extended real-valued function f is convex if and only if*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \quad \text{for all} \quad x, y \in \mathbb{R}^n, \lambda \in [0,1].$$

Definition 3. *A continuously differentiable function f is convex on $\mathbb{R}^n$ if for any $x, y \in \mathbb{R}^n$ we have*

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle.$$

Definition 4. *A twice continuously differentiable function f is convex on $\mathbb{R}^n$ if and only if for any $x \in \mathbb{R}^n$ we have*

$$\nabla^2 f(x) \succeq 0.$$

1.3 Strongly convex function

Definition 5. *A proper extended real-valued function f is μ-strongly convex if and only if*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) - \lambda(1-\lambda)\frac{\mu}{2}\|x-y\|^2 \quad \text{for all} \quad x, y \in \mathbb{R}^n, \lambda \in [0,1].$$

Definition 6. *A continuously differentiable function f is μ-strongly convex on $\mathbb{R}^n$ if for any $x, y \in \mathbb{R}^n$ we have*

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

Definition 7. *A twice continuously differentiable function f is μ-strongly convex on $\mathbb{R}^n$ if and only if for any $x \in \mathbb{R}^n$ we have*

$$\nabla^2 f(x) \succeq \mu I.$$

Lemma 1. *If a continuously differentiable function f is μ-strongly convex on $\mathbb{R}^n$, then we have* *(i) for all $x, y \in \mathbb{R}^n$,*

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

(ii) for all $x, y \in \mathbb{R}^n$,

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

(iii) for all $x, y \in \mathbb{R}^n$,

$$\mu \|x - y\| \leq \|\nabla f(x) - \nabla f(y)\|.$$

2 Big-O notation

Definition 8. *We say $f(x) = \mathcal{O}(g(x))$ if there exists scalars $M > 0$ and $x_0 \in \mathbb{R}$ such that*

$$|f(x)| \leq M g(x) \quad \text{for all} \quad x \geq x_0.$$

We say $f(x) = \Omega(g(x))$ if there exists scalars $M > 0$ and $x_0 \in \mathbb{R}$ such that

$$f(x) \geq M g(x) \quad \text{for all} \quad x \geq x_0.$$

We say $f(x) = \Theta(g(x))$ if $f(x) = \mathcal{O}(g(x))$ and $f(x) = \Omega(g(x))$.

3 Subgradient

Definition 9. *Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper function and let $x \in \text{dom}(f)$. A vector $g \in \mathbb{R}^n$ is called a subgradient of f at x if*

$$f(y) \geq f(x) + \langle g, y - x \rangle \quad \forall y \in \mathbb{R}^n.$$

We denote a subgradient of f at x by $f'(x)$.

Definition 10. *The set of all subgradients of f at x is called the subdifferential of f at x and is denoted by $\partial f(x)$:*

$$\partial f(x) \equiv \{g \in \mathbb{R}^n : f(y) \geq f(x) + \langle g, y - x \rangle \quad \forall y \in \mathbb{R}^n\}.$$

If f is convex, then $\partial f(x) \neq \emptyset$. If f is convex and smooth, then $\partial f(x) = \{\nabla f(x)\}$.

4 Optimization with set constraint

Let us consider now a convex smooth optimization problem with the *set constraint*:

$$\min_{x \in Q} f(x) \tag{1}$$

where Q is a closed convex set.

In the unconstrained case, the optimality condition is

$$\nabla f(x) = 0.$$

But this condition does not work with the set constraint. Consider the following univariate minimization problem:

$$\min_{x \geq 0} x.$$

Here $Q = \{x \in \mathbb{R} : x \geq 0\}$ and $f(x) = x$. Note that $x_* = 0$ but $f'(x_*) = 1 > 0$.

Theorem 1. *Let f be convex and differentiable and Q be closed and convex. A point x_* is as solution to* (1) *if and only if*

$$\langle \nabla f(x_*), x - x_* \rangle \geq 0 \tag{2}$$

for all $x \in Q$.

Proof. Indeed, if (2) is true, then

$$f(x) \geq f(x_*) + \langle \nabla f(x_*), x - x_* \rangle \geq f(x_*)$$

for all $x \in Q$. On the other hand, let x_* be a solution to (1). Assume that there exists some $x \in Q$ such that

$$\langle \nabla f(x_*), x - x_* \rangle < 0.$$

Consider the function

$$\phi(\alpha) = f(x_* + \alpha(x - x_*)), \quad \alpha \in [0, 1].$$

Note that

$$\phi(0) = f(x_*), \quad \phi'(0) = \langle \nabla f(x_*), x - x_* \rangle < 0.$$

Therefore, for α small enough we have

$$f(x_* + \alpha(x - x_*)) = \phi(\alpha) < \phi(0) = f(x_*).$$

This is a contradiction. □

The next statement is often addressed as the growth property of strongly convex functions.

Theorem 2. *If f is μ-strongly convex, then for any $x \in Q$, we have*

$$f(x) \geq f(x_*) + \frac{\mu}{2}\|x - x_*\|^2.$$

Proof. Indeed, by strong convexity and Theorem 1, we have

$$\begin{aligned} f(x) &\geq f(x_*) + \langle \nabla f(x_*), x - x_* \rangle + \frac{\mu}{2}\|x - x_*\|^2 \\ &\geq f(x_*) + \frac{\mu}{2}\|x - x_*\|^2. \end{aligned}$$

□

Theorem 3. *Let f be μ-strongly convex with $\mu > 0$ and the set Q is closed and convex. Then there exists a unique solution x_* to* (1).

Definition 11. *Let Q be a closed set and $x_0 \in \mathbb{R}^n$. Define*

$$\text{proj}_Q(x_0) = \arg\min_{x \in Q} \|x - x_0\|.$$

We call $\text{proj}_Q(x_0)$ the *Euclidean projection* of the point x_0 onto the set Q.

5 Subgradient methods

Consider the following optimization problem

$$\min_{x \in Q} f(x)$$

where f is convex and Q is a closed convex set. We also assume f is M-Lipschitz continuous over Q, i.e.,

$$|f(x) - f(y)| \leq M\|x - y\| \quad \forall x, y \in Q.$$

5.1 Convex function

Algorithm 1 Subradient method

Input: Initial point $x_0 \in Q$
for $k \geq 0$ **do**
  Step 1. Choose $\lambda_k > 0$.
  Step 2. Compute $x_{k+1} = \text{proj}_Q\left(x_k - \lambda_k f'(x_k)\right)$.
end for

Theorem 4. *Assume f is convex and M-Lipschitz continuous over a compact and convex set Q with diameter $D > 0$. (This means for evrey $x, y \in Q$, $\|x - y\| \le D$.) Consider* $\min_{x\in Q} f(x)$ *using the subgradient method with constant stepsize $\lambda_k \equiv \lambda > 0$. Then, we have*

$$\min_{0\le i\le k-1} f(x_i) - f_* \le \frac{D^2}{2\lambda k} + \frac{\lambda M^2}{2}.$$

Moreover, choosing $\lambda = \varepsilon/M^2$, we have

$$\min_{0\le i\le k-1} f(x_i) - f_* \le \frac{M^2 D^2}{2\varepsilon k} + \frac{\varepsilon}{2}.$$

Therefore, the complexity for an ε-solution (i.e., a point $x \in Q$ s.t. $f(x) - f_* \le \varepsilon$) is

$$\mathcal{O}\left(\frac{M^2 D^2}{\varepsilon^2}\right).$$

Proof. It is easy to see

$$x_{k+1} = \arg\min\left\{\ell_f(x; x_k) + \frac{1}{2\lambda}\|x - x_k\|^2 : x \in Q\right\},$$

where $\ell_f(x; y) = f(y) + \langle f'(y), x - y\rangle$. Using Theorem 2 and the fact that the above objective function is λ^{-1}-strongly convex, we have for every $x \in Q$,

$$\ell_f(x; x_k) + \frac{1}{2\lambda}\|x - x_k\|^2 \ge \ell_f(x_{k+1}; x_k) + \frac{1}{2\lambda}\|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda}\|x - x_{k+1}\|^2.$$

It follows from the convexity of f that

$$f(x) + \frac{1}{2\lambda}\|x - x_k\|^2 \ge \ell_f(x_{k+1}; x_k) + \frac{1}{2\lambda}\|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda}\|x - x_{k+1}\|^2.$$

Rearraging the terms and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&f(x_k) - f_* - \frac{1}{2\lambda}\|x_k - x_*\|^2 + \frac{1}{2\lambda}\|x_{k+1} - x_*\|^2 \\
\le&\langle f'(x_k), x_k - x_{k+1}\rangle - \frac{1}{2\lambda}\|x_k - x_{k+1}\|^2 \\
\le&\|f'(x_k)\|\|x_k - x_{k+1}\| - \frac{1}{2\lambda}\|x_k - x_{k+1}\|^2 \\
\le&\frac{\lambda}{2}\|f'(x_k)\|^2 \le \frac{\lambda M^2}{2},
\end{aligned}$$

where the last inequality is due to $x_k \in Q$ and the assumption that f is M-Lipschitz continuous over Q. Summing the above inequality, we have

$$k\left[\min_{0\le i\le k-1} f(x_i) - f_*\right] \le \sum_{i=0}^{k-1}[f(x_i) - f_*] \le \frac{\|x_0 - x_*\|^2}{2\lambda} + \frac{\lambda M^2 k}{2} \le \frac{D^2}{2\lambda} + \frac{\lambda M^2 k}{2}.$$

Thus,

$$\min_{0\le i\le k-1} f(x_i) - f_* \le \frac{D^2}{2\lambda k} + \frac{\lambda M^2}{2}.$$

□

5.2 Strongly convex function

Theorem 5. *Assume f is μ-strongly convex and M-Lipschitz continuous over a closed and convex set Q. Consider $\min_{x\in Q} f(x)$ using the subgradient method with variable stepsize*

$$\lambda_k = \frac{2}{\mu(k+1)} \quad k \geq 0.$$

Then, we have

$$\min_{1\leq i\leq k} f(x_i) - f_* + \frac{\mu}{2}\|x_{k+1} - x_*\|^2 \leq \frac{2M^2}{\mu(k+1)}.$$

Therefore, the complexity for an ε-solution is

$$\mathcal{O}\left(\frac{M^2}{\mu\varepsilon}\right).$$

Proof. Note that it still holds

$$\ell_f(x;x_k) + \frac{1}{2\lambda_k}\|x - x_k\|^2 \geq \ell_f(x_{k+1};x_k) + \frac{1}{2\lambda_k}\|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda_k}\|x - x_{k+1}\|^2.$$

It follows from the μ-strong convexity of f that for every $x \in Q$,

$$f(x) \geq \ell_f(x;x_k) + \frac{\mu}{2}\|x - x_k\|^2.$$

Combining the above two relations, we have

$$f(x) + \frac{1-\lambda_k\mu}{2\lambda_k}\|x - x_k\|^2 \geq \ell_f(x_{k+1};x_k) + \frac{1}{2\lambda_k}\|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda_k}\|x - x_{k+1}\|^2.$$

Proceeding similarly to the proof of Theorem 4, we have

$$f(x_k) - f_* \leq \frac{1-\lambda_k\mu}{2\lambda_k}\|x_k - x_*\|^2 - \frac{1}{2\lambda_k}\|x_{k+1} - x_*\|^2 + \frac{\lambda_k M^2}{2}.$$

Using the stepsize $\lambda_k = 2/[\mu(k+1)]$, we have

$$f(x_k) - f_* \leq \frac{\mu(k-1)}{4}\|x_k - x_*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x_*\|^2 + \frac{M^2}{\mu(k+1)}.$$

Multiplying by k, we get

$$k[f(x_k) - f_*] \leq \frac{\mu(k-1)k}{4}\|x_k - x_*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x_*\|^2 + \frac{M^2}{\mu}.$$

Summing the above inequality, we have

$$\frac{k(k+1)}{2}\left[\min_{1\leq i\leq k} f(x_i) - f_*\right] \leq \sum_{i=1}^{k} i[f(x_i) - f_*] \leq -\frac{\mu k(k+1)}{4}\|x_{k+1} - x_*\|^2 + \frac{kM^2}{\mu}$$

Thus, rearranging the terms and dividing the resulting inequality by $k(k+1)/2$, we obtain

$$\min_{1\leq i\leq k} f(x_i) - f_* + \frac{\mu}{2}\|x_{k+1} - x_*\|^2 \leq \frac{2M^2}{\mu(k+1)}.$$

□