Convex Optimisation

Matthew Tam
(University of Melbourne)
matthew.tam@unimelb.edu.au

ACE Network
S1 2021

---

* Assignment 3 online:

matthewktam.github.io/cvxopt

Due: May 13th @ 5pm

Last time: Frank-Wolfe alg.

convex, diff and ∇f was L-Lip.

$$\min_{x \in \mathbb{R}^n} f(x)$$

s.t. $x \in C \subseteq \mathbb{R}^n$

convex, compact.

Algorithm

Choose $x_0 \in C$.

For $k = 0, 1, 2, \ldots,$ do

$$s_k \in \operatorname*{argmin}_{s \in C} \langle \nabla f(x_k), s \rangle$$

$$x_{k+1} = (1 - \gamma_k) x_k + \gamma s_k$$

where $\gamma_k = \dfrac{2}{k+2}$.

3.2 Proximal gradient descent

In this section, we consider problems of the form:

$$\min_{x \in \mathbb{R}^n} f(x) + g(x) \qquad (3.4)$$

where:

- $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, convex with L-Lipschitz gradient.
- $g : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper, lsc, and convex.

Special case of (3.4):

- $g = 0 \implies \min_{x \in \mathbb{R}^n} f(x)$
- $g = i_C$ for $C \in \mathbb{R}^n$ a closed, convex set $\implies \min_{x \in C} f(x)$

Given an initial point $x_0 \in \mathbb{R}^n$, the <u>proximal gradient method</u> generates a sequence (x_k) according to

$$x_{k+1} = \operatorname{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)) \quad \forall k \in \mathbb{N}, \tag{3.5}$$

where $\lambda > 0$ is called stepsize.

<u>special cases of (3.5)</u>:

- $g = 0 \Rightarrow x_{k+1} = x_k - \lambda \nabla f(x_k)$
- $g = i_C \Rightarrow x_{k+1} = P_C(x_k - \lambda \nabla f(x_k))$
  "projected gradient algorithm".

Lemma 3.2.1

Let $g : \mathbb{R}^n \to (-\infty, +\infty]$ be proper, lsc, convex and $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable with L-Lipschitz gradient. Let $\lambda \in (0, \frac{1}{L}]$ and let (x_k) be given by (3.5). Then

$$(f+g)(x_{k+1}) - (f+g)(\overset{x_k}{z}) \qquad (3.6)$$

$$\le \underbrace{\frac{1}{2\lambda}\|x_k - \underset{x_k}{z}\|^2}_{=0} - \frac{1}{2\lambda}\|x_{k+1} - \underset{x_k}{z}\|^2 \quad \forall z \in \mathbb{R}^n$$

Proof.

Let $z \in \mathbb{R}^n$ be arbitrary. The prox-inequality (Th 2.3.4) to (3.5) gives

$$\lambda g(\cancel{x_{k+1}}) \le \lambda g(z) - \langle z - x_{k+1}, (x_k - \lambda \nabla f(x_k)) - x_{k+1} \rangle$$

$$= \lambda g(z) + \langle x_{k+1} - z, x_k - x_{k+1} \rangle$$

$$+ \lambda \langle \nabla f(x_k), z - x_{k+1} \rangle$$

$$= \cancel{\lambda g}(z) + \frac{1}{2}\left(\|x_k - z\|^2 - \|x_{k+1} - z\|^2 - \cancel{\|x_k - x_{k+1}\|^2} \right) + \lambda \langle \cancel{\nabla f(x_k)}, z - x_{k+1} \rangle .$$

The descent lemma applied to f and the inequality $\lambda \leq \frac{1}{L}$ gives

$$f(x_{k+1}) \leq f(z) + \langle \nabla f(x_k), x_{k+1} - z \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

$$\Rightarrow \lambda f(\cancel{x}_{k+1}) \leq \lambda f(z) + \lambda \langle \nabla f(x_k), x_{k+1} - z \rangle + \frac{\overbrace{L\lambda}^{\leq 1}}{2}\|x_{k+1} - x_k\|^2$$

$$\leq \cancel{\lambda f}(z) + \lambda \langle \cancel{\nabla f(x_k)}, x_{k+1} - z \rangle + \frac{1}{2}\|\cancel{x_{k+1}} - x_k\|^2 .$$

Combining these two inequalities gives

$$\lambda(f+g)(x_{k+1}) - \lambda(f+g)(z)$$
$$\leq \frac{1}{2}\left(\|x_k - z\|^2 - \|x_{k+1} - z\|^2\right),$$

from which (3.6) follows. $\boxtimes$

Theorem 3.2.2

Let $g: \mathbb{R}^n \to (-\infty, +\infty]$ be proper, lsc, convex, let $f: \mathbb{R}^n \to \mathbb{R}$ be convex and diff. with L-Lipschitz gradient, and suppose x^* is a solution of (3.4). Given an initial point $x_0 \in \mathbb{R}^n$ and $\lambda \in (0, \frac{1}{L}]$, let (x_k) be the sequence generated by (3.5).

Then

$$(f+g)(x_k) - (f+g)(x^*) \leq \frac{\|x_0 - x^*\|^2}{2\lambda k} \quad \forall k \in \mathbb{N}.$$

Proof.

Applying Lemma 3.2.1 with $z = x^*$ gives

$$(f+g)(x_{i+1}) - (f+g)(x^*) \le \frac{1}{2\lambda}\|x_i - x^*\|^2 - \frac{1}{2\lambda}\|x_{i+1} - x^*\|^2.$$

Consequently, we have

$$\begin{aligned}
\sum_{i=0}^{k-1} (f+g)(x_{i+1}) - k(f+g)(x^*) &\le \sum_{i=0}^{k-1}\left(\frac{1}{2\lambda}\|x_i - x^*\|^2 - \frac{1}{2\lambda}\|x_{i+1} - x^*\|^2\right) \qquad (3.8)\\
&= \frac{1}{2\lambda}\|x_0 - x^*\|^2 - \frac{1}{2\lambda}\|x_k - x^*\|^2 \\
&\le \frac{1}{2\lambda}\|x_0 - x^*\|^2.
\end{aligned}$$

Applying Lemma 3.2.1 with $z = x_k$ gives

$(f+g)(x_{k+1}) - (f+g)(x_k)$ ⊛

$$\leq -\frac{1}{2\lambda}\|x_{k+1} - x_k\|^2 \leq 0.$$

$\Rightarrow (f+g)(x_{k+1}) \leq (f+g)(x_k) \quad \forall k \in \mathbb{N}.$

* i.e. the sequence $((f+g)(x_k))$ is nonincreasing.

By combining with (3.8), we obtain

$$k(f+g)(x_k) - k(f+g)(x^*)$$

$$\leq \sum_{i=0}^{k-1} (f+g)(x_{i+1}) - k(f+g)(x^*)$$

$$\leq \frac{1}{2\lambda}\|x_0 - x^*\|^2$$

The result follows by dividing both sides by k. ☒

better

Next, we shown a ~~stronger~~ convergence result under stronger assumptions. For simplicity, we consider the case when $g=0$.

Definition 3.2.4

A function $f: \mathbb{R}^n \to (-\infty, +\infty]$ is μ-strongly convex if $\mu > 0$ and $f - \frac{\mu}{2}\|\cdot\|^2$ is convex.

Lemma 3.2.5

Let $f: \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then f is μ-strong convex iff

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

Proof

The function $g(x) := f(x) - \frac{\mu}{2}\|x\|^2$ is differentiable with $\nabla g(x) = \nabla f(x) - \mu x$.

Then we note that:

f is μ-strongly convex $\iff g$ is convex

$$\overset{\text{Prop. 2.2.3(a)}}{\iff} g(x) \geq g(y) + \langle \nabla g(y), x - y\rangle \quad \forall x, y \in \mathbb{R}^n.$$

The last inequality can be written as

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

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

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

which is the desired inequality. $\square$

If we combine the above lemma with the descent lemma, we get:

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

Thus, in this case, $\mu \leq L$.

Theorem 3.2.6

Let $f:\mathbb{R}^n \to \mathbb{R}$ be μ-strongly convex and differentiable with L-Lipschitz gradient, and suppose $x^* \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x)$.
Given an initial point $x_0 \in \mathbb{R}^n$ and $\lambda \in (0, \frac{1}{L}]$, let (x_k) be given by $x_{k+1} = x_k - \lambda \nabla f(x_k)$ $\forall k \in \mathbb{N}$,
then

$$f(x_k)-f(x^*) \leq (1-\lambda\mu)^k \cdot \frac{L\|x_0 - x^*\|^2}{2} \quad \forall k \in \mathbb{N}$$

Proof

By Lemma 3.25, we have.

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

$$\Leftrightarrow f(x^*) - f(x_k) - \frac{\mu}{2} \| x^* - x_k \|^2$$

$$\geq - \langle \nabla f(x_k), x_k - x^* \rangle. \quad (3.9)$$

By (*), we have

$$f(x_{k+1}) - f(x_k) \leq -\frac{1}{2\lambda} \| x_{k+1} - x_k \|^2 \quad (\text{by } (*))$$

$$= -\frac{1}{2\lambda} \| -\lambda \nabla f(x_k) \|^2$$

$$= -\frac{\lambda}{2} \| \nabla f(x_k) \|^2.$$

Since $f(x^*) \leq f(x_{k+1})$, this implies

$$f(x^*) \leq f(x_k) - \frac{\lambda}{2} \| \nabla f(x_k) \|^2 \quad (3.10).$$

With the help of (3.9) and (3.10), we obtain

$$
\begin{aligned}
&\|x_{k+1} - x^*\|^2 \\
&= \|x_k - \lambda \nabla f(x_k) - x^*\|^2 \\
&= \|(x_k - x^*) - \lambda \nabla f(x_k)\|^2 \\
&= \|x_k - x^*\|^2 - 2\lambda \langle x_k - x^*, \nabla f(x_k) \rangle + \lambda^2 \|\nabla f(x_k)\|^2 \\
&\le \|x_k - x^*\|^2 + 2\lambda \left(f(x^*) - f(x_k) - \frac{\mu}{2} \|x^* - x_k\|^2 \right) \\
&\qquad\qquad + 2\lambda \left(f(x_k) - f(x^*) \right) \\
&= (1 - \lambda\mu) \|x_k - x^*\|^2 \\
&\le (1 - \lambda\mu)^{k+1} \|x_0 - x^*\|^2 .
\end{aligned}
$$

Finally, by the descent lemma and Prop. 2.5.4 (i.e. $\nabla f(x^*) = 0$), we obtain

$$
\begin{aligned}
&f(x_{k+1}) - f(x^*) \\
&\le \langle \nabla f(x^*), x_{k+1} - x^* \rangle + \frac{L}{2} \|x_{k+1} - x^*\|^2
\end{aligned}
$$

$$= \underbrace{\cancel{\langle 0, x_{k+1} - x^* \rangle}}_{=0} + \frac{L}{2} \| x_{k+1} - x^* \|^2$$

$$\leq \frac{L}{2} \cdot (1 - \lambda \mu)^{k+1} \| x_0 - x^* \|^2 .$$

$\square$