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

Key results

- **Prop 2.2.3(a)**

Suppose $f : U \to \mathbb{R}$ is convex and differentiable (U open + convex).

Then:

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

$$\forall x, y \in U$$

- **Descent lemma**

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable with L-Lipschitz gradient.
Then

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

$$\forall x, y, z \in \mathbb{R}^n$$

3.1 Frank-Wolf Algorithm

Consider the minimisation problem

$$\min_{x \in \mathbb{R}^n} f(x) \text{ subject to } x \in C. \quad (3.1)$$

where:

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex, differentiable with L-Lipschitz gradient
- C is nonempty, convex, compact

The idea of FW algorithm is to solve a sequence of easier problem by replacing f (potentially nonlinear) in (3.1) with its linearisations.

Given the current iterate x_k, we can approximate f at s using its linearisation at x_k:

$$f(s) \approx \underbrace{f(x_k) + \langle \nabla f(x_k), s - x_k \rangle}_{\text{linearisation at } x_k \text{ eval. at } s}$$

$$= f(x_k) + \langle \nabla f(x_k), s \rangle - \langle \nabla f(x_k), x_k \rangle$$

Substituting this into (3.1) gives

$$\min_{s \in \mathbb{R}^n} \langle \nabla f(x_k), s \rangle \quad \text{st} \quad s \in C.$$

<u>Lemma 3.1.1</u> (extreme value thm)
Let $C \subseteq \mathbb{R}^n$ be nonempty, compact and let $f: \mathbb{R}^n \to \mathbb{R}$ be continuous. Then f attains its maximum and minimum on C. In other words, $\exists x, y \in C$

such that $f(x) = \inf_{z \in C} f(z)$

and that $f(y) = \sup_{z \in C} f(z)$.

Pf.

It suffices to show that f attains its minimum (for max replace f with $-f$).

To this end, consider a sequence $(x_k) \subseteq C$ such that $f(x_k) \to \inf_{z \in C} f(z)$ as $k \to \infty$. Since C is compact, the Bolzano-Weierstraß theorem yields the existence of a subsequence (x_{k_n}) such that $x_{k_n} \to x \in C$ as $n \to \infty$.

Then, since f is continuous, we have

$$\inf_{z \in C} f(z) = \lim_{n \to \infty} f(x_{k_n}) = f\left(\lim_{n \to \infty} x_{k_n}\right) = f(x),$$

which completes the proof. ∎

Theorem 3.1.2 (Frank-Wolfe algorithm)

Let $C \subseteq \mathbb{R}^n$ be a nonempty, compact, convex set, and let $f: \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable with L-Lipschitz gradient.

Given an intial point $x_0 \in C$, set

$$\begin{cases} s_k = \underset{s \in C}{\operatorname{argmin}} \langle \nabla f(x_k), s \rangle \\ x_{k+1} = (1-\gamma_k) x_k + \gamma_k s_k \end{cases} \tag{3.2}$$

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

where $\gamma_k = \dfrac{2}{k+2}$ for all $k \in \mathbb{N}$. as $k \to \infty$

Then

$$f(x_k) - \min_{\bar{x} \in C} f(\bar{x}) \leq \frac{2LD^2}{k+2} \quad \forall k \in \mathbb{N} \tag{3.3}$$

$\to 0$

where $D := \sup_{x,y \in C} \|x-y\| < +\infty$ denotes the diameter of the set C.

(3.3) $\Rightarrow$ $f(x_k) \to \min_{\bar{x} \in C} f(\bar{x})$

Pf.

First, we show that $(x_k) \subseteq C$.

By assumption, $x_0 \in C$. Next, assume $x_k \in C$. Then, since $s_k \in C$, convexity of C implies

$$x_{k+1} = (1-\gamma_k)\underbrace{x_k}_{\in C} + \underbrace{\gamma_k}_{\in (0,1)} \underbrace{s_k}_{\in C} \in C.$$

Hence, by mathematical induction, it follows that $(x_k) \subseteq C$.

Next, we show that D is finite. Since C is compact, it is bounded. That is, there exists $M > 0$ such that $\|x\| \le M$ for all $x \in C$. Thus, for all $x, y \in C$, the triangle inequality implies

$$\|x-y\| \leq \|x\| + \|y\| \leq 2M$$

By taking the supremum of both sides over $x, y \in C$, we deduce

$$D = \sup_{x,y\in C} \|x-y\| \leq \sup_{x,y\in C} (2M) = 2M < +\infty.$$

Let $\bar{x}$ be a solution of (3.1). This always exists, by Lemma 3.1.1, since f is continuous and C is compact.

By the descent lemma,

$$f(x_{n+1}) \leq f(x_n) + \langle \nabla f(x_n), \underbrace{x_{n+1} - x_n}_{= \gamma_n (s_n - x_n)} \rangle + \frac{L}{2} \|x_{n+1} - x_n\|^2$$

$$= f(x_n) + \gamma_n \boxed{\langle \nabla f(x_n), s_n - x_n \rangle} + \frac{L\gamma_n^2}{2} \|s_n - x_n\|^2$$

$$= f(x_k) + \gamma_k \min_{s \in C} \langle \nabla f(x_k), s - x_k \rangle + \frac{L\gamma_k^2}{2} \|s_k - x_k\|^2$$

$$\leq f(x_k) + \gamma_k \underbrace{\boxed{\langle \nabla f(x_k), \bar{x} - x_k \rangle}}_{\text{Prop. 2.2.3(a)}} + \frac{L\gamma_k^2}{2} D^2$$

$$\leq f(x_k) + \gamma_k \left(f(\bar{x}) - f(x_k) \right) + \frac{L\gamma_k^2}{2} D^2$$

Thus altogether we have.

$$f(x_{k+1}) - f(\bar{x}) \leq (1-\gamma_k)\left(f(x_k) - f(\bar{x})\right) + \frac{L\gamma_k^2}{2} D^2 \qquad (*)$$

We now proceed to show (3.3) by induction. First note, $\gamma_0 = 1 \Rightarrow \frac{\gamma_0^2}{2} = \frac{1}{2}$. Thus setting $k=0$ in $(*)$ gives.

$$f(x_1) - f(\bar{x}) \leq 0 \cdot (f(x_0) - f(\bar{x})) + \frac{LD^2}{2}$$
$$= \frac{LD^2}{2}.$$

as claimed. Next suppose (3.3) hold

for index k. Then, ⊛ gives

$$f(x_{k+1}) - f(\bar{x})$$

$$\le (1-\gamma_k)\left(\boxed{f(x_k) - f(\bar{x})}\right) + \frac{L\gamma_k^2}{2} D$$

use (3.3)

$$\le (1-\gamma_k)\cdot\left(\frac{2LD^2}{k+2}\right) + \frac{\gamma_k^2 LD^2}{2}$$

$$\le \frac{2LD^2}{k+3}\cdot\left(\frac{(1-\gamma_k)(k+3)}{k+2} + \frac{\gamma_k^2(k+3)}{4}\right) \quad (**)$$

So, we just need to show that the orange term is ≤ 1. To this end, we observe

$$(**) = \frac{k}{k+2}\cdot\frac{(k+3)}{(k+2)} + \frac{(k+3)}{(k+2)^2}$$

$$= \frac{(k+3)(k+1)}{(k+2)^2}$$

$$= \frac{k^2+4k+3}{k^2+4k+4} \le 1.$$

Thus,

$$f(x_{k+1}) - f(\bar{x}) \le \frac{2LD^2}{k+3},$$

The result then follows by induction. $\boxtimes$

Remark 3.1.7

Suppose the constraint set C consists of linear equality and inequality constraints. By rewriting equality constraints as a pair of inequality constraints, this expressed as

$$C = \{x \in \mathbb{R}^n : Ax \le b\}.$$

Then the subproblem in the Frank-Wolfe algorithm becomes

$$\min_{s \in \mathbb{R}^n} \langle \nabla f(x_k), s \rangle$$

$$\text{st} \quad As \leq b.$$

which is a linear programme (LP).

Thus, the Frank-Wolfe method allows one to solve a linearly constrained nonlinear minimisation problem using only a linear optimisation solver. (eg. simplex).