Convex Optimisation

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

ACE Network
S1 2021

Assignment 2 online:

matthewktam.github.io/cvxopt

Due: April 15th @ 5pm

Last time

Prop 2.2.3(a)

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

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

Prop 2.2.5 (Descent lemma)

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function with L-Lipschitz continuous gradient. Then

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

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

Facts:

- For any matrix A, we have

$$\|Ax\| \leq \|A\|_2 \|x\| \quad \forall x \in \mathbb{R}^n$$

where the operator norm of A is given by

$$\|A\|_2 := \sup_{\|v\|=1} \|Av\| = \sup_{v \in \mathbb{R}^n} \frac{\|Av\|}{\|v\|}$$

- If f is twice differentiable, then

$$(*) \quad \nabla^2 f(x) v = \lim_{t \to 0} \frac{\nabla f(x+tv) - \nabla f(x)}{t}.$$

$$f''(x) v = \lim_{t \to 0} \frac{f'(x+tv) - f'(x)}{t}$$

Prop 2.2.7

Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. If there exists $L \geq 0$ such $\|\nabla^2 f(x)\|_2 \leq L$ $\forall x \in \mathbb{R}^n$, then ∇f is L-Lipschitz continuous.

Proof

Let $x, y, a \in \mathbb{R}^n$ and define

$$h_a(t) = \langle a, \nabla f(x + t(y-x)) \rangle$$

By the mean value theorem, there exists $t_a \in (0,1)$ such that

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

$$= \frac{h_a(1) - h_a(0)}{1 - 0}$$

$\langle x, Ay \rangle = \langle A^T x, y \rangle$

$$= h_a'(t_a)$$

$$\overset{(*)}{=} \langle a, \nabla^2 f(x + t_a(y-x))(y-x) \rangle$$

$$= \langle \nabla^2 f(\overset{z_a}{\cancel{x + t_a(y-x)}})a, y - x \rangle$$

where $z_a = x + t_a(y-x)$ and the last equality hold because $\nabla^2 f(z_x)$ is symmetric.

Next, we compute

$$\|\nabla f(y) - \nabla f(x)\|$$

$$= \left\langle \frac{\nabla f(y) - \nabla f(x)}{\|\nabla f(y) - \nabla f(x)\|}, \nabla f(y) - \nabla f(x) \right\rangle$$

$$\le \sup_{\|a\|=1} \langle a, \nabla f(y) - \nabla f(x) \rangle$$

$$= \sup_{\|a\|=1} \langle \nabla^2 f(z_a) a, y - x \rangle$$

$$\le \sup_{\|a\|=1} \|\nabla^2 f(z_a) a\| \, \|y - x\|$$

$$\le \sup_{\|a\|=1} \underbrace{\|\nabla^2 f(z_a)\|_2}_{\le L} \underbrace{\|a\|}_{=1} \, \|y - x\|$$

$$\le L \|y - x\|,$$

which completes the proof. $\boxed{x}$

2.3 Proximity operator

Consider $f: \mathbb{R}^n \to (-\infty, +\infty]$.

The proximity operator corresponding to f is the operator $\text{prox}_f: \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\text{prox}_f(x) = \underset{y \in \mathbb{R}^n}{\text{argmin}} \left\{ f(y) + \frac{1}{2}\|x-y\|^2 \right\}$$

Our goal: show this is well-defined for reasonable functions.

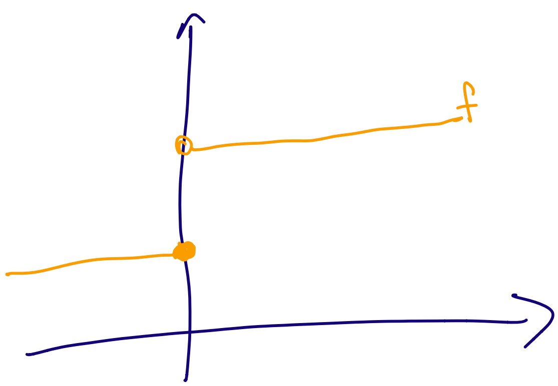
Definition 2.3.1

A function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is lower semicontinuous (lsc) if

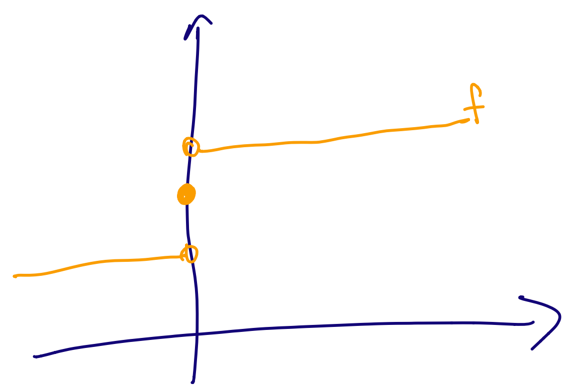
$$\liminf_{z \to x} f(z) \geq f(x) \qquad \forall x \in \mathbb{R}^n$$

f cont would mean

$$\lim_{z \to x} f(z) = f(x)$$



lsc



not lsc.

Exercises

- f is lsc $\iff$ epi f is closed
- ι_C is lsc $\iff$ C is closed

Recall that $f : \mathbb{R}^n \to (-\infty, +\infty]$ is proper if $f \not\equiv +\infty$.

Theorem 2.3.4

Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be proper, lsc and convex. Then prox_f is a well-defined operator. Moreover, $p = \operatorname{prox}_f(x)$ if and only if

$$f(p) + \langle y - p, x - p \rangle \leq f(y) \quad \forall y \in \mathbb{R}^n. \tag{2.3}$$

Proof

(Existence) Let $(p_k) \subseteq \mathbb{R}^n$ be a sequence such that

$$f(p_k) + \frac{1}{2}\|x - p_k\|^2 \to d := \inf_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2}\|x - y\|^2 \right\}$$

[$d > -\infty$, exercise or assignment]

By convexity of f, we have

$$\begin{aligned} f(p_n) + f(p_m) &= \frac{1}{2}(2f)(p_n) + \frac{1}{2}(2f)(p_m) \qquad (2.4) \\ &\geq 2f\left(\frac{p_n + p_m}{2}\right). \end{aligned}$$

Parallelogram law

$$2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2 \quad \forall x, y \in \mathbb{R}^n$$

By the parallelogram law, we have

$$
\begin{aligned}
&2\|p_k - x\|^2 + 2\|x - p_m\|^2 \\
&= \|p_k - p_m\|^2 + \|p_k + p_m - 2x\|^2 \\
&= \|p_k - p_m\|^2 + 4\left\|\frac{p_k + p_m}{2} - x\right\|^2
\end{aligned}
$$

which implies

$$
\frac{1}{2}\|p_k - x\|^2 + \frac{1}{2}\|x - p_m\|^2 = \frac{1}{4}\|p_k - p_m\|^2 + \left\|\frac{p_k + p_m}{2} - x\right\|^2 \qquad (2.5)
$$

By adding (2.4) to (2.5) we obtain

$$
\left[f(p_k) + \frac{1}{2}\|p_k - x\|^2\right] + \left[f(p_m) + \frac{1}{2}\|p_m - x\|^2\right]
\geq \frac{1}{4}\|p_k - p_m\|^2 + 2\left[f\left(\frac{p_k + p_m}{2}\right) + \frac{1}{2}\left\|\frac{p_k + p_m}{2} - x\right\|^2\right]
$$

(2.6)

(2.6)

$$\geq \frac{1}{4}\|p_k - p_m\|^2 + 2d.$$

Taking the limit as $k, m \to \infty$ in (2.6) gives

$$\cancel{2d} \geq \frac{1}{4}\lim_{k,m\to\infty}\|p_k - p_m\|^2 \cancel{+ 2d}$$

$$\Rightarrow \lim_{k,m\to\infty}\|p_k - p_m\|^2 = 0$$

i.e. (p_k) is Cauchy sequence, and hence it converges to a point $p \in \mathbb{R}^n$.

Using lower semicontinuity of f, we deduce

$$d \leq f(p) + \frac{1}{2}\|x - p\|^2$$

$$\leq \liminf_{k\to\infty}\left(f(p_k) + \frac{1}{2}\|x - p_k\|^2\right) \quad \text{conts}$$

$$= \lim_{k\to\infty}\left(f(p_k) + \frac{1}{2}\|x - p_k\|^2\right) = d$$

$$\Rightarrow \quad d = f(p) + \tfrac{1}{2}\|x-p\|^2.$$

i.e. $$p \in \operatorname*{argmin}_{y\in\mathbb{R}^n} f(y) + \tfrac{1}{2}\|x-y\|^2.$$

(Uniqueness) The function $y \mapsto f(y) + \frac{1}{2}\|x-y\|^2$ is strictly convex. (exercise). Hence, by Prop 2.1.6, it has a unique minimiser.

(characterisation)

Let $p = \mathrm{prox}_f(x)$ and denote $p_\lambda = \lambda y + (1-\lambda)p = p + \lambda(y-p)$ for $y \in \mathbb{R}^n$ and $\lambda \in (0,1)$. Then convexity of f gives

$$f(p) + \tfrac{1}{2}\|x-p\|^2$$

$$\leq f(p_\lambda) + \tfrac{1}{2}\|x - p_\lambda\|^2 \qquad (p = \mathrm{prox}_f(x))$$

$$\le \lambda f(y) + (1-\lambda) f(p) + \frac{1}{2}\|(x-p) - \lambda(y-p)\|^2$$

$$= \lambda f(y) + \cancel{(1-\lambda)} f(p)$$

$$+ \cancel{\frac{1}{2}\|x-p\|^2} - \lambda\langle \cancel{x-p}, y-p\rangle + \frac{\lambda^2}{2}\|y-p\|^2$$

Rearranging gives.

$$\lambda f(p) + \lambda\langle x-p, y-p\rangle$$

$$\le \lambda f(y) + \frac{\lambda^2}{2}\|y-p\|^2$$

Divide by $\lambda > 0$ and taking the limit as $\lambda \to 0$ gives

$$f(p) + \langle x-p, y-p\rangle \le f(y) + \cancelto{0}{\frac{\lambda}{2}\|y-p\|^2}$$

which is precisely (2.3).

Conversely, suppose (2.3) holds.

Expanding the norm squared gives

$\|x-y\|^2$

$$= \|(y-p)-(x-p)\|^2$$

$$= \|y-p\|^2 - 2\langle y-p, x-p\rangle + \|x-p\|^2.$$

Substituting this into (2.3) gives

$$f(p) + \frac{1}{2}\|x-p\|^2$$

$$\leq f(y) + \frac{1}{2}\|x-y\|^2 - \frac{1}{2}\|y-p\|^2$$

$$\leq f(y) + \frac{1}{2}\|x-y\|^2$$

for all $y \in \mathbb{R}^n$. Thus $p = \mathrm{prox}_f(x)$ which completes the proof. $\square$

Example 2.3.5 (soft-thresholding)

Let $f(x) = \|x\|_1 = \sum_{i=1}^{n} |x_i|$ and $\lambda > 0$.

Then $p = \mathrm{prox}_{\lambda f}(x)$ is given by

$$p_i = \begin{cases} 0 & |x_i| \le \lambda \\ x_i - \lambda \,\mathrm{sign}(x_i) & |x_i| > \lambda \end{cases}$$

where

$$\mathrm{sign}(\alpha) = \begin{cases} +1 & \alpha > 0 \\ 0 & \alpha = 0 \\ -1 & \alpha < 0 \end{cases}$$

Definition

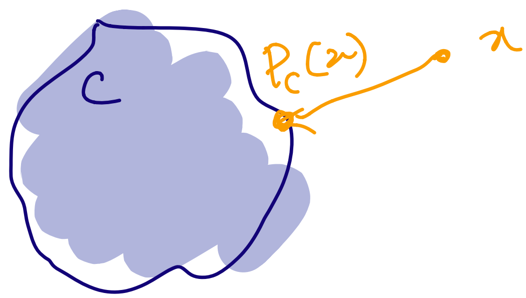
Given a set $C \subseteq \mathbb{R}^n$, the projector onto C is the operator $P_C : \mathbb{R}^n \to C$ given by

$$P_C(x) = \underset{c \in C}{\mathrm{argmin}} \|x - c\|.$$

$$= \mathrm{argmin}\left\{ i_C(c) + \frac{1}{2}\|x - c\|^2 \right\}$$

$C \in \mathbb{R}^n$

$= \text{prox}_{i_C}(x)$



Corollary 2.3.6

Let $C \subseteq \mathbb{R}^n$ be nonempty, closed, and convex. Then P_C is well-defined. Moreover, $p = P_C(x)$ if and only if

$$p \in C \text{ and } \langle x-p, c-p \rangle \le 0 \quad \forall c \in C$$

Proof.

Since i_C is proper, lsc and convex, the result follows by apply Theorem 2.3.4.

$\boxtimes$