Convex Optimisation

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

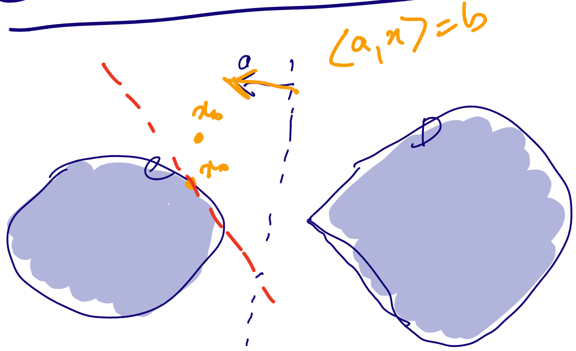
ACE Network
S1 2021

---

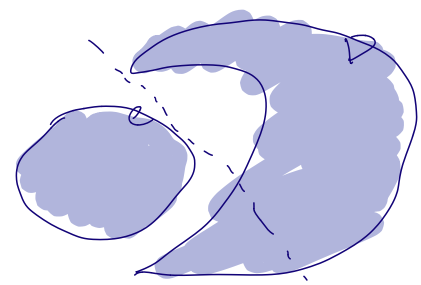
* Assignment 3 online after lecture:
matthewktam.github.io/cvxopt

Due: May 13th @ 5pm

Last time:



Two convex sets that can be separated.



Two sets that cannot be separated by a hyperplane.

Theorem 2.4.1 (supporting hyperplane theorem)

Let $C \subseteq \mathbb{R}^n$ be a nonempty, convex set and let $x_0 \in \mathbb{R}^n$ such that $x_0 \in \operatorname{bdry} C$ or $x_0 \notin C$. Then there exists $a \in \mathbb{R}^n \setminus \{0\}$ such that

$$\langle a, z\rangle \leq \langle a, x_0\rangle \quad \forall z \in C$$

Corollary 2.4.2 (separating hyperplane th).

Let $C, D \subseteq \mathbb{R}^n$ be nonempty, convex sets with $C \cap D = \emptyset$. Then there exists $a \in \mathbb{R}^n \setminus \{0\}$ such

$$\langle a, x\rangle \leq \langle a, y\rangle \quad \forall x \in C, y \in D.$$

2.5 subgradients of cvx functions

Definition 2.5.1

Let $f: \mathbb{R}^n \to (-\infty, +\infty]$. The subdifferential of f at $\bar{x} \in \operatorname{dom} f$ is defined by

$$\partial f(\bar{x}) := \left\{ \phi \in \mathbb{R}^n : \langle \phi, y - \bar{x} \rangle \le f(y) - f(\bar{x}) \quad \forall y \in \mathbb{R}^n \right\}$$

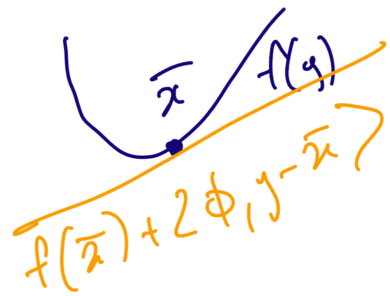
partial

An element $\phi \in \partial f(\bar{x})$ is said to be a subgradient of f at $\bar{x}$.

When $\bar{x} \notin \operatorname{dom} f$, $\partial f(\bar{x}) := \emptyset$.

Intuition: If $\phi \in \partial f(\bar{x})$, then

$$\underbrace{f(\bar{x}) + \langle \phi, y - \bar{x} \rangle}_{\text{"linearisation of } f \text{ at } \bar{x} \text{ evaluated at } y\text{".}} \leq f(y) \quad \forall y \in \mathbb{R}^n$$



Special case: If f is diff and $\phi = \nabla f(\bar{x})$, then this becomes

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

$\uparrow$ Proposition 2.2.3.

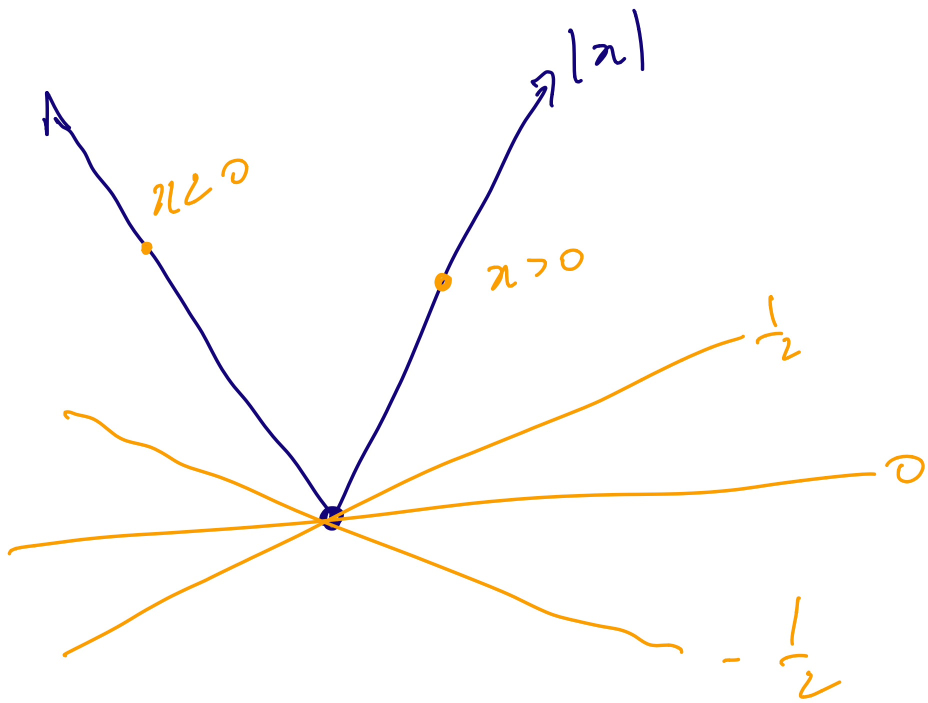
For diff, convex f, we always have

$$\nabla f(x) \in \partial f(x).$$

Example 2.5.2

Let $f(x) = |x|$. Then

$$\partial f(x) = \begin{cases} \{+1\} & \text{if } x > 0 \\ [-1, +1] & \text{if } x = 0 \\ \{-1\} & \text{if } x < 0. \end{cases}$$



Proposition 2.5.3

Suppose $f: \mathbb{R}^n \to (-\infty, +\infty]$ is differentiable at $\bar{x} \in \mathbb{R}^n$. Then
$\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.

Proof. By Prop 2.2.3, $\nabla f(\bar{x}) \in \partial f(\bar{x})$, which establishes one inclusion. For the other inclusion, let $\phi \in \partial f(\bar{x})$. By the definition of the subdifferential, we have

$$\langle \phi, y - \bar{x} \rangle \leq f(y) - f(\bar{x}) \quad \forall y \in \mathbb{R}^n.$$

In particular, setting $y = \bar{x} + td$ for $d \in \mathbb{R}^n$ and $t > 0$ gives

$$\langle \phi, d \rangle \leq \frac{f(\bar{x} + td) - f(\bar{x})}{t}$$

Taking the limit as $t \to 0$ gives

$$\langle \phi, d \rangle \le \lim_{t \to 0} \frac{f(\bar{x}+td) - f(\bar{x})}{t}$$

$$= \langle \nabla f(\bar{x}), d \rangle$$

$$\Rightarrow \langle \phi - \nabla f(\bar{x}), d \rangle \le 0.$$

In particular, setting $d = \phi - \nabla f(\bar{x})$ gives

$$\| \phi - \nabla f(\bar{x}) \|^2 \le 0$$

$$\Rightarrow \phi = \nabla f(\bar{x}).$$

Thus $\partial f(\bar{x}) \subseteq \{\nabla f(\bar{x})\}$ and the proof is complete.

$\square$

Proprosition 2.5.4

Let $f:\mathbb{R}^n \to (-\infty, +\infty]$ be a covex function. Then the following are equivalent.

(a) f has a local minimum at $\bar{x}$.

i.e. $\exists \delta > 0$ st $f(\bar{x}) \le f(x)$ $\forall x \in B_\delta(\bar{x})$.

(b) f has a global minimum at $\bar{x}$

i.e. $f(\bar{x}) \le f(x)$ $\forall x \in \mathbb{R}^n$

(c) $0 \in \partial f(\bar{x})$.

Proof Exercise in assignment 3.

Proprosition 2.55

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

If $\bar{x} \in \mathrm{int}(\mathrm{dom} f)$, then $\partial f(\bar{x}) \neq \phi$.

Proof

Let $\bar{x} \in \mathrm{int}(\mathrm{dom} f)$.

Then $(\bar{x}, f(\bar{x})) \in \mathrm{bdry}(\mathrm{epi}\, f) \subseteq \mathbb{R}^n \times \mathbb{R}$.

Thus, since epi f is convex, the supporting hyperplane theorem (Th 2.4.1) implies the existence of a non-zero vector $(\phi, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that, for all $(x, \alpha) \in \mathrm{epi}\, f$, we have

$$\begin{aligned} \langle \phi, x \rangle + \beta\alpha &= \langle (\phi, \beta), (x, \alpha) \rangle \\ &\leq \langle (\phi, \beta), (\bar{x}, f(\bar{x})) \rangle \end{aligned}$$

$$= \langle \phi, \bar{x} \rangle + \beta f(\bar{x}).$$

In other words,

$$\left\langle \frac{\phi}{-\beta}, x \right\rangle \overset{-f(x)}{\cancel{+\beta\alpha}} \le \left\langle \frac{\phi}{-\beta}, \bar{x} \right\rangle \overset{-}{\cancel{+}} \beta f(\bar{x}) \qquad (*).$$

for all $x \in \operatorname{dom} f$,

$f(x) \le \alpha$.

Since α can be arbitrary large, it follows that $\beta \le 0$. We claim that $\beta < 0$. For if $\beta = 0$, take $\lambda > 0$ such that

$$x = \bar{x} + \lambda \phi \in \operatorname{dom} f$$

(because $\bar{x}$ is an interior pt). Hence, $(*)$ implies

$$\langle \phi, \bar{x} \rangle + \lambda \|\phi\|^2 = \langle \phi, \bar{x} + \lambda \phi \rangle$$

$$\le \langle \phi, \bar{x} \rangle$$

$$\Rightarrow \lambda \|\phi\|^2 \le 0 \qquad \Rightarrow \phi = 0.$$

This contradicts the fact that (ϕ, β) is non-zero.

Thus, dividing ⊛ by $-\beta > 0$ and setting $\alpha = f(x)$, and rearranging gives

$$\left\langle \frac{\phi}{-\beta}, x - \bar{x} \right\rangle \leq f(x) - f(\bar{x}) \quad \forall x \in \operatorname{dom} f.$$

Thus, $\phi' := \frac{\phi}{-\beta} \in \partial f(\bar{x})$ and the proof is complete. $\boxed{}$

Differentiation is linear:

$$\nabla(f + \alpha g) = \nabla f + \alpha \nabla g$$

This does always hold for subdifferentiation

The hypograph of a function h is the set

$$\text{hyp}(h) = \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} : h(x) \geq \alpha\}$$

This conex iff h is concave iff $-h$ convex.

Lemma 2.5.6

Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ and $v \in \mathbb{R}^n$.

Set $h(x) := f(x) + \langle v, x \rangle$. Then

$\partial h(x) = \partial f(x) + v$ for all $x \in \mathbb{R}^n$.

Proof: exercise.

Theorem 2.5.7 (sum rule)

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

(2.8) $\partial(f+g)(\bar{x}) \supseteq \partial f(\bar{x}) + \partial g(\bar{x}) \quad \forall \bar{x} \in \mathbb{R}^n$,

and equality holds if f and g are convex, and $\operatorname{dom} g \cap \operatorname{int}(\operatorname{dom} f) \neq \phi$.

Proof

The inclusion (2.8) follows from the definition of the subdifferentials, so need only show the equality.

To this end, let $\phi \in \partial(f+g)(\bar{x})$ and consider

$$h(x) := f(x) + g(x) - \langle \phi, x \rangle. \quad \geq ? \; \forall x$$

By Lemma 2.5.3, we have

$$\partial h(\bar{x}) = \partial(f+g)(\bar{x}) - \phi \ni 0$$

Hence, Proposition 2.5.4 implies that $\bar{x}$ is a minimum of h. Since adding a constant to a function does not change its subdifferential, we assume WLOG that $h(\bar{x}) = 0$ which implies that

$$f \geq -g + \langle \phi, \cdot \rangle \qquad (2.9)$$

Thus $\operatorname{epi} f$ and $\operatorname{hyp}(-g + \langle \phi, \cdot \rangle)$ are disjoint convex sets (except for maybe the boundary). We can then apply the separating hyperplane theorem to deduce the existence of a non-zero vector $(v, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\langle v, x \rangle + \beta \alpha \leq \langle v, y \rangle + \beta \gamma \qquad (2.10)$$

if $f(x) \leq \alpha$
and $-g(y) + \langle \phi, y \rangle \geq \gamma$

Since α can be arbitrarily large, it follows that $\beta \leq 0$. Further, we claim that $\beta < 0$. For if $\beta = 0$, then $\exists\, \bar{x} \in \operatorname{dom} g \cap \operatorname{int} \operatorname{dom} f$ and $\lambda > 0$ such $x = \bar{x} + \lambda v \in \operatorname{dom} f$ ($\bar{x}$ int pt).
Putting this into (2.10) gives

$$\langle v, \bar{x} + \lambda v \rangle \leq \langle v, \bar{x} \rangle$$

$$\Rightarrow \lambda \|v\|^2 \leq 0$$

$$\Rightarrow v = 0$$

This contradicts the fact that (v, β) is non-zero.

Thus, dividing (2.10) by $-\beta > 0$ and setting $v' = v/-\beta$, $\alpha = f(x)$, $y = \bar{x}$ and $\gamma = -g(\bar{x}) + \langle \phi, \bar{x} \rangle$ gives

$$\langle v', x \rangle - g(x) + \langle \phi, x \rangle$$

Note: There is a typo here that was since been corrected in the lecture notes.

(2.10)

$$\le \langle v', x\rangle - f(x)$$

$$\le \langle v', \bar{x}\rangle + g(\bar{x}) - \langle \phi, \bar{x}\rangle$$

$h(\bar{x}) = 0.$

$$= \langle v', \bar{x}\rangle - f(\bar{x})$$

$$\langle -v', x - \bar{x}\rangle \le f(x) - f(\bar{x})$$

$$\Rightarrow -v' \in \partial f(\bar{x}).$$

$$\langle v' + \phi, x - \bar{x}\rangle \le g(x) - g(\bar{x}).$$

$$\Rightarrow v' + \phi \in \partial g(\bar{x}).$$

---

hence $\phi = -v' + (v' + \phi) \in \partial f(\bar{x}) + \partial g(\bar{x})$.

This completes the proof. $\boxtimes$

Example 2.5.8

Let $C = \{x \in \mathbb{R}^2 : \|x-(1,0)\| \le 1\}$
and $D = \{x \in \mathbb{R}^2 : \|x+(1,0)\| \le 1\}$.

Then

$$\partial(i_C + i_D)(0,0) = \mathbb{R}^2$$

$$\partial i_C(0,0) + \partial i_D(0,0) = \mathbb{R} \times \{0\}$$

Theorem 2.5.9

Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be linear and let $g : \mathbb{R}^m \to (-\infty, +\infty]$. Then

$$A^T \partial g(Ax) \supseteq \partial(g \circ A)(x) \quad \forall x \in \mathbb{R}^n,$$

and equality holds if g is convex and $\operatorname{int}(\operatorname{dom} g) \cap \operatorname{range} A \neq \emptyset$.