Convex Optimisation

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

ACE Network
S1 2021

Course webpage:

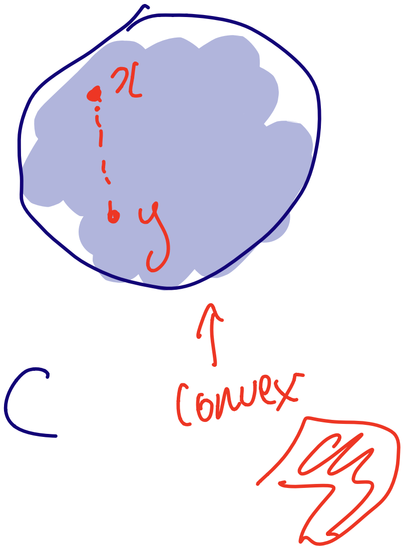
matthewktam.github.io/cvxopt

Last time

- convex set:

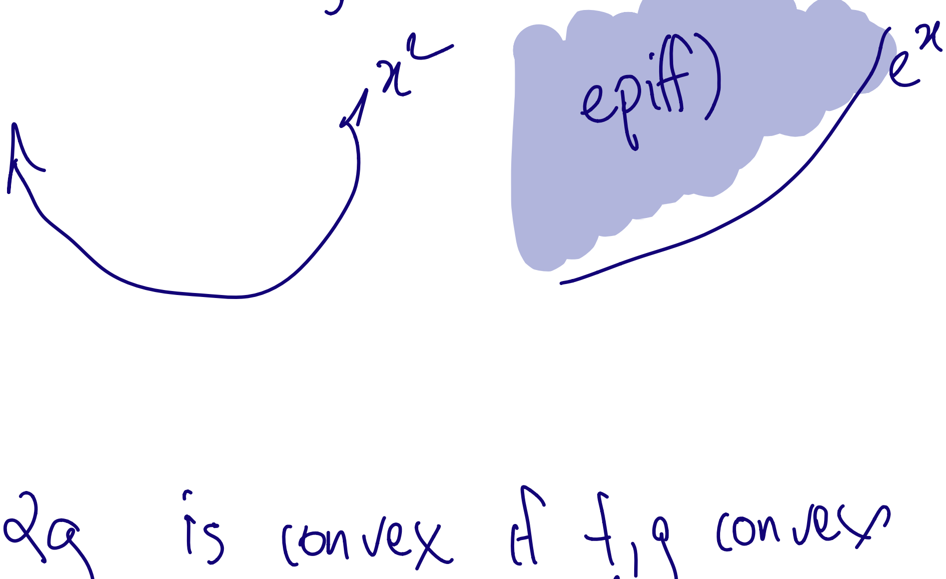
$x, y \in C$ and $\lambda \in [0,1]$

$\Rightarrow \lambda x + (1-\lambda) y \in C$



- convex function

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$$

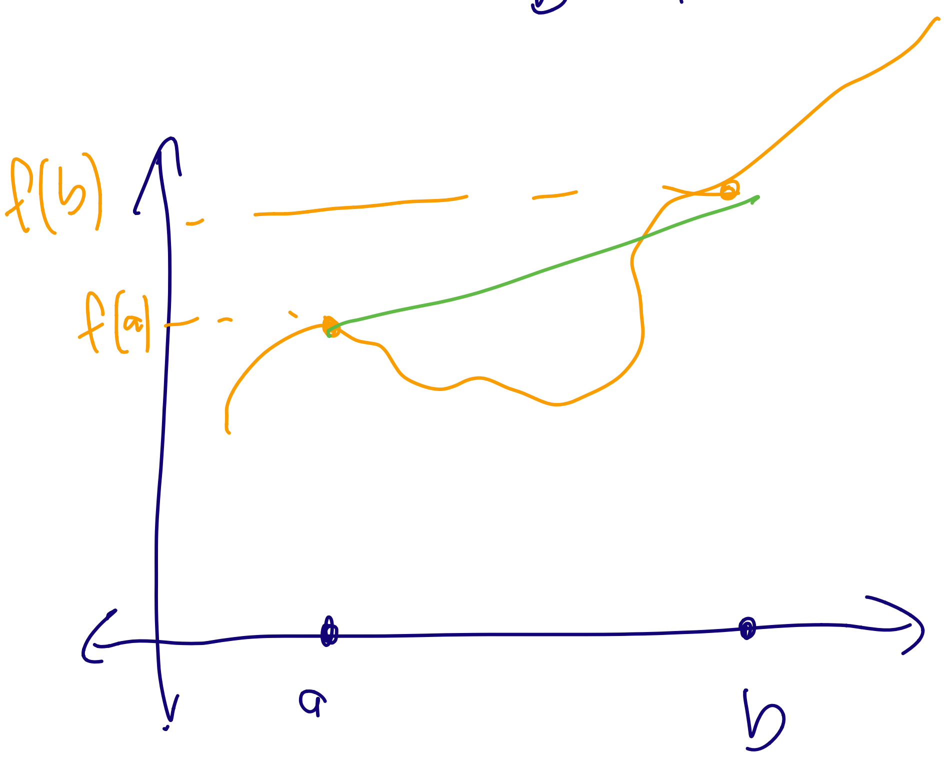


$f + \lambda g$ is convex if f, g convex and $\lambda \geq 0$.

Mean value theorem

Let $f : [a,b] \to \mathbb{R}$ be a continuous function which is differentiable on (a,b). Then $\exists c \in (a,b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b-a}$$



2.2 Differentiability of convex functions.

Lemma 2.2.1

Let $f: I \to \mathbb{R}$ where $I \subseteq \mathbb{R}$ is an interval. Then:

(a) Suppose f is differentiable. Then f is convex iff f' is nondecreasing.

(b) Suppose f is twice differentiable. Then f is convex iff $f''(x) \geq 0 \quad \forall x \in I$.

Pf (a) Let $x, y \in I$ and $\lambda \in (0,1)$.
Since f is convex, we have

$$f(y)+\lambda\big(f(x)-f(y)\big) = \lambda f(x)+(1-\lambda)f(y)$$
$$\geq f\big(\lambda x+(1-\lambda)y\big)$$
$$= f\big(y+\lambda(x-y)\big)$$

Rearranging and taking the limit as $\lambda \to 0$ gives

$$f(x)-f(y) \geq \lim_{\lambda \to 0} \frac{f(y+\lambda(x-y))-f(y)}{\lambda}$$
$$= \left(\lim_{\lambda \to 0} \frac{f(y+\lambda(x-y))-f(y)}{\lambda(x-y)}\right)(x-y)$$
$$= f'(y)(x-y). \quad (*)$$

Interchanging the roles of x, y, we obtain

$$f(y) - f(x) \geq f'(x)(y-x). \quad (**)$$

Adding $(*)$ and $(**)$ together gives

$$0 \geq f'(y)(x-y) + f'(x)(y-x)$$
$$= \big(\underbrace{f'(x) - f'(y)}_{\geq 0}\big)\underbrace{(y-x)}_{\leq 0}$$

Thus, if $x \geq y$, then $f'(x) \geq f'(y)$.
In other words, f' is non-decreasing.

Conversely, suppose f is not convex on I. Then there exist $x, y \in I$ and $t \in (0,1)$

such that

$$f(tx+(1-t)y) > tf(x)+(1-t)f(y)$$

Let $z := tx+(1-t)y = y-t(y-x)$

Then

$$t = \frac{y-z}{y-x} \text{ and } 1-t = \frac{z-x}{y-x}$$

Thus, altogether, we have

$$f(z) > \frac{y-z}{y-x} f(x) + \frac{z-x}{y-x} f(y)$$

which implies

$$\frac{f(z)-f(x)}{z-x} > \frac{f(y)-f(z)}{y-z}.$$

By the mean value theorem, $\exists\, u \in [x,z]$ and $v \in [z,y]$ such that

$$f'(u) = \frac{f(z)-f(x)}{z-x} > \frac{f(y)-f(z)}{y-z} = f'(v)$$

i.e. $f'(u) > f'(v)$, but $u \leq v$.

Thus f' is not nondecreasing.

(b) Since f' is nondecreasing iff f'' is nonnegative, the result follows from (a). $\blacksquare$

Example 2.2.2

The functions $f(x) = e^x$ and $g(x) = -\log(x)$ are convex on their domains.

Pf $f''(x) = e^x > 0 \quad \forall x \in \mathbb{R}$

So convex by Lemma 2.2.1(b).

$$g''(x) = \frac{1}{x^2} > 0 \quad \forall x > 0$$

So convex by Lemma 2.2.1(b).

The function $h(x) = x\log(x)$ is convex on $\mathbb{R}_{++} = (0, +\infty)$.

Pf $h''(x) = \frac{1}{x} > 0$ for all $x > 0$.

So convex by Lemma 2.2.1(b).

Recall that the gradient of a function $f: \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ is given by

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

when it exists, it satisfies

$$D_f(x;h)$$
$$\|$$
$$\lim_{t \to 0} \frac{f(x+th) - f(x)}{t} = \langle \nabla f(x), h \rangle$$

The <u>Hessian</u> of f at x is the matrix $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ whose ij-th coordinate is given

$$\left[\nabla^2 f(x)\right]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \quad .$$

<u>Proposition</u> 2.2.3

Let $f : U \to \mathbb{R}$ where $U \subseteq \mathbb{R}^n$ is an open set. Then the following hold.

(a) Suppose f is differentiable. Then f is convex iff

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \quad \forall x, y \in U \tag{2.1}$$

(b) Suppose f is twice differentiable.
Then f is convex iff

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in U$$

$\equiv \nabla^2 f(x)$ is positive semi-definite.

$\equiv$ eigenvalues of $\nabla^2 f(x)$ nonnegative.

Pf (a). Let $x, y \in U$ and $\lambda \in (0,1)$. Then, by convexity of f, we have

$$\begin{aligned} & f(y) + \lambda(f(x) - f(y)) \\ &= \lambda f(x) + (1-\lambda) f(y) \\ &\geq f(\lambda x + (1-\lambda) y) \\ &= f(y + \lambda(x-y)). \end{aligned}$$

Rearranging and taking the limit as $\lambda \to 0$ gives

$$f(x) - f(y) \geq \lim_{\lambda \to 0} \frac{f(y + \lambda(x-y)) - f(y)}{\lambda}$$

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

which establishes (2.1).

Conversely, suppose (2.1) holds. Then, for all $x, y \in U$, we have

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \text{ and}$$
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

Summing these two inequalities together gives

$$0 \geq \langle \nabla f(y), x-y \rangle + \langle \nabla f(x), y-x \rangle$$

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

$$\forall x, y \in U \quad (2.2).$$

Let $x, y \in U$ be arbitary. To establish convexity of f, it suffices to show that $g : [0,1] \to \mathbb{R}$ is convex where g is given by

$$g(t) := f((1-t)x + ty)$$

$$= f(x + t(y-x))$$

Aside:

$$g'(t) = \lim_{a \to 0} \frac{g(t+a) - g(t)}{a}$$

$$= \lim_{a \to 0} \frac{f(x+t(y-x)+a(y-x)) - f(x+t(y-x))}{a}$$

$$= \left\langle \nabla f(x+t(y-x)),\ y-x \right\rangle$$

Using (2.2), we deduce

$$\left(\overbrace{g'(t)-g'(s)}^{\geqslant 0}\right)\left(\overbrace{t-s}^{\geqslant 0}\right)$$

$$= \left\langle \nabla f(x+t(y-x)) - \nabla f(x+s(x-y)),\ x-y \right\rangle \times (t-s)$$

$$= \Big\langle \nabla f(x+t(y-x)) - \nabla f(x+s(x-y)),\ \underbrace{[x+t(y-x)] - [x+s(y-x)]}_{=(t-s)(x-y)} \Big\rangle$$

$$\geqslant 0 \quad \text{(by (2.2))}.$$

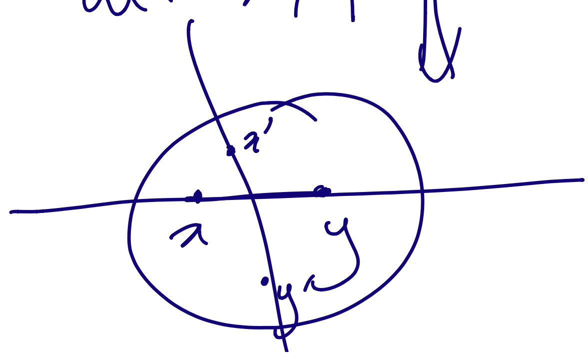
If $t \geq s$, then $g'(t) \geq g'(s)$.
That is, g' is nonincreasing on $[0,1]$
$\Rightarrow$ g is convex by Lemma 2.2.1(a),
Hence, the result follows.

(b) Exercise assignment 2.

$\boxtimes$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for all x, y



\

Definition 2.2.4

An operator $T: \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous (where $L \geq 0$) if

$$\|T(x) - T(y)\| \leq L \|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

Proposition 2.2.5 (descent lemma)

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

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

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

Pf. Let $x, y, z \in \mathbb{R}^n$ and set

$$\phi(t) := f(z + t(y-z)), \quad t \in [0,1].$$

Since f is differentiable, we have:

$$\phi'(t) = \langle y-z, \nabla f(z + t(y-z)) \rangle$$

By the Fundamental Theorem of Calculus applied to ϕ, we have.

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)\,dt$$

Thus, we have

$$f(y) - f(z)$$

$$= \int_0^1 \langle y-z, \nabla f(z + t(y-z)) \rangle dt$$

$$= \int_0^1 \langle y-z, \nabla f(z+t(y-z)) - \nabla f(z) \rangle$$

$$+\langle y-z, \nabla f(z)\rangle \, dt$$

$$= \int_0^1 \langle y-z, \nabla f(z+t(y-z)) - \nabla f(z)\rangle dt$$

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

$$= \int_0^1 \frac{1}{t} \langle \cancel{z} + t(y-z) - \cancel{z}, \nabla f(z+t(y-z)) - \nabla f(z)\rangle dt$$

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

C-S

$$\leq \int_0^1 \frac{1}{t} \|t(y-z)\| \|\nabla f(z+t(y-z)) - \nabla f(z)\| dt$$

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

Lip

$$\leq \int_0^1 \frac{1}{t} \|t(y-z)\| \cdot L \|\cancel{z} + t(y-z) - \cancel{z}\| dt$$

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

$$= \int_0^1 t L \|y-z\|^2 dt + \langle y-z, \nabla f(z)\rangle$$

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

Altogether, we have.

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

Since f is differentiable and convex, Proposition 2.2.3(a) implies

$$f(z) + \langle \nabla f(z), x-z\rangle \le f(x) \quad \left(\text{i.e. } \begin{matrix}\text{this is}\\ (2.1)\end{matrix}\right)$$

$$\Rightarrow 0 \le f(x) - f(z) + \langle \nabla f(z), z-x\rangle \quad (**)$$

Adding (*) and (**) gives.

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

Which completes the proof. $\boxed{\text{Ei}}$.

Example 2.2.6

Let $f(x) = \frac{1}{2}\|x\|^2$. Then $\nabla f(x) = x$.

which is 1-Lipschitz cont since

$$\|\nabla f(x) - \nabla f(y)\| = \|x-y\|$$
$$\leq L\|x-y\|$$
with $L=1$.