Convex Functions, part 2

Thursday, January 28, 2021 5:59 PM

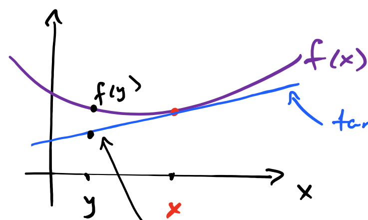
First-order conditions §3.1.3 BV'04

i.e., characterize convexity of f by looking at f' (i.e., ∇f)

why? differentiability at boundary is sketchy

Fact 1 If $f:\mathbb{R}^n \to \mathbb{R}$ is differentiable on dom(f), and if dom(f) is open and convex,

then f is convex iff ① $\underbrace{\forall x,y \in \mathrm{dom}(f),\ f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle}_{\text{i.e., Taylor approximation is an underestimate}}$



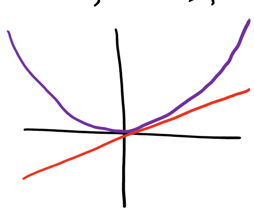
Fact 2 "..." (same setup)

then f is convex iff ② $\forall\ x, y \in \mathrm{dom}(f)$,

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

i.e., $\boxed{\nabla f \text{ is monotone}}$

Recall, in 1D, $f:\mathbb{R}\to\mathbb{R}$ is convex if slope is non-decreasing, i.e., monotone



i.e., if $x \geq y$

then $f'(x) \geq f'(y)$

i.e., $x - y \geq 0 \Rightarrow f'(x) - f'(y) \geq 0$

$x - y \leq 0 \Rightarrow f'(x) - f'(y) \leq 0$

and since positive · positive ≥ 0
negative · negative ≥ 0

$$(x-y)\cdot(f'(x) - f'(y)) \geq 0$$

So we can see the connection w/ our definition of monotone.

Fact 3 (2nd order condition) $f:\mathbb{R}^n \to \mathbb{R}$

If $\nabla^2 f(x)$ ("Hessian", like f'') exists for all $x \in \mathrm{dom}(f)$

then ③a f is convex iff $\nabla^2 f(x) \succeq 0\ \ \forall x \in \mathrm{dom}(f)$ (pos. semidef) and (3b) f is μ-strongly convex (w.r.t. $\|\cdot\|_2$) iff $\nabla^2 f(x) \succeq \mu \cdot I$.

⚠ $\nabla^2 f(x) \succ I \implies$ strictly convex (sufficient but not necessary)

Remark Change the $\leq$ to $<$ for (1) and (2) for strict convexity

⚠ f can be convex but ∇f and $\nabla^2 f$ need not exist!

So... what about when f isn't differentiable?

Def Let $f: \mathbb{R}^n \to (-\infty, \infty]$ be proper, then we define the subdifferential of f at x to be

$$\underbrace{\partial f(x)}_{\text{subdiff.}} := \{ \underbrace{d}_{\text{subgradient}} \in \mathbb{R}^n \mid \forall y \in \mathbb{R}^n,\ f(y) \geq f(x) + \langle d, y - x \rangle \}$$

(this may be an empty set... but:

Fact If f is proper and convex then (if domain is $\mathbb{R}^n$)

if $x \in \operatorname{relint}(\operatorname{dom}(f)) \implies \partial f(x) \neq \emptyset$)

(Related to separating/supporting hyperplanes)

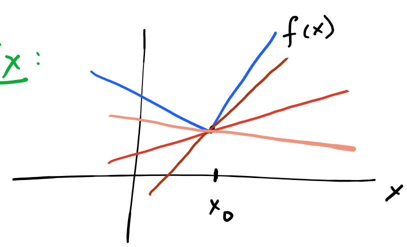
This should remind you of (1). In fact,

roughly, $\partial f(x)$ is a singleton, i.e. $|\partial f(x)| = 1$, iff f differentiable at x

(precisely, cf Prop. 17.26 Bauschke + Combettes '11, f proper and convex, $x \in \operatorname{dom}(f)$, then 1) f differentiable at $x \implies \partial f(x) = \{\nabla f(x)\}$

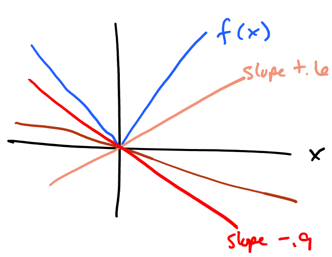
2) f continuous at x (e.g., $x \in \operatorname{relint}(\operatorname{dom}(f))$), and $\partial f(x) = \{u\}$, then f differentiable at x and $\nabla f(x) = u$)

Ex:



For $x \neq x_0$, $\partial f(x) = \{ f'(x) \}$ since differentiable

For $x = x_0$, $\partial f(x_0)$ has more than one entry



More specifically, let $f(x) = |x|$, $f: \mathbb{R}^1 \to \mathbb{R}$

Then if $x \neq 0$, $f'(x) = \operatorname{sign}(x)$, and $\partial f(x) = \{f'(x)\}$

If $x = 0$, $f'(0)$ DNE

but $\partial f(0) = [-1, 1]$

Fundamental Theorem of Convex Optimization*

* no one else calls it this

Thm: Fermat's Rule If f is a proper function, then

$$\operatorname*{argmin}_x f(x) = \underbrace{\{ x : 0 \in \partial f(x) \}}$$

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

i.e., $f(x) \leq f(y)\ \forall y$

i.e., x is a global minimizer

proof: triviality

[This covers constrained optimization too, since f can be extended valued]

Generalizes 1D idea of finding x s.t. $f'(x) = 0$

or smooth and unconstrained notion of finding x s.t. $\nabla f(x) = 0$.

Subdifferentials are a global notion ... yet gradients are a local notion.

For convex functions, the convexity (a global notion) links the two.

... so, all we need to do is invert ∂f, i.e., $\operatorname{argmin} f(x) = \partial f^{-1}(0)$

this is usually not directly possible for interesting problems (though it may be possible for subproblems).

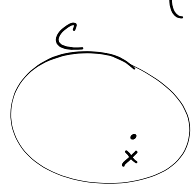
Ex Let $C \neq \emptyset$ be convex, so I_C is a proper convex function.

Then $\partial I_C = N_C$, the normal cone

Def The normal cone to a set C at the point x is

$$N_C(x) = \begin{cases} \{ d \mid \langle d, y-x \rangle \leq 0 \quad \forall y \in C \} & \text{if } x \in C \\ \emptyset & \text{if } x \notin C \end{cases}$$

Ex



$x \in \operatorname{int}(C) \Rightarrow N_C(x) = \{0\}$

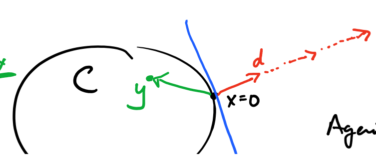
why? wlog, shift so $x=0$

If $\langle d, y \rangle \leq 0$ for all $y \in C$

then $x \in \operatorname{int}(C) \Rightarrow y = \varepsilon d \in C$ for ε sufficiently small but nonzero

$\Rightarrow \varepsilon \|d\|^2 \leq 0 \Rightarrow d = 0$.

Ex



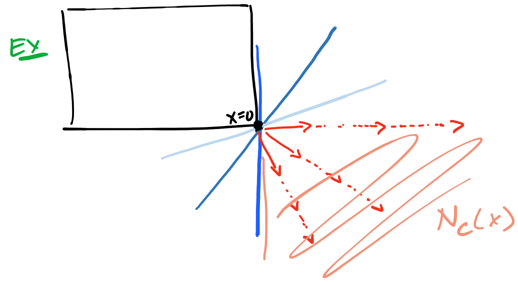
$x \in \partial C$ ← overloaded notation! means "boundary"

Again wlog let $x=0$, want d s.t. $\langle d, y \rangle \leq 0\ \forall y \in C$

hyperplane defined by d

(In fact, equivalent definition, p.66 BV'04
$N_C(x)$ = set of all vectors that define a supporting hyperplane to C, passing through x)

Ex

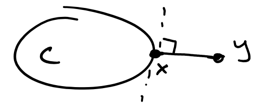


Ex. where $N_C(x)$ is solid

Ex If C is a vector (sub-)space, $N_C(x) = \begin{cases} C^\perp & x \in C \\ \phi & x \notin C \end{cases}$

Fact If $C \neq \phi$ is closed and convex then
$x = P_C(y)$ iff $y - x \in N_C(x)$

orthogonal projection



cf Prop. 6.46 Bauschke + Combettes '11

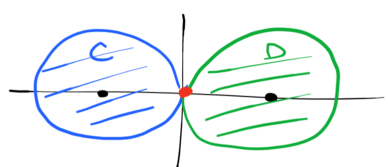
Calculus means a set of rules we can use to calculate

"Rule #1 from Calc. I that is so simple you never even have to cite it"

$\frac{d}{dx}(f+g) = f' + g'$, or in $\mathbb{R}^n$, $\nabla(f+g) = \nabla f + \nabla g$

Is it true that $\partial(f+g) = \partial f + \partial g$? No! (but often true)

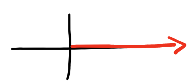
Ex: $f = I_C$, $g = I_D$ in $\mathbb{R}^2$



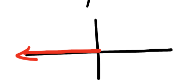
C, D balls of radius 1 centered at $(-1,0)$ and $(+1,0)$, resp.

Then $\partial(f+g)(x) = \partial f(x) + \partial g(x)$ for all x except at $x=0$

At $x=0$, $\partial f(0) = N_C(0) = \mathbb{R}_+ \times \{0\}$

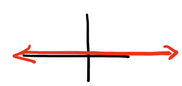


$\partial g(0) = N_D(0) = \mathbb{R}_- \times \{0\}$



so $\partial f(0) + \partial g(0) = \mathbb{R} \times \{0\}$ (A)

Minkowski sum



but... $\partial(f+g)(0) = N_{C \cap D}(0) = N_{\{\vec{0}\}}$ $C \cap D = \{\vec{0}\}$

$:= \{ d \mid \underbrace{\langle d, y - \vec{0} \rangle \le 0 \ \forall y \in \{0\}}_{\text{vacuous constraint}} \}$

$= \mathbb{R}^2$ (B)

(A) $\neq$ (B)

It's often true $\partial(f+g) = \partial f + \partial g$.

Sufficient conditions to guarantee when that is true are called constraint "CQ" qualifications.

i.e., ex: Slater

Cor. 16.38 (iv) Bauschke + Combettes '11

If $f, g \in \Gamma_0(\mathcal{H})$, and $\mathcal{H} = \mathbb{R}^n$, then if

1) $\operatorname{relint}(\operatorname{dom}(f)) \cap \operatorname{relint}(\operatorname{dom}(g)) \neq \emptyset$ } fancy

or 2) $\operatorname{dom}(f) \cap \operatorname{int}(\operatorname{dom}(g)) \neq \emptyset$

or 3) * either f or g has full domain (all of $\mathbb{R}^n$) } most commonly used

then $\partial(f+g) = \partial f + \partial g$.

... back to example $f = I_C$, $g = I_D$. This did not satisfy a CQ

$\operatorname{dom}(f) = C$

$\operatorname{dom}(g) = D$, $C \cap D = \vec{0}$

$\operatorname{int}(C) \cap \operatorname{int}(D) = \emptyset$

none of 1) 2) 3) above hold.

⚠ A lot of people are unaware of CQ and their importance.

Lack of CQ lead to "degenerate" optimization problems.

You hear, "MAXCUT, a NP-Hard problem, has a SDP relaxation (which is solvable in polynomial time) that guarantees a good approximation."

but is only "sorta" true: degenerate SDP are not efficiently solvable

(also, "polynomial time" depends on your model of computation, and if LP / SDP are strongly polynomial is an open question)