Minimizers, intro to convexity

Tuesday, January 19, 2021 5:26 PM

First, go over discussion questions from last time

Solutions

Generic problem: $\min_{x \in C} f(x)$, $C \subseteq \mathbb{R}^n$

- A feasible point x means $x \in C$

- A solution or minimizer or global minimizer x^* means
  1) $x^* \in C$
  2) $\forall y \in C,\ f(x^*) \le f(y)$ (NOT $f(x^*) < f(y)$)

  This is what we usually really want

  i.e., $x^* \in \operatorname{argmin}_{x \in C} f(x)$

⚠ Solutions may not exist ... even for convex problems

Ex $\min_{x \in \mathbb{R}} x$ Ex $\min_{x \in (0,1)} x^2$

abuse of notation. it's more proper to write $\inf_{x \in (0,1)} x^2$ but optimizers are too lazy

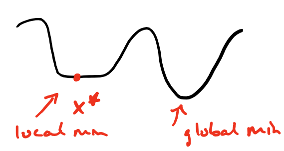
⚠ Solutions need not be unique ... even for convex problems

Ex $\min_{x \in \mathbb{R}} f(x)$ where $f(x) = 0\ \forall x$

- x^* is a local minimizer if x^* is feasible and $\exists\, \varepsilon > 0$ s.t.

  $f(x^*) \le f(y)\quad \forall y \in C \cap \underbrace{B_\varepsilon(x^*)}_{\text{open ball of radius } \varepsilon} := \{ y : \|y - x^*\| < \varepsilon \}$

Ex:



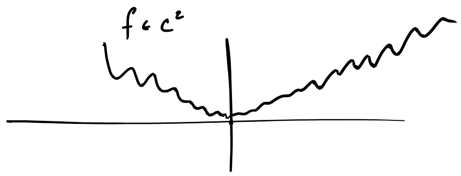
- x^* is a strict local minimizer if x^* is feasible and $\exists\, \varepsilon > 0$ s.t.

  $f(x^*) < f(y)\quad \forall y \in B_\varepsilon(x^*) \cap C$

- x^* is an isolated (and strict) local min if a local min, and no other local min are nearby

  FAIRLY ESOTERIC

Ex $f(x) = \begin{cases} x^4 \cos(1/x) + 2x^4 & x \ne 0 \\ 0 & x = 0 \end{cases}$ $x = 0$ is a strict but not isolated local min



i.e., $\exists$ sub-optimal local min x_j w/ $x_j \to 0$

B/w, notation

$f \in C^3$ means f, f', f'', f''' all exist and are continuous

or $f \in C^3(\mathbb{R}^n)$ means $f, \nabla f, \nabla^2 f, \nabla^3 f$ all exist and are continuous

Connections w/ Calc. I

$\min f(x)$
$x \in [a,b]$

The procedure: 1) find all points x_i such that $f'(x_i) = 0$ "critical numbers" or where f is discontinuous or not differentiable

2) throw in $\{a, b\}$ boundary

3) minimize $\{a, b, x_i\}$

in textbook exercises, $\{x_i\}$ was always finite

(Ex: Minimize $\{3, -2, 5, 17, -31, 21\}$
It's easy: it's -31)

generalization

- A critical or stationary point of $f(x)$ is a point where $\nabla f(x) = 0$

... but this isn't enough usually.

1) it doesn't take into account constraints.

If $x \in \partial C$ then don't expect $\nabla f(x) = 0$ here

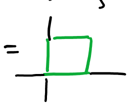
$\underbrace{\partial C}_{\text{boundary}}$ boundary, $= \overline{C} \setminus \text{int}(C)$, $\text{int}(C) =$ largest open set in C

$\overline{C} = \text{closure}(C)$
$=$ smallest closed set containing C

$= \{x \in C : \exists\, \varepsilon > 0 \;\; B_\varepsilon(x) \subseteq C\}$

2) we can't just check boundary separately

in $\mathbb{R}$, if $C = [0,1]$, $\partial C = \{0, 1\}$, so $|\partial C| = 2$ ← cardinality

in $\mathbb{R}^2$, if $C = [0,1]^2$, $\partial C =$  so $|\partial C| = \infty$

3) often there are an ∞ # (or at least a large #) of critical points, so can't check all

- If x^* is a local or global minimizer and $C = \mathbb{R}^n$ (i.e., unconstrained) HW exerc.

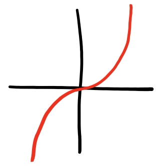
then x^* is a critical point ← "necessary"

minimizer $\Rightarrow$ crit. pt.

but if x^* is a critical point, it need not be a local or global minimizer

(so not sufficient) in which case we often call it a saddle point

Ex: $f(x) = x^3$



$x^* = 0$ is a saddle point

($f'(0) = 3x^2\big|_{x=0} = 0$ so it's a crit. pt., but not a local min.)

19th century analysis theorem

Sometimes called "Weierstrass thm":

ie, in $\mathbb{R}^n$, means closed and bounded

if f is continuous and C is compact then

f achieves its infimum over C, ie, $\inf\limits_{x \in C} f(x) = \min\limits_{x \in C} f(x)$

(our first existence of minimizers result)

proof

Q1

on your own, using these ingredients:

1) in $\mathbb{R}^n$, compact $\Leftrightarrow$ sequentially compact (i.e., every sequence in C has a convergent subseq. whose limit is in C)

2) in $\mathbb{R}^n$ w/ usual topology (ie, induced by a norm),

a function f is cts $\Leftrightarrow$ sequentially cts (ie, $x_n \to x \Rightarrow f(x_n) \to f(x)$)

shorthand for continuous

One way to prove it:

First, note the set $f(C) := \{ f(x),\ x \in C \}$ is bounded (compact in fact)

since if not bounded, pick $(\alpha_k) \subseteq f(C)$ s.t. $|\alpha_k| \geq k$

but then $\exists\ x_k$ st. $\alpha_k = f(x_k)$, and $\exists$ convergent subseq. $x_{k_j} \to x \in C$

so f cts $\Rightarrow$ $\lim_{j\to\infty} \underbrace{f(x_{k_j})}_{= \lim \alpha_{k_j}} = f(x)$ ← limit

$= \lim \alpha_{k_j}$ ← unbounded

Convergent sequences are bdd. Contradiction

So this guarantees $\inf_{x \in C} f(x) =: f^*$ is finite.

So, second step, by "infimum", pick $\alpha_k \in f(C)$ st. $\alpha_k \to f^*$

again, $\exists\ x_k$ st $\alpha_k = f(x_k)$, again pick convergent subseq. x_{k_j},

again use continuity of f. $\square$

So... it's nice when our constraints C are compact

It turns out that's often too much to ask for, but at the very least, we like our constraint sets to be closed

Ex: $\min \|x\|^2$ (1) $\quad\quad$ $\min \|x\|^2$ (2)

s.t. $\|Ax - b\| < \varepsilon$ $\quad\quad$ s.t. $\|Ax - b\| \leq \varepsilon$

very similar, but (1) probably doesn't have a solution, whereas (2) does.

Other high-level properties of the feasible set C:

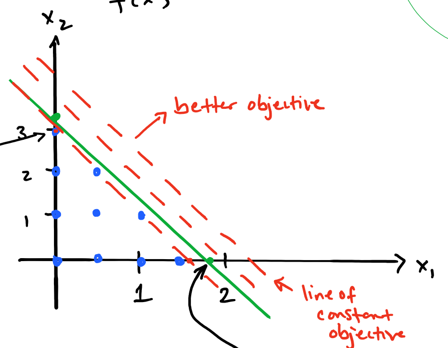
- If $C = \emptyset$, the problem is "infeasible."

  Not always obvious to spot, ex: $C = \{ x : Ax \leq b \}$ may or may not be empty.

• In this class, C will usually be convex ... in particular, <u>not</u> <u>integral</u>. $\mathbb{Z}^n$

ex: $C = \{ x \in \mathbb{R}^2,\ x \geq 0,\ 7 \cdot x_1 + 4 \cdot x_2 \leq 13,\ \overbrace{x_1, x_2 \in \mathbb{Z}}^{\text{makes it hard}} \}$ $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$

$\max\ \underbrace{21 x_1 + 11 x_2}_{f(x)}$ s.t. $x \in C$ "integer linear program"



the line $7x_1 + 4x_2 = 13$ is drawn in green (intercepts: $x_2 = 0,\ x_1 = \frac{13}{7} \approx 1.85$
$x_1 = 0,\ x_2 = 13/4 = 3.25$)

C shown in blue

<u>level sets</u> of objective $f(x)$ in red

eg. $f(x) = \alpha$

$\alpha = 7$: $21 x_1 + 11 x_2 = 33$

intercepts $(0, 3)$ and $(\frac{33}{21}, 0)$ ≈ 1.57

w/o $\mathbb{Z}^2$ constraints,
max is at $(13/7, 0)$
w/ $\mathbb{Z}^2$ constraint,
max is at $(0, 3)$

Far away.

Rounding $(13/7, 0)$ to nearest integers is <u>not</u> optimal
Done all the time in practice.
Sometimes it works sorta ok... often it doesn't work well at all.

(More on this later in semester)

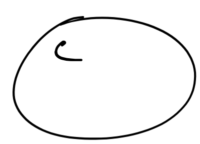
Convexity

<u>Def</u>. A <u>set</u> C $\overbrace{\text{inside any vector space}}^{\text{from now on, this is always implicitly assumed in this class}}$ is **convex** if $\forall\ x, y \in C$ and $\forall\ t \in [0,1]$,

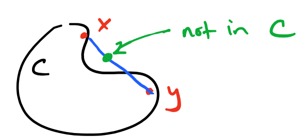
then $t \cdot x + (1-t) y \in C$

i.e.,

$\overline{xy} = \{ t x + (1-t) y : t \in [0,1] \}$



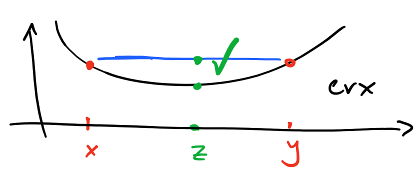
Convex



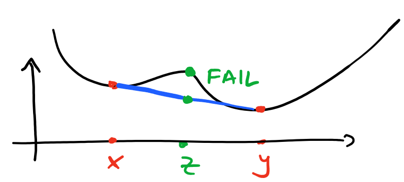
not convex

<u>Def</u> A <u>function</u> $f: \mathbb{R}^n \to \mathbb{R}$ is **convex** if $\forall\ x, y \in \mathbb{R}^n$, $\forall\ t \in [0,1]$

then $f(\underbrace{t x + (1-t) y}_{z}) \leq t \cdot f(x) + (1-t) f(y)$



cvx



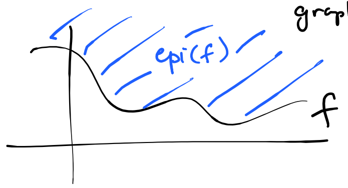
not convex

or, since mathematicians don't like to write/talk more than necessary,

a function f is **convex** iff the set "epi(f)" is **convex** where

$\mathrm{epi}(f) = \text{epigraph}(f) = \{ (x,s) : x \in \mathbb{R}^n, s \in \mathbb{R}, s \geq f(x) \}$

$\text{graph}(f) = \{ (x,s) : \text{" ... "} \quad s = f(x) \}$



Why do we like convexity?

Basic thm #1: If f is convex and C is convex, then any local minimizer of $\min_{x \in C} f(x)$ is in fact global. (Also, $X := \operatorname{argmin}_{x \in C} f(x)$, the set of global sol'n, is convex, an in particular, connected)

proof

$\boxed{Q2}$ on your own (scroll down for answer)

proof: let x be a local sol'n. For contradiction, assume it's not global, so $\exists\, y \in C$ s.t. $f(y) < f(x)$.

Since $x \in C$ is a local sol'n, $\exists\, \varepsilon$ st. $f(x) \leq f(z)$ $\forall z \in B_\varepsilon(x) \cap C$

Pick $t \in [0,1]$ sufficiently close to 1 to guarantee

$z = tx + (1-t)y$ is within ε of x

Since C is convex, $z \in C$, thus by we know $f(x) \leq f(z)$. $(*)$

But by convexity of f,

$$
\begin{aligned}
f(z) &= f(tx + (1-t)y) \\
&\leq t \cdot f(x) + (1-t) f(y) \\
&< t \cdot f(x) + (1-t) f(x) \quad \text{since } (1-t) \geq 0 \text{ and } f(\ ) < f(\ ) \\
&= f(x)
\end{aligned}
$$

so $f(z) < f(x)$ which contradicts $(*)$.