Optima

Lecture 04

September 7, 2021

Beck, section 2.1

Optimization

Goal: Solve minimization/maximization problems.

Given $f: S \to \mathbb{R}$, $S \subset \mathbb{R}^n$

$$\max_{\underline{x} \in S} f(\underline{x}) \quad \text{or} \quad \min_{\underline{x} \in S} f(\underline{x})$$

Challenges: (i) f can be "complicated"

(ii) S can be "complicated"

f: "objective" / "design" / "loss"

S: "feasible set"

There can be many solutions to optimization problems:

$$\{\underline{x} \in S : f(\underline{x}) \text{ is maximized over } S\} = \underset{\underline{x} \in S}{\operatorname{argmax}} f(\underline{x})$$

Similarly, $\underset{\underline{x} \in S}{\operatorname{argmin}} f(\underline{x})$

argmax's and argmin's can contain many points in $\mathbb{R}^n$.

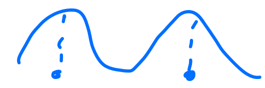
<u>Goal</u>: (i) compute an extremal value of f.

(ii) compute/characterize argmin/argmax set or point.

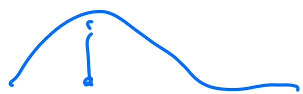
Global extrema/optima

Given $f: S \to \mathbb{R}$, $S \subset \mathbb{R}^n$,

(i) A feasible point $\underline{x} \in S$ is a global maximum if
$f(\underline{y}) \le f(\underline{x}) \quad \forall\, \underline{y} \in S.$



(ii) A feasible point $\underline{x} \in S$ is a strict global maximum if
$f(\underline{y}) < f(\underline{x}) \quad \forall\, \underline{y} \in S,\ \underline{y} \ne \underline{x}.$

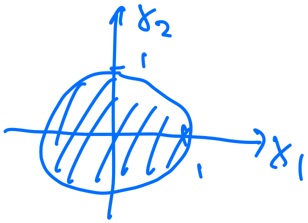


(iii) A feasible point $\underline{x} \in S$ is a global minimum if
$f(\underline{y}) \ge f(\underline{x}) \quad \forall\, \underline{y} \in S.$

(iv) Similarly for "strict global minimum".

Any point $\underline{x}$ satisfying any of the above conditions is a global extremum or global optimum.

Example



Example

For $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$, compute

$$\arg\max_{\boldsymbol{x} \in B[\mathbf{0},1]} f(\boldsymbol{x}), \qquad f(\boldsymbol{x}) = x_1 + x_2.$$

closed unit ball, using ℓ^2 norm.

Determine the extremal point(s) and discuss uniqueness.

Note: $f(\underline{x}) = x_1 + x_2 = \langle \underline{x}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle$

Use Cauchy-Schwarz inequality: $|\langle \underline{x}, \underline{y} \rangle| \leq \|\underline{x}\| \cdot \|\underline{y}\|$

equality iff $\underline{x}, \underline{y}$ are parallel

$$f(\underline{x}) = \left\langle \underline{x}, \begin{pmatrix}1\\1\end{pmatrix} \right\rangle \le \|\underline{x}\| \cdot \left\| \begin{pmatrix}1\\1\end{pmatrix} \right\|$$

equality
iff $\underline{x}$ is parallel to $\begin{pmatrix}1\\1\end{pmatrix}$.

i.e. $|f(\underline{x})| = \left\| \begin{pmatrix}1\\1\end{pmatrix} \right\| \cdot \|\underline{x}\|$ if $\underline{x}$ is parallel to $\begin{pmatrix}1\\1\end{pmatrix}$

$\underline{x}$ parallel to $\begin{pmatrix}1\\1\end{pmatrix} \Longrightarrow \underline{x} = \alpha \begin{pmatrix}1\\1\end{pmatrix}$, $\alpha \in \mathbb{R}$.

global extrema of f occur when $\underline{x} = \alpha \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}x_1\\x_2\end{pmatrix}$

$$f(\underline{x}) = x_1 + x_2 = 2\alpha$$

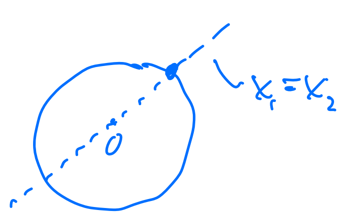
$$\max_{\underline{x} \in B[\underline{0},1]} f(\underline{x}) = \max_{\underline{x} \in B[\underline{0},1]} 2\alpha = \max_{\underline{x} \in B[\underline{0},1]} \frac{2}{\sqrt{2}} \|\underline{x}\|$$

$\|\underline{x}\| = \alpha\sqrt{2}$

$$= \frac{2}{\sqrt{2}} = \sqrt{2}$$

Summary: $\max\limits_{\underline{x} \in B[\underline{0},1]} f(\underline{x}) = \sqrt{2}$, $\operatorname{argmax}\limits_{\underline{x} \in B[\underline{0},1]} f(\underline{x}) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

$\alpha = 1/\sqrt{2}$



$\underline{x} = (1/\sqrt{2}, 1/\sqrt{2})$ is a strict global maximum
(strict because of Cauchy-Schwarz)

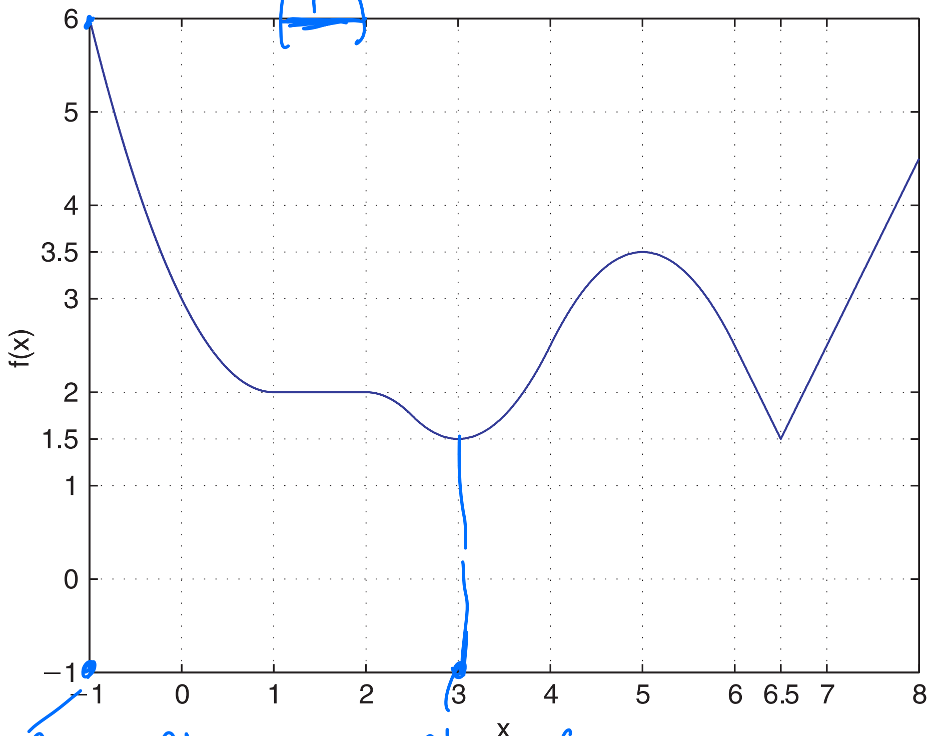
Also: $\underline{x} = (-1/\sqrt{2}, -1/\sqrt{2})$ is a strict global minimum

Local extrema/optima

Example

Lmax, Lmin

Gmax
Gmin
SGmax
SGmin



SGmax, SLmax

SLmin, Gmin

Figure 2.2. *Local and global optimum points of a one-dimensional function.*

First-order optimality

Recall: if $f:(a,b) \to \mathbb{R}$ and f is differentiable on (a,b), then any local optimum of f on (a,b) satisfies $f'(x) = 0$.

Lemma: (First-order optimality)

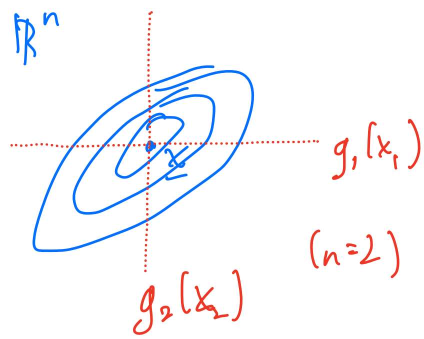
Suppose $f: S \to \mathbb{R}$, $S \subset \mathbb{R}^n$ is open, is differentiable on S. Then if $\underline{x} \in S$ is a local optimum of f, then $\nabla f(\underline{x}) = \underline{0}$.

("Necessary condition for local optimality")

Proof (sketch):

$\underline{x} \in S$ is a local optimum

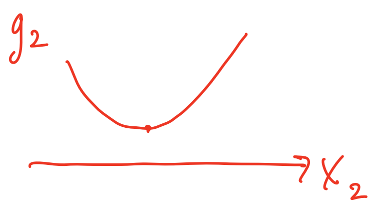
$x_1, \dots x_n$ are coordinates of $\underline{x}$.



I.e., define $g_j(x_j) = f(x_1, \dots x_j, \dots x_n)$

Idea: $g_j(x_j)$ is a 1D function.

And it has a local optimum @ x_j.



I.e., $g_j'(x_j) = 0$.

But $g_j'(x_j) = \frac{\partial f}{\partial x_j}(\underline{x}) = 0$

$$\Longrightarrow \frac{\partial f}{\partial x_j}(\underline{x}) = 0 \Longrightarrow \nabla f(\underline{x}) = \underline{0} \quad \blacksquare$$

Example

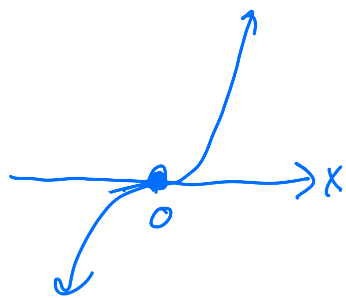
Example

Classify the stationary points of $f(x) = x^3$ on $\mathbb{R}$.

Def: If f is differentiable, then $\underline{x}$ is a stationary point if $\nabla f(\underline{x}) = 0$.

$f'(x) = 3x^2 = 0 \Rightarrow$ stationary point @ $x = 0$

$x = 0$ is not a local optimum



Example

Example

Compute the global optima of

$$f(\boldsymbol{x}) = \frac{x_1 + x_2}{x_1^2 + x_2^2 + 1},$$

for $\boldsymbol{x} \in \mathbb{R}^2$.