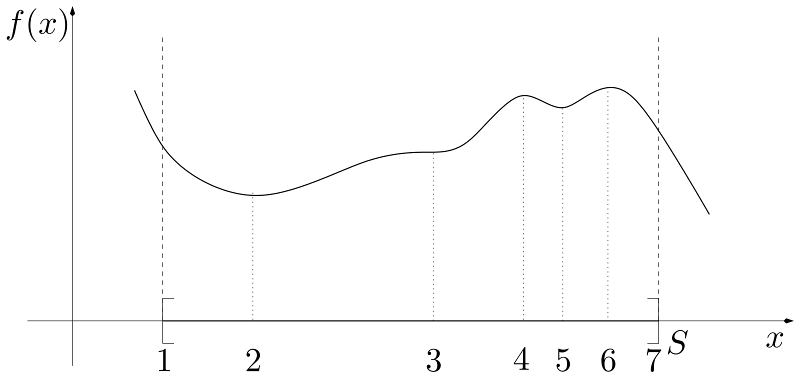
Lecture 3: Optimality conditions with applications

Local and global optimality, $\mathbb{R}$ and $\mathbb{R}^n$

$$\text{minimize } f(\boldsymbol{x}), \tag{1a}$$
$$\text{subject to } \boldsymbol{x} \in S, \tag{1b}$$

$S \subseteq \mathbb{R}^n$ nonempty set, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ a given function



Interesting points

(i) boundary points of S

(ii) stationary points, that is, where $f'(x) = 0$

(iii) discontinuities in f or f'

Here:

(i) 1, 7

(ii) 2, 3, 4, 5, 6

(iii) none

Global and local minimum

- $\boldsymbol{x}^* \in S$ is a *global minimum* of f over S if it attains the lowest value of f over S:

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \qquad \boldsymbol{x} \in S$$

- $\boldsymbol{x}^* \in S$ is a *local minimum* of f over S if there exists a small enough ball intersected with S around $\boldsymbol{x}^*$ such that it is an optimal solution in that smaller set: with $B_\varepsilon(\boldsymbol{x}^*) := \{\, \boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y} - \boldsymbol{x}^*\| < \varepsilon \,\}$ being the Euclidean ball with radius ε centered at $\boldsymbol{x}^*$, we get

$$\exists \varepsilon > 0 \text{ such that } f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \qquad \boldsymbol{x} \in S \cap B_\varepsilon(\boldsymbol{x}^*).$$

- $\boldsymbol{x}^* \in S$ is a *strict local minimum* of f over S if $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$ holds above for $\boldsymbol{x} \neq \boldsymbol{x}^*$.

Fundamental Theorem of global optimality

Consider the problem (1), *where S is a convex set and f is convex on S. Then, every local minimum of f over S is also a global minimum.*

Proof.

Intuitive image: If $\boldsymbol{x}^*$ is a local minimum, then f cannot go down-hill from $\boldsymbol{x}^*$ in any direction, but if $\bar{\boldsymbol{x}}$ has a lower value, then f has to go down-hill sooner or later. This cannot be the shape of any convex function.

Weak coercivity

$S \subseteq \mathbb{R}^n$ nonempty and closed, $f : S \to \mathbb{R}$

- f is *weakly coercive* with respect to the set S if either S is bounded or

$$\lim_{\substack{\|\boldsymbol{x}\| \to \infty \\ \boldsymbol{x} \in S}} f(\boldsymbol{x}) = \infty$$

  holds.

- The weak coercivity of $f : S \to \mathbb{R}$ is equivalent to the property that f has bounded level sets (Why?)

Lower semi-continuity

$S \subseteq \mathbb{R}^n$ nonempty and closed, $f : S \to \mathbb{R}$

- f is lower semi-continuous at $\bar{\boldsymbol{x}} \in S$ if the value $f(\bar{\boldsymbol{x}})$ is less than or equal to every limit of f as $\{\boldsymbol{x}_k\} \to \bar{\boldsymbol{x}}$.
  In other words, f is lower semi-continuous at $\bar{\boldsymbol{x}} \in S$ if

$$\{\boldsymbol{x}_k\} \to \bar{\boldsymbol{x}} \qquad \Longrightarrow \qquad f(\bar{\boldsymbol{x}}) = \liminf_{k\to\infty} f(\boldsymbol{x}_k).$$

- Lower semi-continuity of f is equivalent to the closedness of all its level sets $\operatorname{lev}_f(b)$, $b \in \mathbb{R}$, as well as the closedness of its epigraph (Why?)
- Lower semi-continuous functions in one variable have the appearance shown in Figure 1.
- $\exists$ corresponding notion of upper semi-continuity. Continuity = both upper and lower semi-continuity

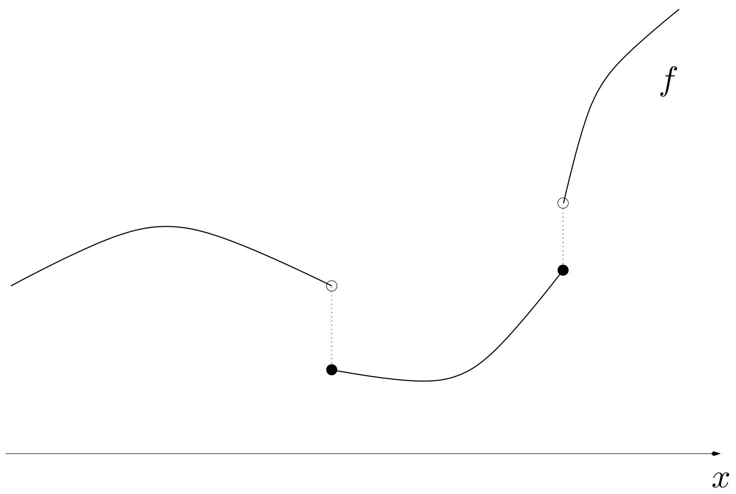


Figure 1: A lower semi-continuous function in one variable

Existence of optimal solutions, I: Weierstrass

Weierstrass' Theorem *Let $S \subseteq \mathbb{R}^n$ be a nonempty and closed set, and $f : S \to \mathbb{R}$ be a lower semi-continuous function on S. If f is weakly coercive with respect to S, then there exists a nonempty, closed and bounded (thus compact) set of optimal solutions to the problem* (1).

Proof.

Under the given assumptions there is an $\bar{\boldsymbol{x}} \in S$ with $f(\bar{\boldsymbol{x}}) < \infty$. The set

$$\{ (\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(\boldsymbol{x}) \leq \alpha; \quad \alpha \leq f(\bar{\boldsymbol{x}}) \}$$

is nonempty, closed and bounded. To minimize f over S is the same as finding the minimum value of α over all pairs $(\boldsymbol{x}, \alpha)$ in the above set; α ranges over a closed and bounded interval.

Existence of solutions, II: Specialization to LP

- Suppose $S = \{ \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}; \quad \boldsymbol{E}\boldsymbol{x} = \boldsymbol{d} \}$; $f(\boldsymbol{x}) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$
- *Three equivalent statements:*

  (a) *The problem* (1) *has a nonempty (polyhedral) set of optimal solutions*

  (b) f *is lower bounded on* S

  (c) *For every feasible direction* $\boldsymbol{p}$ *of* S, *it holds that* $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{p} \geq 0$
- Stronger result than the above
- Lower bounded not enough in general; cf. $f(x) = 1/x$ on $x \geq 1$

Optimality over $\mathbb{R}^n$, $f \in C^1$

- $\boldsymbol{x}^*$ *is a local minimum of* f *on* $\mathbb{R}^n \quad \Longrightarrow \quad \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n$.
- Proof by Taylor expansion, contradiction.
- Direction $\Longleftarrow$ not true: $f(x) = x^3$, $x = 0$.
- Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ be given. Let $\boldsymbol{x} \in \mathbb{R}^n$ be a vector such that $f(\boldsymbol{x})$ is finite. Let $\boldsymbol{p} \in \mathbb{R}^n$. We say that the vector $\boldsymbol{p} \in \mathbb{R}^n$ is a *descent direction* with respect to f at $\boldsymbol{x}$ if

$$\exists \delta > 0 \text{ such that } f(\boldsymbol{x} + \alpha \boldsymbol{p}) < f(\boldsymbol{x}) \text{ for every } \alpha \in (0, \delta].$$

- Sufficient condition: *Suppose that* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is in* C^1 *around a point* $\boldsymbol{x}$ *for which* $f(\boldsymbol{x}) < +\infty$, *and that* $\boldsymbol{p} \in \mathbb{R}^n$. *If* $\nabla f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} < 0$ *then the vector* $\boldsymbol{p}$ *defines a direction of descent with respect to* f *at* $\boldsymbol{x}$.

Optimality over $\mathbb{R}^n$, $f \in C^2$

- $\boldsymbol{x}^*$ *is a local minimum of* f *on* $\mathbb{R}^n \Longrightarrow$
$$\begin{cases} \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n; \\ \nabla^2 f(\boldsymbol{x}^*) \text{ is positive semi-definite.} \end{cases}$$
- [*Note:* $n = 1$: $x^* \in \mathbb{R}$ is a local minimum $\Longrightarrow f'(x^*) = 0$ and $f''(x^*) \geq 0$.]
- $$\left.\begin{array}{r} \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n \\ \nabla^2 f(\boldsymbol{x}^*) \text{ is positive definite} \end{array}\right\} \Longrightarrow$$
$\boldsymbol{x}^*$ *is a strict local minimum of* f *on* $\mathbb{R}^n$.
- [*Note:* $n = 1$: $f'(x^*) = 0$ and $f''(x^*) > 0 \Longrightarrow x^* \in \mathbb{R}$ is a strict local minimum.]

Optimality over $\mathbb{R}^n$, f convex in C^1

- *Let $f \in C^1$, and f be convex. Then,*

$$\boldsymbol{x}^* \text{ is a global minimum of } f \text{ on } \mathbb{R}^n \quad \Longleftrightarrow \quad \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n.$$

- *Proof.*

Active constraints

Consider

$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}),\\ &\text{subject to } \boldsymbol{x} \in S \end{aligned}$$

$S \subseteq \mathbb{R}^n$ nonempty, closed, convex, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ in C^1 on S.

- Feasible directions at $\boldsymbol{x}^*$ depend on active constraints.
- Let $\boldsymbol{x} \in S$, where $S \subseteq \mathbb{R}^n$, and that $\boldsymbol{p} \in \mathbb{R}^n$. Then, $\boldsymbol{p}$ defines a *feasible direction* at $\boldsymbol{x}$ if

$$\exists \delta > 0 \text{ such that } \boldsymbol{x} + \alpha \boldsymbol{p} \in S \text{ for all } \alpha \in [0, \delta].$$

- Suppose

$$S = \{\, \boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) = 0, \quad i \in \mathcal{E}; \quad g_i(\boldsymbol{x}) \le 0, \quad i \in \mathcal{I} \,\}.$$

  Suppose $\boldsymbol{x} \in S$. The set of *active constraints* is the union of all the equality constraints and the set of inequality constraints that are satisfied with equality, that is, the set $\mathcal{E} \cup \mathcal{I}(\boldsymbol{x})$, where $\mathcal{I}(\boldsymbol{x}) := \{\, i \in \mathcal{I} \mid g_i(\boldsymbol{x}) = 0 \,\}$.
- Linear constraints: $g_i(\boldsymbol{x}) := \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{x} - d_i$ $(i \in \mathcal{E})$, $g_i(\boldsymbol{x}) := \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i$ $(i \in \mathcal{I})$.
- Matrix notation: $S = \{\, \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{E}\boldsymbol{x} = \boldsymbol{d}; \quad \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b} \,\}$.
- Feasible directions at $\boldsymbol{x} \in S$:

$$\{\, \boldsymbol{p} \in \mathbb{R}^n \mid \boldsymbol{E}\boldsymbol{p} = \boldsymbol{0}^\ell; \quad \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{p} \le 0,\ i \in \mathcal{I}(\boldsymbol{x}) \,\}.$$

- For nonlinear constraints: more technical! Later! KKT!

Necessary optimality conditions, I: VIP

- *Suppose* $S \subseteq \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is in* C^1 *around* $\boldsymbol{x} \in S$ *for which* $f(\boldsymbol{x}) < +\infty$.
  (a) *If* $\boldsymbol{x}^* \in S$ *is a local minimum of* f *on* S *then* $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} \ge 0$ *holds for every feasible direction* $\boldsymbol{p}$ *at* $\boldsymbol{x}^*$.
  (b) *Suppose that* S *is convex and that* f *is in* C^1 *on* S. *If* $\boldsymbol{x}^* \in S$ *is a local minimum of* f *on* S *then*

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \ge 0, \qquad \boldsymbol{x} \in S. \tag{2}$$

- *Proof.*

Convex case

- We refer to (2) as a *variational inequality*.
- *Suppose* $S \subseteq \mathbb{R}^n$ *is nonempty and convex. Let* $f \in C^1$ *on* S, *convex. Then,*

$$\boldsymbol{x}^* \text{ is a global minimum of } f \text{ on } S \quad \Longleftrightarrow \quad (2) \text{ holds.}$$

- *Proof.*
- Compare with the case $S = \mathbb{R}^n$!

Necessary optimality conditions, III: LP

- $\boldsymbol{x}^*$ *is stationary iff*

$$\underset{\boldsymbol{x} \in S}{\text{minimum}} \; \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) = 0.$$

- *Proof.*
- Method basis: given $\boldsymbol{x}_k \in S$, find out if we are stationary by minimizing $\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k)$ over $\boldsymbol{x} \in S$. In some sense, we find the $\boldsymbol{x} \in S$ which "violates optimality the most." Perform a line search in the direction from $\boldsymbol{x}_k$ towards that point. Repeat until convergence.
- Names: *Frank–Wolfe, Simplicial decomposition.* Chapter 12.

Necessary optimality conditions, II: Projection

- $\boldsymbol{x}^*$ *is stationary iff*

$$\boldsymbol{x}^* = \mathrm{Proj}_S[\boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*)].$$

- In other words, $\boldsymbol{x}^*$ is stationary if and only if a step in the direction of the steepest descent direction followed by a Euclidean projection onto S means that we have not moved at all. (If not, then we obtain a descent direction towards that projected point—basis for the *projection method* in Chapter 12.)
- *Proof.*

Necessary optimality conditions, IV: Normal cone

- If we wish to project $\boldsymbol{z} \in \mathbb{R}^n$ onto S, then the resulting (unique) projection is the vector $\boldsymbol{x}$ for which the following holds:

$$[\boldsymbol{x} - \boldsymbol{z}]^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \geq 0, \qquad \boldsymbol{y} \in S$$

  that is

$$[\boldsymbol{z} - \boldsymbol{x}]^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \leq 0, \qquad \boldsymbol{y} \in S.$$

- Interpretation: the angle between the two vectors $\boldsymbol{z} - \boldsymbol{x}$ (the direction towards the point being projected) and the vector $\boldsymbol{y} - \boldsymbol{x}$ (the direction towards any vector $\boldsymbol{y} \in S$) is $\geq 90°$. So, the projection operation has the characterization

$$[\boldsymbol{z} - \mathrm{Proj}_S(\boldsymbol{z})]^{\mathrm{T}}(\boldsymbol{y} - \mathrm{Proj}_S(\boldsymbol{z})) \leq 0, \qquad \boldsymbol{y} \in S. \tag{3}$$

- Interesting with $\boldsymbol{z} = \boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*)$:

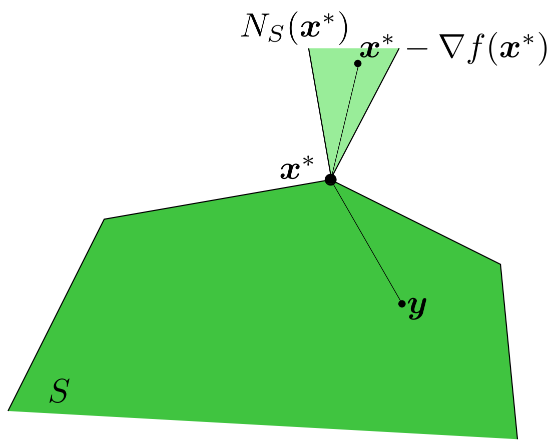


Figure 2: Normal cone characterization of a stationary point.

- Suppose $S \subseteq \mathbb{R}^n$ is closed and convex. Let $\boldsymbol{x} \in \mathbb{R}^n$. Then, the *normal cone* to S at $\boldsymbol{x}$ is the set

$$N_S(\boldsymbol{x}) := \begin{cases} \{\, \boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \leq 0, \qquad \boldsymbol{y} \in S \,\}, & \text{if } \boldsymbol{x} \in S, \\ \emptyset & \text{otherwise} \end{cases}$$

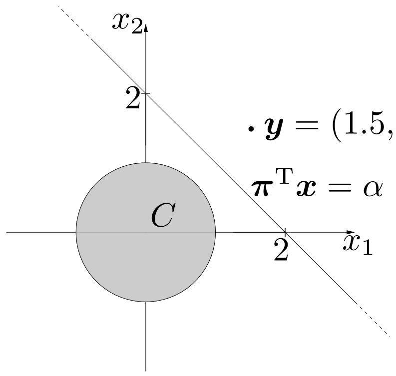
- Characterization of stationary point at $\boldsymbol{x}^*$, number IV:

$$-\nabla f(\boldsymbol{x}^*) \in N_S(\boldsymbol{x}^*) \tag{4}$$

- Geometry: the angle between $-\nabla f(\boldsymbol{x}^*)$ and any feasible direction is $\geq 90°$ ($\nexists$ feasible descent directions)
- S is a subspace $\Longrightarrow$ $\nabla f(\boldsymbol{x}^*)$ is a normal to the subspace!
- Note: $\boldsymbol{x}^*$ interior point $\Longrightarrow$ $N_S(\boldsymbol{x}^*) = \{\mathbf{0}^n\}$ ($S = \mathbb{R}^n \Longrightarrow \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n$)
- (4) to be extended to the KKT conditions

Separation Theorem revisited (proof)

- *Suppose that $C \subseteq \mathbb{R}^n$ is closed and convex, and that the point $\boldsymbol{y}$ does not lie in C. Then there exist a vector $\boldsymbol{\pi} \neq \mathbf{0}^n$ and $\alpha \in \mathbb{R}$ such that $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{y} > \alpha$ and $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x} \le \alpha$ for all $\boldsymbol{x} \in C$.*



- *Proof.*

- Set $\boldsymbol{\pi} := \boldsymbol{y} - \boldsymbol{x}^*$ and $\alpha := \boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x}^*$
- $\boldsymbol{y} \notin C$: $0 < \boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{y} - \alpha = (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{y} - (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x}^* = \|\boldsymbol{y} - \boldsymbol{x}^*\|^2$.
- $\boldsymbol{x} \in C$: $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x} \le \alpha \iff (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x} \le (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x}^* \iff (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \le 0$.
- The hyperplane is a tangent to C, the normal is $\boldsymbol{y} - \boldsymbol{x}^*$.

Non-expansiveness property of the projection operation

- Suppose $S \subseteq \mathbb{R}^n$ is closed and convex. Let $P : S \to S$ denote a vector-valued operator from S to S. We say that P is *non-expansive* if, as a result of applying the mapping P, the distance between any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in S does not increase:

$$\|P(\boldsymbol{x}) - P(\boldsymbol{y})\| \leq \|\boldsymbol{x} - \boldsymbol{y}\|, \qquad \boldsymbol{x}, \boldsymbol{y} \in S.$$

- *For every* $\boldsymbol{x} \in \mathbb{R}^n$*, its projection* $\mathrm{Proj}_S(\boldsymbol{x})$ *is uniquely defined. The operator* $\mathrm{Proj}_S : \mathbb{R}^n \to S$ *is non-expansive, and therefore in particular continuous.*
- *Proof.*

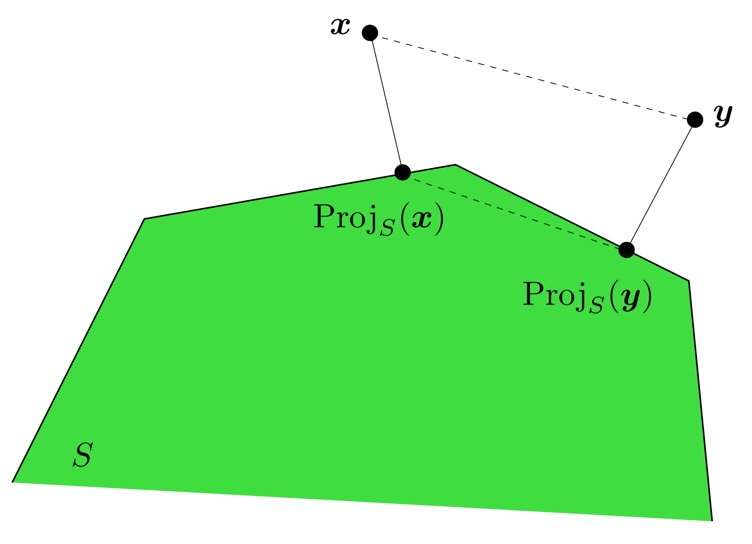


Figure 3: The projection operation is non-expansive.

Fixed-point theory, I: Contractions

Suppose $S \subseteq \mathbb{R}^n$ is closed and convex. Let $P : S \to S$ denote a vector-valued operator from S to S. We say that P is a *contraction* if, as a result of applying the mapping P, the distance between any two distinct vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in S decreases: $\exists \alpha \in [0, 1)$ such that

$$\|P(\boldsymbol{x}) - P(\boldsymbol{y})\| \leq \alpha \|\boldsymbol{x} - \boldsymbol{y}\|, \qquad \boldsymbol{x}, \boldsymbol{y} \in S$$

Fixed-point theory, II: Theorems

Let S be a nonempty, closed and convex set in $\mathbb{R}^n$.

(a) [Banach's Theorem] *Let $\boldsymbol{f}$ be a contraction mapping from S to S. Then, $\boldsymbol{f}$ has a unique fixed point $\boldsymbol{x}^* \in S$. Further, for every initial vector $\boldsymbol{x}_0 \in S$, the iteration sequence $\{\boldsymbol{x}_k\}$ defined by the fixed-point iteration*

$$\boldsymbol{x}_{k+1} := \boldsymbol{f}(\boldsymbol{x}_k), \qquad k = 0, 1, \dots,$$

converges to the unique fixed point $\boldsymbol{x}^*$. In particular,

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \leq \alpha^k \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|, \qquad k = 0, 1, \dots .$$

(b) [Brouwer's Theorem] *Let S further be bounded, and assume merely that $\boldsymbol{f}$ is continuous. Then, $\boldsymbol{f}$ has a fixed point.*

Proof [*of* (a)].

Fixed-point theory, III: Interpretation

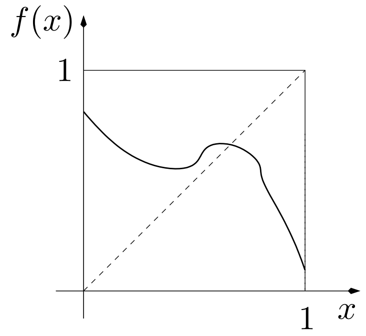


Figure 4: Let $S = [0, 1]$, and $f : S \to S$ be continuous. Brouwer's Theorem states that there exists an $x^* \in S$ with $f(x^*) = x^*$. This is the same as saying that the continuous curve starting at $(0, f(0))$ and ending at $(1, f(1))$ must pass through the line $y = x$ inside the square.

Fixed-point theory, IV: Applications

- [Raking of gravel] Suppose you wish to rake the gravel in your garden; if the area is, say, circular, then any continuous raking will leave at least one tiny stone (which one is a function of time) in the same place.
- [Maps] Suppose you have two city maps over Gothenburg, which are not of the same scale. You crumple one of them up into a loose ball and place it on top of the other map entirely within the borders of the Gothenburg region on the flat map. Then, there is a point on the crumpled map (that represents the same place in Gothenburg on both maps) that is directly over its twin on the flat map. (A more simple problem is defined by a non-crumpled map and the city of Gothenburg itself; lay down the map anywhere in Gothenburg, and at least one point on the map will lie over that exact spot in real-life Gothenburg.)

- [Stirring coffee] Stirring the contents of a (convex) coffee cup in a continuous way, no matter how long you stir, some particle (which one is a function of time) will stay in the same position as it did before you began stirring.
- [Meteorology] Even as the wind blows across the Earth there will be one location where the wind is perfectly vertical (or, perfectly calm). This fact actually implies the existence of cyclones; not to mention whorls, or crowns, in your hair no matter how you comb it. (It even bears its own name: The Hairy Ball Theorem.)
- Consider finding $x^* \in \mathbb{R}$ with $f(x^*) = 0$, where $f \in C^2$ on $\mathbb{R}$. The Newton–Raphson algorithm has an iteration formula of the form

$$x_0 \in \mathbb{R}; \qquad x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, \ldots .$$

Assume that $f'(x^*) > 0$, and prove that the algorithm locally converges since it is a contraction in a small interval around x^*.