Chapter 1

Week1

1.1. Monday

1.1.1. Course Introduction

Convex Optimization. The basic optimization model is as follows:

$$\begin{aligned} \min \quad & f(x) : \mathbb{R}^n \to \mathbb{R} \\ \text{s.t.} \quad & x \in X \subseteq \mathbb{R}^n \end{aligned} \tag{1.1}$$

For instance, the constraint X can be union of some inequality constraints:

$$X = \left\{ x \in \mathbb{R}^n \middle| f_i(x) \le b_i, \quad i = 1, \dots, m \right\}$$

We only consider **convex** problems in this course, which means that in (1.1):

- the objective function f is convex;
- the constraint set X is a convex set.

The goal of optimization is to find $x^* \in X$ such that

$$f(x^*) \le f(x), \quad \forall x \in X.$$

or determine whether such an x^* exists or not.

■ **Example 1.1** The least squares problem is a convex problem:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 \\ \text{s.t.} \quad & x \in \mathbb{R}^n \end{aligned}$$

The optimal solution can be compactly written as

$$\boldsymbol{x}^* = \boldsymbol{A}^* \boldsymbol{b}$$

where $\boldsymbol{A}^*$ denotes the pseudo inverse of $\boldsymbol{A}$. ■

■ **Example 1.2** The linear programming is also a convex problem:

$$\begin{aligned} \min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} \le b_i, \quad i = 1,2,\dots,m. \end{aligned}$$

The linear programming is a special case of the conic programming. We will cover conic programming in this course. ■

Definition 1.1 [Line Segments] The line segment between $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ is defined as

$$\{\theta\boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \mid \theta \in [0,1]\} = \{\boldsymbol{x}_2 + \theta(\boldsymbol{x}_1 - \boldsymbol{x}_2) \mid \theta \in [0,1]\}$$

■

Definition 1.2 [Convex Set in $\mathbb{R}^n$] The set $X \subseteq \mathbb{R}^n$ is convex if for any $x_1, x_2 \in X$,

$$\theta x_1 + (1-\theta)x_2 \in X, \qquad \forall \theta \in [0,1].$$

■

Here are two examples for convex sets over matrices:

$$\mathcal{S}^n = \{\boldsymbol{W} \in \mathbb{R}^n \times \mathbb{R}^n \mid \boldsymbol{W} = \boldsymbol{W}^{\mathrm{T}}\}, \quad \mathcal{S}_+^n = \{\boldsymbol{W} \in \mathcal{S}^n \mid \boldsymbol{W} \succeq 0\},$$

R Sometimes the convex set could be a collection of matrices, but we can vectorize a matrix into a vector.

Definition 1.3 [Convex Function] A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if

- $\text{dorm}(f)$ is a convex set, and
- $\forall x, y \in \text{dorm}(f)$, it holds that for $\theta \in [0,1]$,

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y),$$

which means that the secant line betweeen any two points is above the function.

Proposition 1.1 $f(x) = \max_i f_i(x)$ is convex if $f_i(x)$ is convex for each i.

Proof. For any $\theta \in [0,1]$, the following inequality holds:

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max_i f_i(\theta x + (1-\theta)y) \\ &\le \max_i \theta f_i(x) + (1-\theta)f_i(y) \\ &\le \max_i \theta f_i(x) + \max_i (1-\theta)f_i(y) \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

■

■ **Example 1.3** Define the function

$$\begin{aligned} & f : \mathbb{S}^n \to \mathbb{R} \\ \text{with} \quad & f(\boldsymbol{X}) = \lambda_{\max}(\boldsymbol{X}) \triangleq \max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v} \end{aligned}$$

This function is convex since f can be written as the maximization of a collection of affines (in terms of $\boldsymbol{X}$):

$$f(\boldsymbol{X}) = \max_{\|\boldsymbol{v}\|=1} f_{\boldsymbol{v}}(\boldsymbol{X}), \qquad \text{with } f_{\boldsymbol{v}}(\boldsymbol{X}) = \boldsymbol{v}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{v}.$$

1.2. Friday

Before we give a proof of Schroder-Bernstein theorem, we'd better review the definitions for one-to-one mapping and onto mapping.

Definition 1.4 [One-to-One/Onto Mapping] If $f : A \mapsto B$, then

- f is said to be **onto** mapping if

$$\forall b \in B, \exists a \in A \text{ s.t. } f(a) = b;$$

- f is said to be **one-to-one** mapping if

$$\forall a, b, \in A, f(a) = f(b) \implies a = b.$$

The Fig.(1.1) shows the examples of one-to-one/onto mappings.

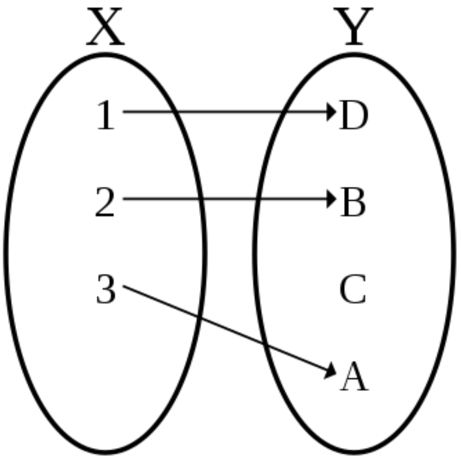
1.2.1. Proof of Schroder-Bernstein Theorem

Before the proof, note that in this lecture we abuse the notation fg to denote the composite function $f \circ g$, but in the future fg will refer to other meanings.

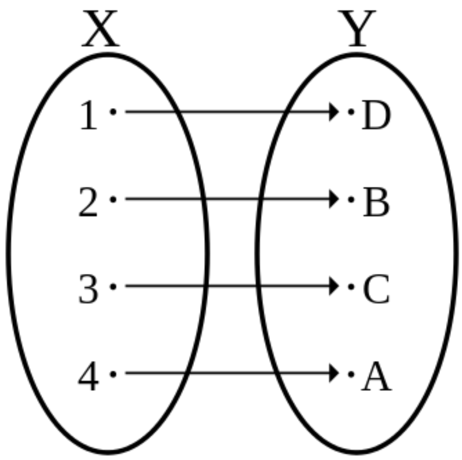
Intuition from Fig.(1.2). The proof for this theorem is constructive. Firstly Fig.(1.2) gives us the intuition of the proof for this theorem. Let $f : A \mapsto B$ and $g : B \mapsto A$ be two one-to-one mappings, and D, C are the image from A, B respectively. Note that

> if the set $B \setminus D$ is empty, then $D = B = f(A)$ with f being the one-to-one mapping, which implies f is one-to-one onto mapping. In this case the proof is complete.

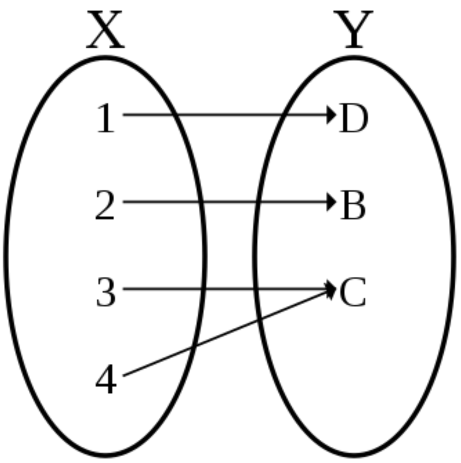
Hence it suffices to consider the case $B \setminus D$ is non-empty. Thus $B \setminus D$ is the "**trouble-maker**". To construct a one-to-one onto mapping from A, we should study the subset $g(B \setminus D)$ of A (which can also be viewed as a *trouble-maker*). Moreove, we should study



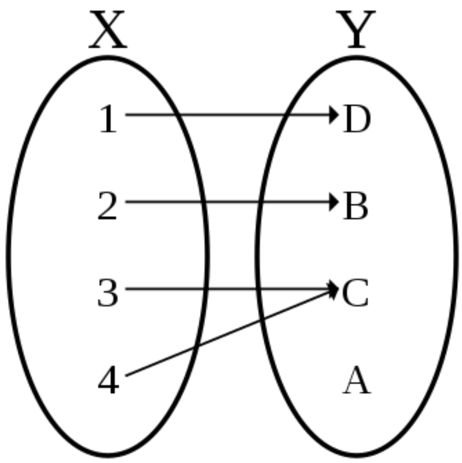
(a) A one-to-one but not onto mapping



(b) A one-to-one onto mapping



(c) A onto but not one-to-one mapping



(d) Neither a one-to-one nor onto mapping

Figure 1.1: Illustrations of one-to-one/onto mappings

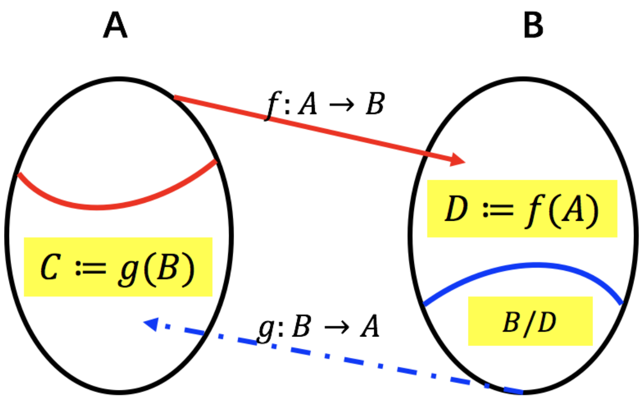


Figure 1.2: Illustration of Schroder-Bernstein Theorem

the subset $gf[g(B \setminus D)]$ (which is also a *trouble-maker*)... so on and so forth. Therefore, we should study the *union of these trouble makers*, i.e., we define

$$A_1 := g(B \setminus D), \quad A_2 := gf(A_1), \quad \cdots, \quad A_n := gf(A_{n-1}),$$

Then we study the union of infinite sets

$$S := A_1 \bigcup A_2 \bigcup \cdots \bigcup A_n \bigcup \cdots$$

Define

$$F(a) = \begin{cases} f(a), & a \in A \setminus S \\ g^{-1}(a), & a \in S \end{cases}$$

We claim that $F : A \mapsto B$ is one-to-one onto mapping.

F is onto mapping. Given any element $b \in B$, it follows two cases:

1. $g(b) \in S$. It implies $F(g(b)) = g^{-1}(g(b)) = b$.
2. $g(b) \notin S$. It implies $b \in D$, since otherwise $b \in B \setminus D \implies g(b) \in g(B \setminus D) \subseteq S$, which is a contradiction. $b \in D$ implies that $\exists a \in A$ s.t. $f(a) = b$.

   Then we study the relationship between $gf(S)$ and S. Verify by yourself that

   $$S = g(B \setminus D) \bigcup gf(S)$$

   With this relationship, we claim $a \notin S$, since otherwise $a \in S \implies gf(a) \in S$, but $gf(a) = g(b) \notin S$, which is a contradiction.

   Hence, $F(a) = f(a) = b$.

Hence, for any element $b \in B$, we can find a element from A such that the mapping for which is equal to b, i.e., F is onto mapping.

F is one-to-one mapping. Assume not, verify by yourself that the only possibility is that $\exists a_1 \in A \setminus S$ and $a_2 \in S$ such that $F(a_1) = F(a_2)$, i.e., $f(a_1) = g^{-1}(a_2)$, which follows

$$gf(a_1) = a_2 \in S = A_1 \bigcup A_2 \bigcup \cdots \tag{1.2}$$

We claim that Eq.(1.2) is false. Note that $gf(a_1) \notin A_1 := g(B \setminus D)$, since otherwise $f(a_1) \in B \setminus D$, which is a contradiction; note that $gf(a_1) \notin A_2$, since otherwise $gf(a_1) \in gfg(B \setminus D) \implies a_1 \in g(B \setminus D) = A_1 \subseteq S$, which is a contradiction.

Applying the similar trick, we wil show that $gf(a_1) \notin A_k$ for $k \geq 1$. Hence, Eq.(1.2) is false, the proof is complete.

■ **Example 1.4** Given two sets $A := (0,1]$ and $B := [0,1)$. Now we apply the idea in the proof above to construct a one-to-one onto mapping from A to B:

- Firstly we construct two one-to-one mappings:

$$\begin{aligned} f :&A \mapsto B & g :&B \mapsto A \\ f(x) &= \frac{1}{2}x & g(x) &= x \end{aligned}$$

- It follows that $B \setminus D = (\frac{1}{2}, 1)$, $gf(B \setminus D) = (\frac{1}{4}, 1)$, so on and so forth.

$$S = (\frac{1}{2}, 1) \bigcup (\frac{1}{4}, 1) \bigcup \cdots$$

- Hence, the one-to-one onto mapping we construct is

$$F(x) = \begin{cases} \frac{1}{2}x, & x \in A \setminus S \\ x, & x \in S \end{cases}$$

- Conversely, to construct the inverse mapping, we define

$$f(x) = x \quad g(x) = \tfrac{1}{2}x$$

- It follows that $D = (0,1)$, $B \setminus D = \{1\}$. Then

$$S = \left\{\frac{1}{2}\right\} \bigcup \cdots = \left\{\frac{1}{2}, \frac{1}{4}, \cdots\right\}$$

- Hence, the function we construct for inverse mapping is

$$F(x) = \begin{cases} x, & x \neq \frac{1}{2^m} \\ 2x, & x = \frac{1}{2^m} \end{cases} \qquad (m = 1, 2, 3, \dots)$$

■

1.2.2. Connectedness of Real Numbers

There are two approaches to construct real numbers. Let's take $\sqrt{2}$ as an example.

1. The first way is to use **Dedekind Cut**, i.e., every non-empty subset has a least upper bound. Therefore, $\sqrt{2}$ is actually the least upper bound of a non-empty subset

$$\{x \in \mathbb{Q} \mid x^2 < 2\}.$$

2. Another way is to use **Cauchy Sequence**, i.e., every Cauchy sequence is convergent. Therefore, $\sqrt{2}$ is actually the limit of the given sequence of decimal approximations below:

$$\{1, 1.4, 1.41, 1.414, 1.4142, \dots\}$$

We will use the second approach to define real numbers. Every real number r essentially represents a collection of cauchy sequences with limit r, i.e.,

$$r \in \mathbb{R} \implies \left\{ \{x_n\}_{n=1}^{\infty} \middle| \lim_{n \to \infty} x_n = r \right\}$$

Let's give a formal definition for cauchy sequence and a formal definition for real number.

Definition 1.5 [Cauchy Sequence]

- Any sequence of rational numbers $\{x_1, x_2, \cdots\}$ is said to be a **cauchy sequence** if for every $\epsilon > 0$, $\exists N$ s.t. $|x_n - x_m| < \epsilon$, $\forall m, n \geq N$

- Two cauchy sequences $\{x_1, x_2, \dots\}$ and $\{y_1, y_2, \dots\}$ are said to be **equivalent** if for every $\epsilon > 0$, there $\exists N$ s.t. $|x_n - y_n| < \epsilon$ for $\forall n \geq N$.
- A real number is a **collection** of **equivalent** cauchy sequences. It can be represented by a cauchy sequence:

$$x \in \mathbb{R} \sim \{x_1, x_2, \dots, x_n, \dots\},$$

where x_j is a rational number.

R Let $\xi_{\mathbb{Q}}$ denote a collection of any cauchy sequences. Then once we have equivalence relation, the whole collection $\xi_{\mathbb{Q}}$ is partitioned into several disjoint subsets, i.e., equivalence classes. Hence, the real number space $\mathbb{R}$ are the equivalence classes of $\xi_{\mathbb{Q}}$.

The real numbers are well-defined, i.e., given two real numbers $x \sim \{x_1, x_2, \dots\}$ $y \sim \{y_1, y_2, \dots\}$, we can define add and multiplication operator.

$$x + y \sim \{x_1 + y_1, x_2 + y_2, \dots\}$$
$$x \cdot y \sim \{x_1 \cdot y_1, x_2 \cdot y_2, \dots\}$$

We will show how to define $x > 0$ in next lecture, this construction essentially leads to the lemma below:

Proposition 1.2 $\mathbb{Q}$ are dense in $\mathbb{R}$.

In the next lecture we will also show the completeness of $\mathbb{R}$:

Theorem 1.1 $\mathbb{R}$ is complete, i.e., every cauchy sequence of real numbers converges.

Recommended Reading:

Prof. Katrin Wehrheim, MIT Open Course, Fall 2010, Analysis I Course Notes, Online avaiable:

https://ocw.mit.edu/courses/mathematics
/18-100b-analysis-i-fall-2010/readings-notes/MIT18_100BF10_Const_of_R.pdf