MATH11111: Fundamentals of Optimization **Fall 2022**

Lecture 1 Introduction to Operational Research and Optimization

Lecturer: E. Alper Yıldırım *Week: 1*

1.1 Outline

- What is Operational Research (OR)?
- What is Optimization?
- Mathematical Model and Terminology
- Review Problems

1.2 Operational Research and Optimization

Definition 1.1 Operational Research (OR) *is a branch of applied mathematics that employs analytical models and tools for solving decision problems arising from many different fields such as energy, healthcare, engineering, finance, machine learning, and transportation.*

Definition 1.2 Optimization *is concerned with finding the* best solution *of a given decision problem among a set of* candidate solutions*.*

Remark 1.1 *Optimization is one of the fundamental analytical tools in operational research.*

Other tools of OR include dynamic programming, stochastic modelling, simulation, queueing theory, and inventory control models.

1.2.1 Ingredients of an Optimization Problem

An optimization problem consists of the following four ingredients:

- A set of controllable inputs: Set of quantities whose best values we wish to compute (Decision Variables)
- A set of uncontrollable inputs: Set of quantities whose values cannot be changed (Parameters)
- A description of the set of candidate solutions: Usually expressed by some functional relations (Feasible Region)
- A measure of the *goodness* of a candidate solution: Usually measured by a real-valued function (Objective Function)

1.2.2 An Example

Problem 1.3 *What is the shortest (fastest) route from the King's Buildings to the Edinburgh Airport?*

- Decision variables: Which route to choose
- Parameters: Length (travel time) of each route
- Candidate solutions: Set of all possible routes from the King's Buildings to the Edinburgh Airport
- Objective function: Length (travel time) of the route from the King's Buildings to the Edinburgh Airport

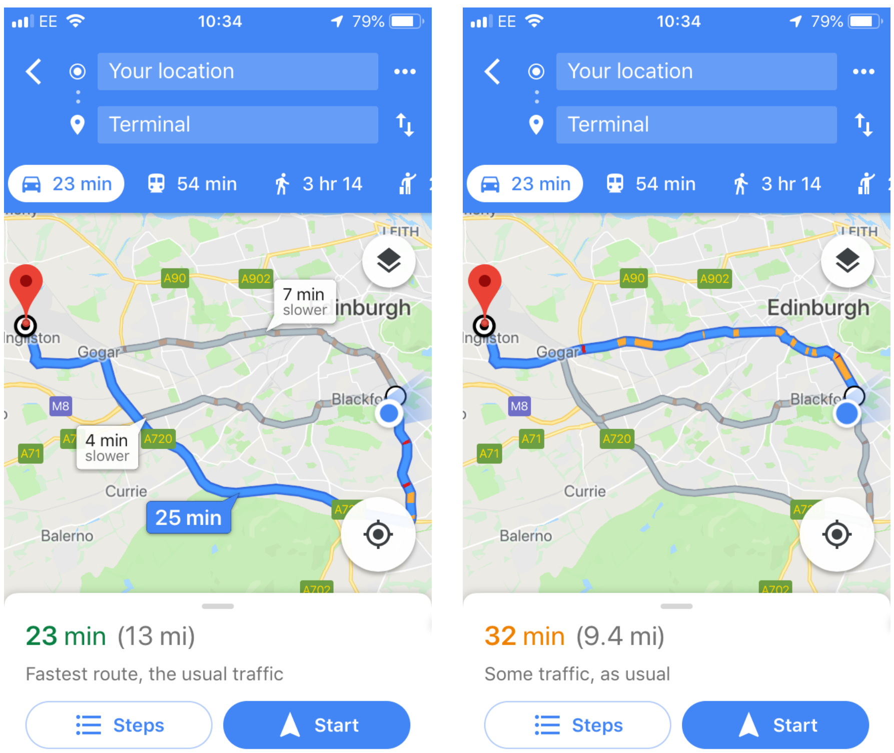

Figure 1.1: The fastest (on the left) and the shortest route (on the right) from King's Buildings to the Edinburgh Airport

As illustrated by Figure 1.1, the best solution depends on the objective function, and one may get different best solutions for different objective functions.

Optimization problems arise in numerous applications, some of which are listed below:

- Manufacturing (cost minimization/profit maximization)
- Wireless networks (throughput maximization/interference minimization)
- Finance (portfolio optimization)

- Physical systems (energy minimization/entropy maximization)
- Machine learning (classification error minimization)
- Transportation and logistics (cost/emission minimization)
- Medicine (treatment optimization/drug design/personalized medicine)
- Many other applications

1.2.3 Mathematical Model and Representation

The mathematical model of an optimization problem consists of the following four ingredients:

- **Decision variables:** Set of quantities whose best values we wish to compute (usually denoted by $x \in \mathbb{R}^n$)
- **Parameters:** Needed to identify the feasible region and the objective function
- **Feasible Region:** $x \in \mathcal{S}$, where $\mathcal{S} \subseteq \mathbb{R}^n$
- **Objective function:** $f : \mathbb{R}^n \to \mathbb{R}$

An optimization problem can therefore be expressed as

$$\text{(P)} \quad \min\{f(x) : x \in \mathcal{S}\}$$

or alternatively as

$$\begin{array}{lll} \text{(P)} & \min & f(x) \\ & \text{s.t. (subject to)} & \\ & & x \in \mathcal{S} \end{array}$$

The first notation is more compact and is, in fact, a set notation. The latter one is more explicit. However, both representations are equivalent and will be used interchangeably throughout the course.

It is customary to use minimization (as opposed to maximization) in optimization problems. However, from a mathematical point of view, one can be converted into the other one easily since

$$f(x^1) \geq f(x^2) \quad \text{if and only if} - f(x^1) \leq -f(x^2),$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $x^1 \in \mathbb{R}^n$, and $x^2 \in \mathbb{R}^n$. Therefore, by negating a function, the ranking of all objective function values is reversed. It follows that we may always consider a minimization problem without loss of generality. This is stated more formally in the following remark.

Remark 1.2 *Maximizing $g(x)$ is equivalent to minimizing $-g(x)$. Therefore, every maximization problem can be converted to an equivalent minimization problem by simply negating the objective function.*

1.2.4 Terminology

Consider an optimization problem given by

$$\text{(P)} \quad \min\{f(x) : x \in \mathcal{S}\}$$

- The set $\mathcal{S}$ of candidate solutions is called the *feasible region* or the *feasible set*.
- Any solution $\hat{x} \in \mathcal{S}$ is called a *feasible solution*.
- A feasible solution $x^* \in \mathcal{S}$ is called an *optimal solution* of (P) if

$$f(x^*) \leq f(\hat{x}), \quad \forall \hat{x} \in \mathcal{S}.$$

  We replace $\leq$ by $\geq$ for a maximization problem.

- The set of all optimal solutions of (P) is denoted by $\mathcal{S}^*$, i.e.,

$$\mathcal{S}^* = \{x^* \in \mathcal{S} : f(x^*) \leq f(\hat{x}), \quad \forall \hat{x} \in \mathcal{S}\}$$

- $z^* \in \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ is the *optimal value* of (P) if $z^* \leq f(\hat{x})$ for all $\hat{x} \in \mathcal{S}$ and there exists a sequence $x^k \in \mathcal{S},\ k = 1, 2, \ldots$ such that $f(x^k) \to z^*$ as $k \to \infty$ (i.e., for a minimization problem, the optimal value is the largest lower bound on the objective function values of all feasible solutions, known as the *infimum* of the objective function values of all feasible solutions).
- If $\mathcal{S}^* \neq \emptyset$, then $z^* = f(x^*)$ for any $x^* \in \mathcal{S}^*$. In this case, we say that the optimal value is *attained*, i.e., there is at least one feasible solution whose objective function value is the same as the optimal value.
- (P) is said to be an *unbounded problem* if there exists a sequence $x^k \in \mathcal{S},\ k = 1, 2, \ldots$ such that $f(x^k) \to -\infty$. In this case, $\mathcal{S}^* = \emptyset$, and we define $z^* = -\infty$. (For a maximization problem, we simply replace $-\infty$ by $+\infty$.)
- If $\mathcal{S} = \emptyset$, then (P) is said to be an *infeasible problem*. In this case, $\mathcal{S}^* = \emptyset$, and we define $z^* = +\infty$. (For a maximization problem, we simply replace $+\infty$ by $-\infty$.)
- If (P) is neither infeasible nor unbounded, then $-\infty < z^* < +\infty$. In this case, $\mathcal{S}^*$ may still be the empty set if there is no feasible solution $\hat{x} \in \mathcal{S}$ such that $f(\hat{x}) = z^*$.

Remarks

1. Every optimization problem has an optimal value $z^* \in \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$, even if it is infeasible or unbounded. Note that $-\infty$ and $+\infty$ are **not** real numbers. The set $\mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ is referred to as the *extended real numbers*.
2. For some optimization problems, the optimal value may be finite (i.e., a real number), but it may not be attained. Here is a simple example:

$$\min \left\{ \frac{1}{x} : x \geq 1 \right\}$$

   Note that $1/x \geq 0$ for each $x \geq 1$ and if we define $x^k = k,\ k = 1, 2, \ldots$, then $f(x^k) = 1/k \to 0$ as $k \to \infty$. Therefore, the optimal value is given by $z^* = 0$. However, no feasible solution achieves the optimal value. In this example, the optimal value is finite but is not attained.

3. By the previous example, the feasible region of an optimization problem may be unbounded and yet the optimal value may be finite (i.e., the optimization problem is not unbounded). Please pay attention to the difference between the unboundedness of an optimization problem and the unboundedness of its feasible region.

Exercises

Question 1.1 *What are the four main ingredients in an optimization problem?*

Question 1.2 *Consider the following optimization problem:*

$$\min\{x : x \geq 1\}$$

1. *What are the decision variables?*
2. *What is the objective function?*
3. *What is the feasible region?*
4. *What is the optimal value?*
5. *Find the set of all optimal solutions.*

Question 1.3 *Consider the following optimization problem:*

$$\min\{0 : x \geq 1\}$$

1. *What are the decision variables?*
2. *What is the objective function?*
3. *What is the feasible region?*
4. *What is the optimal value?*
5. *Find the set of all optimal solutions.*

Question 1.4 *Consider the following optimization problem:*

$$\min\{-x : x \geq 1\}$$

1. *What are the decision variables?*
2. *What is the objective function?*
3. *What is the feasible region?*
4. *What is the optimal value?*
5. *Find the set of all optimal solutions.*

Question 1.5 *Consider the following optimization problem:*

$$\min\left\{\frac{1}{x} : x \geq 1\right\}$$

1. *What are the decision variables?*
2. *What is the objective function?*
3. *What is the feasible region?*
4. *What is the optimal value?*
5. *Find the set of all optimal solutions.*

Question 1.6 *Consider the following optimization problem:*

$$\min \left\{ x : x^2 \leq -1 \right\}$$

1. *What are the decision variables?*
2. *What is the objective function?*
3. *What is the feasible region?*
4. *What is the optimal value?*
5. *Find the set of all optimal solutions.*

MATH11111: Fundamentals of Optimization **Fall 2022**

Lecture 2 — Constrained and Unconstrained Optimization and Introduction to Linear Programming

Lecturer: E. Alper Yıldırım *Week: 1*

2.1 Outline

- Constrained vs Unconstrained Optimization
- Linear Functions
- Linear Programming
- Review Problems

2.2 Constrained vs Unconstrained Optimization Problem

Recall our generic optimization problem:

$$(\mathrm{P}) \quad \min \{f(x) : x \in \mathcal{S}\},$$

where $x \in \mathbb{R}^n$ denotes the decision variables, $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function and $\mathcal{S} \subseteq \mathbb{R}^n$ is the feasible region.

If $\mathcal{S} = \mathbb{R}^n$, then (P) is called an *unconstrained optimization problem*. In this case, any $x \in \mathbb{R}^n$ is a feasible solution.

Otherwise, $\mathcal{S} \subset \mathbb{R}^n$, i.e., $\mathcal{S}$ is a proper subset of $\mathbb{R}^n$ and (P) is called a *constrained optimization problem*. In this case, the decision variable x is *constrained* to belong to a smaller set $\mathcal{S}$.

2.3 Constrained Optimization

In constrained optimization, the feasible region $\mathcal{S}$ is usually expressed by functional relations:

$$\mathcal{S} = \{x \in \mathbb{R}^n : g_i(x) \geq b_i,\ i \in M_1; \quad \ell_i(x) \leq b_i,\ i \in M_2; \quad h_i(x) = b_i,\ i \in M_3\},$$

where M_1, M_2, and M_3 are finite index sets, some of which may possibly be empty; $g_i : \mathbb{R}^n \to \mathbb{R},\ i \in M_1$; $\ell_i : \mathbb{R}^n \to \mathbb{R},\ i \in M_2$; $h_i : \mathbb{R}^n \to \mathbb{R},\ i \in M_3$; and $b_i \in \mathbb{R}$ for each $i \in M_1 \cup M_2 \cup M_3$.

Each of the functional relations $g_i(x) \geq b_i,\ i \in M_1$; $\ell_i(x) \leq b_i,\ i \in M_2$; $h_i(x) = b_i,\ i \in M_3$ is called a *constraint*. Note that, for a given $\hat{x} \in \mathbb{R}^n$, $\hat{x} \in \mathcal{S}$ if and only if $\hat{x}$ satisfies *all* of the above constraints *simultaneously*. On the other hand, $\hat{x} \notin \mathcal{S}$ if and only if $\hat{x}$ violates at least one of these constraints.

In the constraints, we generally prefer to have weak inequalities (i.e., $\leq$ and $\geq$) as opposed to strict inequalities (i.e., $<$ and $>$) since the feasible region $\mathcal{S}$ may otherwise not contain its boundary points (i.e., may not be a closed set).

Example 2.1. *If $\mathcal{S} = \{x \in \mathbb{R} : x > 1\}$, then $1 \notin \mathcal{S}$. If we minimize x over $\mathcal{S}$, then the optimal value is given by $z^* = 1$ but $\mathcal{S}^* = \emptyset$ since it is not attained. Therefore, the optimal value is finite but not attained.*

2.3.1 Representation of Constrained Optimization Problems

A constrained optimization problem can be represented using the more compact set notation:

$$(P) \quad \min \{f(x) : g_i(x) \geq b_i, \ i \in M_1; \quad \ell_i(x) \leq b_i, \ i \in M_2; \quad h_i(x) = b_i, \ i \in M_3\},$$

where $M_1, M_2,$ and M_3 are finite index sets and each of $g_i : \mathbb{R}^n \to \mathbb{R}, \ i \in M_1$, $\ell_i : \mathbb{R}^n \to \mathbb{R}, \ i \in M_2$, and $h_i : \mathbb{R}^n \to \mathbb{R}, \ i \in M_3$ is a constraint, and $b_i \in \mathbb{R}$ for each $i \in M_1 \cup M_2 \cup M_3$.

Alternatively, we may use a more explicit representation:

$$\begin{array}{lllll} \text{(P)} & \min & f(x) & & \\ & \text{s.t.} & & & \\ & & g_i(x) & \geq & b_i, \quad i \in M_1, \\ & & \ell_i(x) & \leq & b_i, \quad i \in M_2, \\ & & h_i(x) & = & b_i, \quad i \in M_3, \end{array}$$

Note that $\hat{x} \in \mathcal{S}$ (i.e., a feasible solution of (P)) if and only if $g_i(\hat{x}) \geq b_i$ **for each** $i \in M_1$ **and** $\ell_i(\hat{x}) \leq b_i$ **for each** $i \in M_2$ **and** $h_i(x) = b_i$ **for each** $i \in M_3$ (i.e., $\hat{x}$ should satisfy each and every constraint). Therefore, by adding more constraints to (P), the feasible region $\mathcal{S}$ cannot get larger (i.e., either remains unchanged or shrinks).

2.4 Linear Functions

Definition 2.1. *A function $\rho : \mathbb{R}^n \to \mathbb{R}$ is a linear function if*

1. *for all $x \in \mathbb{R}^n$ and for all $y \in \mathbb{R}^n$, $\rho(x + y) = \rho(x) + \rho(y)$;*
2. *for all $x \in \mathbb{R}^n$ and for all $\alpha \in \mathbb{R}$, $\rho(\alpha x) = \alpha \rho(x)$.*

Remark 2.1. *For every linear function $\rho : \mathbb{R}^n \to \mathbb{R}$, we have $\rho(\mathbf{0}) = 0$ by Property 2, where $\mathbf{0} \in \mathbb{R}^n$ denotes the n-dimensional vector of all zeroes.*

2.4.1 Characterisation of Linear Functions

The next proposition gives a complete characterisation of linear functions.

Proposition 2.1. *A function $\rho : \mathbb{R}^n \to \mathbb{R}$ is a linear function if and only if there exists a vector $a \in \mathbb{R}^n$ such that*

$$\rho(x) = a^T x = \sum_{j=1}^{n} a_j x_j.$$

Proof. $\Leftarrow$: If $\rho(x) = a^T x$, then $\rho(x + y) = a^T(x + y) = a^T x + a^T y = \rho(x) + \rho(y)$ for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. Similarly, $\rho(\alpha x) = a^T(\alpha x) = \alpha a^T x = \alpha \rho(x)$ for any $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

⇒: Let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a linear function and let $x \in \mathbb{R}^n$. Then, $x = \sum_{j=1}^{n} x_j e^j$. Therefore, $\rho(x) = \rho\left(\sum_{j=1}^{n} x_j e^j\right) = \sum_{j=1}^{n} x_j \rho(e^j)$. Therefore, define $a_j = \rho(e^j),\ j = 1, \ldots, n$ and we are done. □

2.5 Linear Programming Problem

Recall our generic constrained optimization problem:

$$\begin{array}{lll} \text{(P)} & \min & f(x) \\ & \text{subject to (s.t.)} & \\ & & \begin{array}{rcll} g_i(x) & \geq & b_i, & i \in M_1, \\ \ell_i(x) & \leq & b_i, & i \in M_2, \\ h_i(x) & = & b_i, & i \in M_3. \end{array} \end{array}$$

Definition 2.2. *(P) is called a* linear programming problem *if each of* $f : \mathbb{R}^n \to \mathbb{R},\ g_i : \mathbb{R}^n \to \mathbb{R},\ i \in M_1$, $\ell_i : \mathbb{R}^n \to \mathbb{R},\ i \in M_2$, *and* $h_i : \mathbb{R}^n \to \mathbb{R},\ i \in M_3$ *is a linear function.*

By Proposition 2.1, $\rho : \mathbb{R}^n \to \mathbb{R}$ is a linear function if and only if there exists a vector $a \in \mathbb{R}^n$ such that $\rho(x) = a^T x = \sum_{j=1}^{n} a_j x_j$.

A linear programming problem can therefore be represented as follows:

$$\begin{array}{lll} \text{(P)} & \min & c^T x \\ & \text{subject to (s.t.)} & \\ & & \begin{array}{rcll} (a^i)^T x & \geq & b_i, & i \in M_1, \\ (a^i)^T x & \leq & b_i, & i \in M_2, \\ (a^i)^T x & = & b_i, & i \in M_3, \end{array} \end{array}$$

where $c \in \mathbb{R}^n$ and $a^i \in \mathbb{R}^n$ for each $i \in M_1 \cup M_2 \cup M_3$, and $b_i \in \mathbb{R}$ for each $i \in M_1 \cup M_2 \cup M_3$. Note that some of these index sets may possibly be empty.

Exercises

Question 2.1. *What is the difference between constrained and unconstrained optimization?*

Question 2.2. *Consider the constrained optimization problem* $\min\{f(x) : x \in \mathcal{S}\}$, *where* $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$, *and* $\mathcal{S} \subset \mathbb{R}^n$ *is given by*

$$\mathcal{S} = \{x \in \mathbb{R}^n : g_i(x) \geq b_i,\ i \in M_1; \quad \ell_i(x) \leq b_i,\ i \in M_2; \quad h_i(x) = b_i,\ i \in M_3\}.$$

(i) How do the feasible region $\mathcal{S}$ *and the optimal value* z^* *change if we add a new constraint?*

(ii) How do the feasible region $\mathcal{S}$ *and the optimal value* z^* *change if we remove an existing constraint?*

Question 2.3. *For each of the following functions $f : \mathbb{R}^2 \to \mathbb{R}$, determine whether f is a linear function or not.*

- $f(x) = 3x_1 - 4x_2$
- $f(x) = 3|x_1| - 4x_2$
- $f(x) = 3x_1^2 - 4x_2$
- $f(x) = \max\{3x_1, -4x_2\}$

Question 2.4. *For each of the following optimization problems, determine whether it is a linear programming problem or not.*

- $\min\{x_1^2 + x_2 : x_1 - x_2 \geq 5, \quad x_1 + 2x_2 = 3\}$
- $\min\{|x_1| - x_2 : 3x_1 + 2x_2 \leq -1, \quad x_1 \geq 0\}$
- $\min\{x_1 - x_2 : 3x_1 + 2x_2 \leq -1, \quad x_1^2 + x_2^2 \geq 1\}$
- $\min\{x_1 - 3x_2 : x_1 = 3, \quad x_2 = -2\}$

MATH11111: Fundamentals of Optimization **Fall 2022**

Lecture 3 — Convex Sets and Convex Functions

Lecturer: E. Alper Yıldırım *Week: 1*

3.1 Outline

- Convex Sets
- Convex Functions
- Concave Functions
- Review Problems

3.2 Convex Sets

Definition 3.1. *Let $\mathcal{C} \subseteq \mathbb{R}^n$. We say that $\mathcal{C}$ is a* convex set *if $\forall x \in \mathcal{C}$, $\forall y \in \mathcal{C}$, $\forall \lambda \in [0,1]$, we have $\lambda x + (1-\lambda) y \in \mathcal{C}$. A set that is not convex is said to be* nonconvex.

Geometrically, a set $\mathcal{C} \subseteq \mathbb{R}^n$ is a convex set if and only if, for every two points $x \in \mathcal{C}$ and $y \in \mathcal{C}$, the line segment that joins these two points is entirely contained in $\mathcal{C}$. It is nonconvex if and only if there exist two points $x \in \mathcal{C}$ and $y \in \mathcal{C}$ such that the line segment that joins these two points is not entirely contained in $\mathcal{C}$. Figure 3.1 depicts an example of a convex set and a nonconvex set.

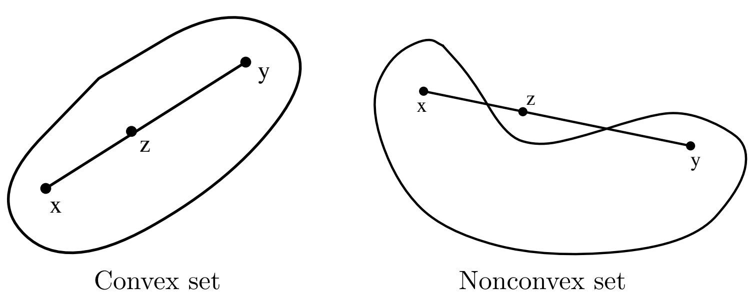

Figure 3.1: An example of a convex set (on the left) and a nonconvex set (on the right)

Remarks

1. Note that $\emptyset$ and $\mathbb{R}^n$ are both convex sets as the definition is trivially satisfied in both cases.
2. Any set $\mathcal{C} \subseteq \mathbb{R}^n$ that consists of a single element (i.e., $\mathcal{C} = \{\hat{x}\}$, where $\hat{x} \in \mathbb{R}^n$) is a convex set since the definition is trivially satisfied.

3. The intersection of any collection of convex sets is a convex set, i.e., convexity is preserved under taking intersections.
4. However, convexity is not necessarily preserved under taking unions.

3.3 Convex Functions

Definition 3.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We say that f is a* convex function *if $\forall x \in \mathbb{R}^n$, $\forall y \in \mathbb{R}^n$, $\forall \lambda \in [0, 1]$, we have*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

If $n = 1$, a function $f : \mathbb{R} \to \mathbb{R}$ is a convex function if and only if, for every two points $x \in \mathbb{R}$ and $y \in \mathbb{R}$, the line segment that joins the two points $(x, f(x)) \in \mathbb{R} \times \mathbb{R}$ and $(y, f(y)) \in \mathbb{R} \times \mathbb{R}$ lies on or above the graph of the function between x and y. It is a nonconvex function if and only if there exist two points $x \in \mathbb{R}$ and $y \in \mathbb{R}$ that violate this property. Figure 3.2 illustrates an example of a convex function and a nonconvex function.

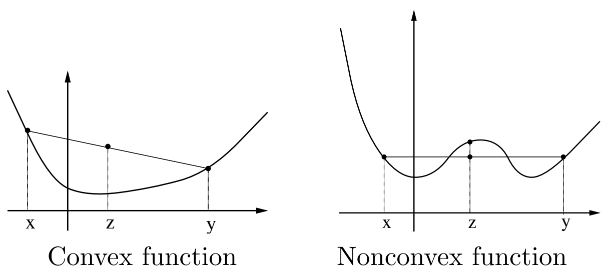

Figure 3.2: An example of a convex function (on the left) and a nonconvex function (on the right)

For a function $f : \mathbb{R}^n \to \mathbb{R}$, where $n > 2$, one can define another function $g : \mathbb{R} \to \mathbb{R}$ by

$$g(\lambda) = f(\lambda x + (1 - \lambda)y),$$

where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. Then, you can easily verify that the definition of convexity of the function $f : \mathbb{R}^n \to \mathbb{R}$ is equivalent to the convexity of the one-dimensional function $g : \mathbb{R} \to \mathbb{R}$ between 0 and 1 for every two points $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. Therefore, Figure 3.3 essentially captures the definition of convexity for any value of n.

3.4 Epigraphs

Definition 3.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. The* epigraph *of f, denoted by $epi(f) \subseteq \mathbb{R}^{n+1}$, is the region above the graph of the function f, i.e.,*

$$epi(f) = \{(x, z) \in \mathbb{R}^n \times \mathbb{R} : x \in \mathbb{R}^n, \quad z \geq f(x)\}.$$

Note that the epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a subset of $\mathbb{R}^{n+1}$. For $n = 1$, this is easy to visualise as illustrated by Figure 3.3.

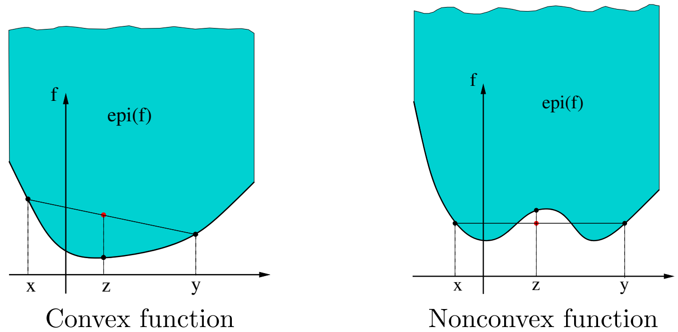

Figure 3.3: An example of the epigraph of a convex function (on the left) and the epigraph of a nonconvex function (on the right)

3.5 Convex Functions and Convex Sets

Proposition 3.1. *$f : \mathbb{R}^n \to \mathbb{R}$ is a convex function if and only if epi$(f) \in \mathbb{R}^{n+1}$ is a convex set.*

Proof. $\Rightarrow$: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $(x, z_1) \in \text{epi}(f)$ and $(y, z_2) \in \text{epi}(f)$. Therefore, $f(x) \leq z_1$ and $f(y) \leq z_2$. Then, for any $\lambda \in [0, 1]$,

$$f\left(\lambda x + (1-\lambda)y\right) \leq \underbrace{\lambda}_{\geq 0} f(x) + \underbrace{(1-\lambda)}_{\geq 0} f(y) \leq \lambda z_1 + (1-\lambda)z_2.$$

Therefore, $(\lambda x + (1-\lambda)y, \lambda z_1 + (1-\lambda)z_2) = \lambda(x, z_1) + (1-\lambda)(y, z_2) \in \text{epi}(f)$, i.e., $\text{epi}(f) \in \mathbb{R}^{n+1}$ is a convex set.

$\Leftarrow$: Let $\text{epi}(f) \in \mathbb{R}^{n+1}$ be a convex set. For any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, we have $(x, f(x)) \in \text{epi}(f)$ and $(y, f(y)) \in \text{epi}(f)$. Since $\text{epi}(f) \in \mathbb{R}^{n+1}$ is a convex set, for any $\lambda \in [0, 1]$, we have $\lambda(x, f(x)) + (1-\lambda)(y, f(y)) = (\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda)f(y)) \in \text{epi}(f)$. Therefore,

$$f\left(\lambda x + (1-\lambda)y\right) \leq \lambda f(x) + (1-\lambda)f(y),$$

i.e., $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. $\square$

3.6 Concave Functions

Definition 3.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We say that f is a* concave function *if $\forall x \in \mathbb{R}^n$, $\forall y \in \mathbb{R}^n$, $\forall \lambda \in [0, 1]$, we have*

$$f\left(\lambda x + (1-\lambda)y\right) \geq \lambda f(x) + (1-\lambda)f(y).$$

As opposed to convexity, note that concavity is only defined for functions. There is no analogous definition for sets. A set that is not convex is called a *nonconvex set.*

Remark 3.1. *$f : \mathbb{R}^n \to \mathbb{R}$ is a concave function if and only if $-f : \mathbb{R}^n \to \mathbb{R}$ is a convex function.*

Exercises

Question 3.1. *For each of the following subsets of* $\mathbb{R}^2$*, determine whether it is a convex set or a nonconvex set.*

- $\mathcal{C} = \{x \in \mathbb{R}^2 : x_2 \leq |x_1|\}$
- $\mathcal{C} = \{x \in \mathbb{R}^2 : x_2 \leq |x_1|, \quad x_1 \geq 0\}$

Question 3.2. *For each of the following functions* $f : \mathbb{R}^2 \to \mathbb{R}$*, determine whether it is a convex function, concave function, both, or neither.*

- $f(x) = 3x_1 - 2x_2$
- $f(x) = 3|x_1| + 2x_2$
- $f(x) = -3|x_1| + 2|x_2|$

MATH11111: Fundamentals of Optimization	**Fall 2022**
Lecture 4	Level Sets, Sublevel and Superlevel Sets
Lecturer: E. Alper Yıldırım	*Week: 1*

4.1 Outline

- Level Sets
- Hyperplanes
- Sublevel and Superlevel Sets
- Halfspaces
- Review Problems

4.2 Level Sets

Definition 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. For each $\alpha \in \mathbb{R}$, the level set of f is given by*

$$\mathcal{L}_\alpha(f) = \{x \in \mathbb{R}^n : f(x) = \alpha\}$$

Remarks

1. For every function $f : \mathbb{R}^n \to \mathbb{R}$ and every $\alpha \in \mathbb{R}$, the level set of f given by $\mathcal{L}_\alpha(f)$ is a subset of the domain of the function, i.e., $\mathcal{L}_\alpha(f) \subseteq \mathbb{R}^n$. It is, in fact, the inverse image of α. The union of the sets $\mathcal{L}_\alpha(f)$ over all values of $\alpha \in \mathbb{R}$ is equal to $\mathbb{R}^n$, i.e.,
$$\bigcup_{\alpha \in \mathbb{R}} \mathcal{L}_\alpha(f) = \mathbb{R}^n.$$
2. For a given function $f : \mathbb{R}^n \to \mathbb{R}$, the *range* of f is given by the set of possible output values, i.e., the range of f is given by
$$\bigcup_{x \in \mathbb{R}^n} \{f(x)\} \subseteq \mathbb{R}.$$
3. For a given function $f : \mathbb{R}^n \to \mathbb{R}$ and a given $\alpha \in \mathbb{R}$, the level set $\mathcal{L}_\alpha(f)$ can be equal to the empty set. This is true if and only if α is not in the range of the function f.

4.2.1 Level Sets of Linear Functions

Definition 4.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function given by $f(x) = a^T x$, where $a \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. For each $\alpha \in \mathbb{R}$, the level set of f given by*

$$\mathcal{L}_\alpha(f) = \left\{x \in \mathbb{R}^n : a^T x = \alpha\right\}$$

is called a hyperplane.

Remarks

1. If $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = ax$, where $a \in \mathbb{R}\backslash\{0\}$, then $\mathcal{L}_\alpha(f)$ is a point on the real line for any $\alpha \in \mathbb{R}$ (zero degrees of freedom).
2. If $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x) = a^T x$, where $a \in \mathbb{R}^2\backslash\{\mathbf{0}\}$, then $\mathcal{L}_\alpha(f)$ is a line in $\mathbb{R}^2$ for any $\alpha \in \mathbb{R}$ (one degree of freedom).
3. If $f : \mathbb{R}^3 \to \mathbb{R}$ is given by $f(x) = a^T x$, where $a \in \mathbb{R}^3\backslash\{\mathbf{0}\}$, then $\mathcal{L}_\alpha(f)$ is a plane in $\mathbb{R}^3$ for any $\alpha \in \mathbb{R}$ (two degrees of freedom).
4. More generally, for $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = a^T x$, where $a \in \mathbb{R}^n\backslash\{\mathbf{0}\}$, then $\mathcal{L}_\alpha(f)$ is a hyperplane with $n-1$ degrees of freedom.

4.2.2 Hyperplanes and Convexity

Proposition 4.1. *Every level set of a linear function is a convex set. Therefore, every hyperplane is a convex set.*

Proof. Let $\mathcal{H} \subset \mathbb{R}^n$ be a hyperplane. Then, there exists a linear function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = a^T x$, where $a \in \mathbb{R}^n\backslash\{\mathbf{0}\}$, and an $\alpha \in \mathbb{R}$ such that $\mathcal{H}$ is given by the level set of f:

$$\mathcal{H} = L_\alpha(f) = \left\{x \in \mathbb{R}^n : a^T x = \alpha\right\}.$$

For any $x^1 \in \mathcal{H}$, $x^2 \in \mathcal{H}$, and any $\lambda \in [0, 1]$, $a^T(\lambda x^1 + (1-\lambda)x^2) = \lambda a^T x^1 + (1-\lambda)a^T x^2 = \lambda\alpha + (1-\lambda)\alpha = \alpha$. Therefore, $\lambda x^1 + (1-\lambda)x^2 \in \mathcal{H}$, which implies that $\mathcal{H}$ is a convex set. □

4.3 Sublevel Sets and Superlevel Sets

Definition 4.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. For each $\alpha \in \mathbb{R}$, the sublevel set of f is given by*

$$\mathcal{L}_\alpha^-(f) = \{x \in \mathbb{R}^n : f(x) \le \alpha\}.$$

Definition 4.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. For each $\alpha \in \mathbb{R}$, the superlevel set of f is given by*

$$\mathcal{L}_\alpha^+(f) = \{x \in \mathbb{R}^n : f(x) \ge \alpha\}.$$

Remarks

1. For a given $\alpha \in \mathbb{R}$, similarly to the level set of f given by $\mathcal{L}_\alpha(f)$, each of the sublevel set $\mathcal{L}_\alpha^-(f)$ and the superlevel set $\mathcal{L}_\alpha^+(f)$ is a subset of $\mathbb{R}^n$. For a given $\alpha \in \mathbb{R}$, the sublevel set $\mathcal{L}_\alpha^-(f)$ or the superlevel set $\mathcal{L}_\alpha^+(f)$ can be the empty set, but not both of them can be empty at the same time for the same value of $\alpha \in \mathbb{R}$.

2. For every function $f : \mathbb{R}^n \to \mathbb{R}$ and every $\alpha \in \mathbb{R}$,

$$\mathcal{L}_\alpha^-(f) \cup \mathcal{L}_\alpha^+(f) = \mathbb{R}^n, \qquad \mathcal{L}_\alpha^-(f) \cap \mathcal{L}_\alpha^+(f) = \mathcal{L}_\alpha(f).$$

4.3.1 Sublevel Sets of Convex Functions

Proposition 4.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. For each $\alpha \in \mathbb{R}$, the sublevel set of f given by*

$$\mathcal{L}_\alpha^-(f) = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$$

is a convex set.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $\alpha \in \mathbb{R}$. Let $x^1 \in \mathcal{L}_\alpha^-(f)$ and $x^2 \in \mathcal{L}_\alpha^-(f)$. Then, for any $\lambda \in [0,1]$,

$$f\left(\lambda x^1 + (1-\lambda)x^2\right) \leq \underbrace{\lambda}_{\geq 0} \underbrace{f(x^1)}_{\leq \alpha} + \underbrace{(1-\lambda)}_{\geq 0} \underbrace{f(x^2)}_{\leq \alpha} \leq \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

Therefore, $\lambda x^1 + (1-\lambda)x^2 \in \mathcal{L}_\alpha^-(f)$, which implies that $\mathcal{L}_\alpha^-(f)$ is a convex set. □

4.3.2 Superlevel Sets of Concave Functions

Corollary 4.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a concave function. For each $\alpha \in \mathbb{R}$, the superlevel set of f given by*

$$\mathcal{L}_\alpha^+(f) = \{x \in \mathbb{R}^n : f(x) \geq \alpha\}$$

is a convex set.

Proof. Since f is a concave function, $-f$ is a convex function and $\mathcal{L}_\alpha^+(f) = \mathcal{L}_{-\alpha}^-(-f)$. The result follows from Proposition 4.2. □

4.4 Convexity and Concavity of Linear Functions

Proposition 4.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function given by $f(x) = a^T x$, where $a \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. Then, f is both a convex and a concave function.*

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function given by $f(x) = a^T x$, where $a \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. Then, $x^1 \in \mathbb{R}^n$, $x^2 \in \mathbb{R}^n$, and any $\lambda \in [0,1]$, $f(\lambda x^1 + (1-\lambda)x^2) = a^T(\lambda x^1 + (1-\lambda)x^2) = \lambda a^T x^1 + (1-\lambda)a^T x^2 = \lambda f(x^1) + (1-\lambda)f(x^2)$. Therefore, f is both convex and concave. □

4.4.1 Halfspaces

Definition 4.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function given by $f(x) = a^T x$, where $a \in \mathbb{R}^n \backslash \{\mathbf{0}\}$. For each $\alpha \in \mathbb{R}$, each of the sublevel set of f given by*

$$\mathcal{L}_\alpha^-(f) = \{x \in \mathbb{R}^n : a^T x \leq \alpha\}$$

and the superlevel set of f given by

$$\mathcal{L}_\alpha^+(f) = \{x \in \mathbb{R}^n : a^T x \geq \alpha\}$$

is called a halfspace.

4.4.2 Convexity of Halfspaces

Corollary 4.7. *Every halfspace is a convex set.*

Proof. Since a halfspace is given by the sublevel set of superlevel set of a linear function and since every linear function is convex and concave by Proposition 4.3, the result follows from Proposition 4.2 and Corollary 4.5. □

4.4.3 Properties of Halfspaces

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function given by $f(x) = a^T x$, where $a \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and let $\mathcal{H} = \mathcal{L}_\alpha(f) = \{x \in \mathbb{R}^n : a^T x = \alpha\}$ be a hyperplane, where $\alpha \in \mathbb{R}$. Let

$$\begin{aligned} \mathcal{H}^- &= \mathcal{L}_\alpha^-(f) = \{x \in \mathbb{R}^n : a^T x \leq \alpha\} \\ \mathcal{H}^+ &= \mathcal{L}_\alpha^+(f) = \{x \in \mathbb{R}^n : a^T x \geq \alpha\} \end{aligned}$$

be the corresponding halfspaces.

1. $\mathcal{H}$ is the boundary of each of $\mathcal{H}^-$ and $\mathcal{H}^+$. Furthermore,

$$\mathcal{H} = \mathcal{H}^- \cap \mathcal{H}^+.$$

2. The vector $a \in \mathbb{R}^n$ is perpendicular to $\mathcal{H}$, i.e., for any $x^1 \in \mathcal{H}$ and $x^2 \in \mathcal{H}$, $a^T(x^1 - x^2) = \alpha - \alpha = 0$.

Exercises

Question 4.1. *If $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x) = x_1^2 + x_2^2$ and $\alpha = 1$, then give a description of $\mathcal{L}_\alpha(f)$.*

Question 4.2. *If $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = |x|$ and $\alpha = 1$, then give a description of $\mathcal{L}_\alpha^+(f)$.*

Question 4.3. *If $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x) = -x_1 + x_2$ and $\alpha = 1$, then give a description of $\mathcal{L}_\alpha^-(f)$.*

Question 4.4. *Show that the intersection of any number of halfspaces and hyperplanes in $\mathbb{R}^n$ is a convex set.*