1 Linear programming: history and overview

History

Linear programming problems (LPPs) are a specific class of mathematical problems, in which a linear function is maximised (for example, maximal profit) or minimised (minimal cost) subject to a given set of linear constraints (resource restriction or contract requirements). This problem class is rich enough to cover many interesting and important applications, yet most LPPs can be solved efficiently even if the number of variables and constraints is large.

LPPs were first seriously studied in the late 1930s by the Soviet mathematician Leonid Kantorovich (1912-1986) and by the Russian-born American economist Wassily Leontief (1906-1999) in the areas of manufacturing schedules and of economics, respectively. Kantorovich authored several books including "The Mathematical Method of Production Planning and Organization" and "The Best Uses of Economic Resources". For his work (published in 1939, when he was just 27 years old), Kantorovich was awarded The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel (Nobel Prize), which he shared with another mathematician Tjalling Koopmans. This prize, was given "for their contributions to the theory of optimum allocation of resources".

During World War II, linear programming was used extensively in wartime operations: modelling efficient scheduling and resources allocation subject to certain natural restrictions (for example, costs and availability). In 1947 the American mathematician George Dantzig (1914-2005) introduced the famous *simplex method, which greatly simplified the solution of linear programming problems. In the same year, another famous mathematician John von Neumann (1903-1957) established the theory of duality.*

There are a number of algorithms for linear programming (other than the simplex method). They include Khachiyan's ellipsoidal algorithm, Karmarkar's projective algorithm, and path-following algorithms.

Modern application

Many modern industries use linear programming as a standard rigorous tool to allocate their available resources in a best possible (optimal) way. Examples of important application areas include staff scheduling (including airline crew shift allocation), shipping, telecommunication, oil refining and blending, stock and bond portfolio selection, data fitting, signal processing and many others.

Linear programming is part of a wider area of modern mathematics, called mathematical programming (also known as "mathematical optimisation" or just "optimisation"). In this unit, you will learn about linear programming and a bit beyond it (integer programming and convex optimisa-

tion). You will learn how to recognise such kind of problems, formulate them mathematically and solve using modern optimisation techniques.

2 Linear functions

In this section we provide a brief review of results from linear algebra that are used in this unit.

Linear spaces, vectors and matrices

We start by introducing a number of definitions.

Definition 2.1 *A* matrix of dimensions $m \times n$ is an array of real numbers a_{ij}, such that

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

Matrices are always denoted by upper-case boldface characters. To refer to an $(i, j)-th$ entry of a matrix $\mathbf{A}$ we use the notation a_{ij}.

Definition 2.2 *An $m \times n$-dimensional matrix is a* square matrix, if $m = n$. In this case, this matrix is also called an *$n-$dimensional square matrix.*

Definition 2.3 *An n-dimensional square matrix is* symmetric matrix, if $a_{ij} = a_{ji},\ i, j = 1, \dots, n$.

Definition 2.4 *A* row vector is a matrix with $m = 1$ and a *column vector is a matrix with $n = 1$.*

The word vector always means *column vector.* Vectors are usually denoted by low-case boldface character. The notation $\mathbb{R}^n$ is used to indicate the space of all $n-$dimensional vectors. For any vector $\mathbf{x} \in \mathbb{R}^n$ we use $x_1, x_2, \dots, x_n$ to indicate its components:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

We use $\mathbf{O}$ to denote vectors and matrices with all components equal to zero.

Definition 2.5 *The* transpose $\mathbf{A}^T$ of an $m \times n$ matrix $\mathbf{A}$ is the $n \times n$ matrix

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix}.$$

Therefore, the $(i, j)-$th entry of $\mathbf{A}$ is also the $(j, i)-$th entry of $\mathbf{A}^T$. If $\mathbf{x}$ is a column vector then $\mathbf{x}^T$ is a row vector and vice versa.

Definition 2.6 *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are two vectors in* $\mathbb{R}^n$ *then the quantity*

$$\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = \sum_{i=1}^{n} x_i y_i$$

is called inner product of $\mathbf{x}$ and $\mathbf{y}$.

Definition 2.7 *Two vectors are called* orthogonal is their inner product is zero.

Definition 2.8 *The expression* $\sqrt{\mathbf{x}^T\mathbf{x}}$ *is called the* Euclidean norm and denoted by $\|\mathbf{x}\|$.

Note the following.

1. $\|\mathbf{x}\| \geq 0$ and the equality holds if and only if $\mathbf{x} = \mathbf{O}$.
2. *Schwartz inequality:* $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$. *The equality holds if and only if one of the vectors is a scalar multiple of another one (that is,* $\mathbf{x} = a\mathbf{y}$*).*

If $\mathbf{A}$ is an $m \times n$ matrix, we use $\mathbf{A}_j$ to indicate its $j-$th column and $\mathbf{a}_i$ to indicate the entries of its $i-$th row.

Definition 2.9 *Given two matrices* $\mathbf{A}$ *of dimension* $m \times n$ *and* $\mathbf{B}$ *of dimension* $n \times k$ *the* matrix product $\mathbf{AB}$ is an $m \times k$ dimensional matrix whose (ij) entry is

$$\sum_{l=1}^{n} a_{il} b_{lj} = \mathbf{a}_i^T \mathbf{B}_j,$$

where $\mathbf{a}_i^T$ is the $i-$th row of $\mathbf{A}$ and $\mathbf{B}_j$ is the $j-$th column of $\mathbf{B}$.

Recall that a matrix is called *square matrix if the number of rows is the same as the number of columns.*

Definition 2.10 *A square matrix with all the diagonal elements equal to one and all the off-diagonal elements equal to zero is called the* identity matrix, denoted by $\mathbf{I}$.

Consider the following matrix multiplication properties.

1. Matrix multiplication is associative: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.
2. Matrix multiplication, in general, is **not** commutative: $\mathbf{AB} = \mathbf{BA}$ is **not** always true.
3. Matrix product transposition: $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

4. Identity matrix properties: $\mathbf{IA} = \mathbf{A}$ and $\mathbf{BI} = \mathbf{B}$ for any matrix $\mathbf{A}$ and $\mathbf{B}$ of dimensions compatible with those of $\mathbf{I}$.

Definition 2.11 *If $\mathbf{x}$ is a vector ($\mathbf{A}$ is a matrix), the notation $\mathbf{x} \geq \mathbf{O}$ or $\mathbf{x} > \mathbf{O}$ ($\mathbf{A} \geq \mathbf{O}$ or $\mathbf{A} > \mathbf{O}$) means that all the components of $\mathbf{x}$ ($\mathbf{A}$) are non-negative or positive respectively.*

Definition 2.12 *A square matrix $\mathbf{A}$ is* invertible or *nonsingular if there exists a matrix $\mathbf{B}$ (same dimension as $\mathbf{A}$), such that*

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

Such a matrix $\mathbf{B}$ is unique and called the inverse of $\mathbf{A}$ and denoted $\mathbf{A}^{-1}$.

Note the following important properties of invertible matrices.

1. If $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices of the same dimension, then the product $\mathbf{AB}$ is also invertible and

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

2. For an invertible matrix $\mathbf{A}$ the following holds:

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T.$$

Definition 2.13 *Consider a finite collection of vectors $\mathbf{x}^1, \ldots, \mathbf{x}^l \in \mathbb{R}^n$. We say that these vectors are* linearly dependent if there exist real numbers $a_1, \ldots, a_l$, such that not all of them are zero and

$$\sum_{i=1}^{l} a_i \mathbf{x}^i = \mathbf{O}. \qquad (1)$$

Otherwise, they are called *linearly independent.*

Definition 2.14 *A* linear combination of a finite collection of vectors

$$\mathbf{x}^1, \ldots, \mathbf{x}^l \in \mathbb{R}^n$$

is any expression of the form

$$\mathbf{x} = \sum_{i=1}^{l} a_i \mathbf{x}^i,$$

where $a_1, \ldots, a_l$ are any real numbers (*scalars).*

Theorem 2.1 *Vectors $\mathbf{x}^1, \ldots, \mathbf{x}^l \in \mathbb{R}^n$ are linearly dependent if and only if at least one of these vectors is a linear combination of the remaining vectors.*

Proof: We have to proof both directions:

1. if vectors $\mathbf{x}^1, \ldots, \mathbf{x}^l \in \mathbb{R}^n$ are linearly dependent then at least one of them is a linear combination of the remaining vectors;

2. if at least one of the vectors $\mathbf{x}^1, \ldots, \mathbf{x}^l$ is a linear combination of the remaining vectors, then these vectors are linearly dependent.

If vectors $\mathbf{x}^1, \ldots, \mathbf{x}^l \in \mathbb{R}^n$ are linearly dependent, then their linear combination $\sum_{i=1}^{l} a_i \mathbf{x}^i$ is zero while at least one of the numbers $a_1, \ldots, a_l$ is not zero. Assume that $a_k \neq 0,\ 1 \leq k \leq l$, then, since $a_k \neq 0$,

$$\mathbf{x}^k = \sum_{i=1, i \neq k}^{l} \frac{a_i}{a_k} \mathbf{x}^i.$$

Therefore, vector $\mathbf{x}^k$ is a linear combination of the remaining vectors.

Now assume that vector $\mathbf{x}^k$ is a linear combination of the remaining vectors. Then

$$\sum_{i=1}^{l} a_i \mathbf{x}^i = \mathbf{O},$$

where $a_k = -1 \neq 0$. Therefore, these vectors are linearly dependent.

□

It can be shown that

1. a square matrix $\mathbf{A}$ is invertible if and only if its columns (rows) are linearly independent;

2. any $n+1$ or more vectors in $\mathbb{R}^n$ are linearly dependent.

Linear functions and systems of linear equations

Definition 2.15 *An* affine function represents a function of the form

$$f(x_1, \ldots, x_n) = a_1 x_1 + \cdots + a_n x_n + b = \mathbf{a}\mathbf{x} + b,$$

where $\mathbf{x} \in \mathbb{R}^n$ represents the variables, $\mathbf{a} = (a_1, \ldots, a_n)$ is a fixed row vector and b is a fixed number (constant term).

A constant function $f(x_1, \ldots, x_n) = b$ is also considered affine in this context. Its graph is a horizontal line (if there is only one independent variable).

Definition 2.16 *A* linear function represents a function of the form

$$f(x_1, \ldots, x_n) = a_1 x_1 + \cdots + a_n x_n = \mathbf{a}\mathbf{x}, \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^n$ represents the variables, $\mathbf{a} = (a_1, \ldots, a_n)$ is a fixed row vector (constant term $b = 0$).

It can be shown that:

1. the sum of a finite number of linear (affine) functions is linear (affine);
2. if $f(\mathbf{x})$ is linear (affine), then $kf(\mathbf{x})$ is also linear (affine), where k is a real number.

Definition 2.17 *A* linear equation is an algebraic equation where the left-hand side is represented by a linear function and the right-hand side is a constant.

It is clear that an equation in which the left-hand side is represented by an affine function and the right-hand side is a constant is also linear. In general, linear equations have the following form:

$$a_1x_1 + \cdots + a_nx_n = \mathbf{ax} = b, \tag{3}$$

where $\mathbf{x} \in \mathbb{R}^n$ represents the variables, $\mathbf{a} = (a_1, \ldots, a_n)$ is a fixed row vector, b is a constant.

Definition 2.18 *A* system of linear equations (or *linear system) is a set of linear equations with the same variables.*

Definition 2.19 *A solution to a linear system is a set of numbers (assigned to the variables) such that all the equations are simultaneously satisfied.*

Definition 2.20 *The set of all possible solutions is called the* solution set.

There are three different possibilities.

1. The solution set contains infinitely many elements, means that the system has infinitely many solutions.
2. The solution set contains only one element, means that the system has a single (unique) solution.
3. The solution set is empty, means that the system has no solution.

A system of linear equations can be written as

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1; \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{4}$$

The corresponding matrix form is

$$\mathbf{Ax} = \mathbf{b}, \tag{5}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ (that is $m \times n$ matrix), $\mathbf{x} \in \mathbb{R}^n$ is a column vector (variables) and $\mathbf{b} \in \mathbb{R}^m$ is a fixed column vector (right-hand side).

Definition 2.21 *If in a linear system all the equality signs are substituted by the same inequality sign, then this system is said to be a* system of linear inequalities.

Similarly, a *system of linear inequalities can be written as*

$$\mathbf{Ax} < (or \ \ \geq, \ \ >, \ \ \leq) \ \mathbf{b}, \tag{6}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$ *is a column vector (variables) and* $\mathbf{b} \in \mathbb{R}^m$ *is a fixed column vector (right-hand side).*

Hyperplanes, halfspaces, polyhedra and polytopes

We start with several definitions.

Definition 2.22 *A set* $S \in \mathbb{R}^n$ *is said to be bounded if there exists a constant* M*, such that the absolute value of every component of every element of this set does not exceed* M*. Otherwise,* S *is said to be unbounded.*

Definition 2.23 *A* convex polyhedron (plural polyhedra) is a set that can be expressed in the form

$$\{\mathbf{x} \in \mathbb{R}^n | \ \mathbf{Ax} \geq \mathbf{b}\},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

Therefore, any convex polyhedron is formed by all the possible solution to the corresponding system of linear inequalities. If such a system contains only one inequality and thus $\mathbf{b}$ *is a scalar, then there are two possibilities. In this unit, we will be working with convex polyhedra, therefore, in some cases, we will refer to them as just* polyhedra for simplicity.

1. The set $\{x \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b\}$ is called a *hyperplane.*
2. *The set* $\{x \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} \leq b\}$ *(*$\{x \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} \geq b\}$*) is called a* halfspace.

Each hyperplane divides the original space into two halfspaces (positive and negative, see Figure 1).

Note that

- hyperplanes and halfspaces are unbounded;
- a hyperplane is the boundary of a corresponding halfspace;
- a convex polyhedron is an intersection of a finite number of halspaces;
- bounded polyhedra are also called polytopes.

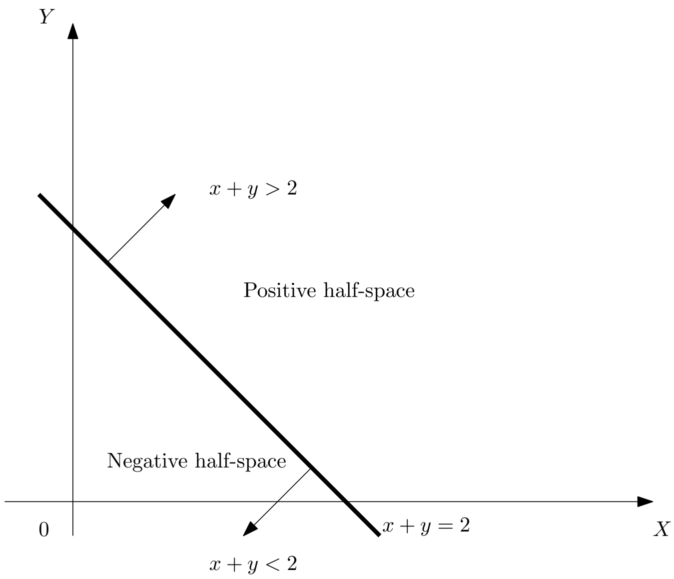

Figure 1: Positive and negative halfspaces.

3 Convex functions and convex sets

Convex functions

In section 2 we were working with linear and affine function. Now we are ready to make one more step forward and introduce *convex functions.*

Definition 3.1 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called* convex if for every pair of vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and every $\lambda \in [0, 1]$, we have

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

Definition 3.2 *A function f is called* concave if the function $-f$ is convex.

It is possible to prove that linear and affine functions are convex and concave simultaneously. Moreover, they are the only functions, that are both convex and concave.

Theorem 3.1 *Let $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ are convex function, then the function*

$$f(\mathbf{x}) = \max_{i=1,\ldots,m} f_i(\mathbf{x})$$

is also convex (maximum of convex functions is convex).

Proof:

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \max_{i=1,\ldots,m} f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &\leq \max_{i=1,\ldots,m} (\lambda f_i(\mathbf{x}) + (1-\lambda) f_i(\mathbf{y})) \\ &\leq \max_{i=1,\ldots,m} \lambda f_i(\mathbf{x}) + (1-\lambda) \max_{i=1,\ldots,m} f_i(\mathbf{y}) \\ &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}). \end{aligned}$$

□

Example 3.1 *Show that the function $f(x) = |x|$ is convex.*

First note that $f(x)$ can be expressed as

$$f(x) = \max\{x, -x\},$$

that is, the maximum of two linear (and therefore convex) functions $f_1(x) = x$ and $f_2(x) = -x$. Applying Theorem 3.1, obtain that $f(x) = |x|$ is convex.

Theorem 3.2 *Multiplication by a nonnegative constant and summation preserve convexity.*

1. *If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and α is a nonnegative number, then $g(\mathbf{x}) = \alpha f(\mathbf{x})$ is also convex.*

2. *If* $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ *are convex functions, then* $g(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ *is also convex.*

Proof: *Consider each case separately.*

1. *Multiplication by a nonnegative constant.*

$$\begin{aligned} g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \alpha f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &\leq \alpha(\lambda f(\mathbf{x})) + \alpha(1-\lambda)f(\mathbf{y}) \\ &= \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}). \end{aligned}$$

2. *Sum of convex functions.*

$$\begin{aligned} g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= f_1(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + f_2(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &\leq \lambda f_1(\mathbf{x}) + (1-\lambda)f_1(\mathbf{y}) + \lambda f_2(\mathbf{x}) + (1-\lambda)f_2(\mathbf{y}) \\ &= \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}). \end{aligned}$$

□ *Based on Theorems 3.1 and 3.2, one can conclude that if*

$$f_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, \ldots, m$$

are convex and the corresponding scalars $\alpha_i,\ i = 1, \ldots, m$ *are nonnegative, then*

$$f = \sum_{i=1}^{m} \alpha_i f_i$$

is convex.

Definition 3.3 *The* domain of a function $f(\mathbf{x})$ (notation: $\mathrm{dom}(f)$) is the complete set of possible values of the independent variable $\mathbf{x}$.

From Calculus:

- *Assume that* $f(x)$ *is* **twice continuously differentiable** *and the domain is the real line, then* $f(x)$ *convex if and only if* $f''(x) \geq 0$ *for all* x.
- *A twice differentiable function* $f(x)$ *of one variable is convex on an interval* $[a, b]$, $a \neq b$ *if and only if its second derivative is non-negative there:* $f''(x) \geq 0$ *for all* $x \in [a, b]$.

Example 3.2 *Consider the following functions and their second derivatives:*

- *$f_1(x) = x^3$, then f_1'' changes sign at $x = 0$ and therefore $f_1(x)$ is neither convex nor concave;*
- *$f_2 = x^4$, then $f_2'' \geq 0$ and therefore $f_2(x)$ is neither convex;*
- *$f_3 = -x^4$, then $f_3'' \leq 0$ and therefore $f_2(x)$ is neither concave (or just simply because $f_2(x) = -f_3(x)$ is convex).*

Theorem 3.3 *(Composition with an affine mapping) Let $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{\mathbf{n\times m}}$ and $\mathbf{b} \in \mathbb{R}^n$. Define $g : \mathbb{R}^m \to \mathbb{R}$ by*

$$g(\mathbf{x}) = f(\mathbf{Ax} + \mathbf{b})$$

with

$$\mathrm{dom}(g) = \{\mathbf{x} : \mathbf{Ax} + \mathbf{b} \in \mathrm{dom}(f)\}.$$

Then if f is convex, so is g.

Proof: *Let $\mathbf{x},\ \mathbf{y} \in \mathbb{R}$ and $\lambda \in [0, 1]$, then:*

$$\begin{aligned} g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= f(\mathbf{A}(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + \mathbf{b}) \\ &= f(\lambda(\mathbf{Ax} + \mathbf{b}) + (1-\lambda)\mathbf{Ay} + \mathbf{b}) \\ &\leq \lambda f(\mathbf{Ax} + \mathbf{b}) + (1-\lambda)f(\mathbf{Ay} + \mathbf{b}) \\ &= \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}). \end{aligned}$$

Therefore, g is convex.

□

Definition 3.4 *A vector* $\mathbf{x}$ *is a* local minimum of f if $f(\mathbf{x}) \leq f(\mathbf{y})$ for all $\mathbf{y}$ in a neighbourhood of $\mathbf{x}$.

Definition 3.5 *A vector* $\mathbf{x}$ *is a* global minimum of f if $f(\mathbf{x}) \leq f(\mathbf{y})$ for all $\mathbf{y}$.

Figure 2 illustrates the difference between a local minimum ($x = 20$) and a global one ($x = 80$).

If f is convex, any local minimum is also a global one. This special property of convex functions helps to design powerful optimisation algorithms. Note that convex functions may have more than one global minimum (see Figure 3, where any $x \in [30, 70]$ is a global minimum).

Example 3.3 *Show that function*

$$F(t) = \min\{g_1(t), \ldots, g_n(t)\},$$

where $g_i(t),\ i = 1, \ldots, n$ are concave functions on $[a, b]$ is concave on $[a, b]$.

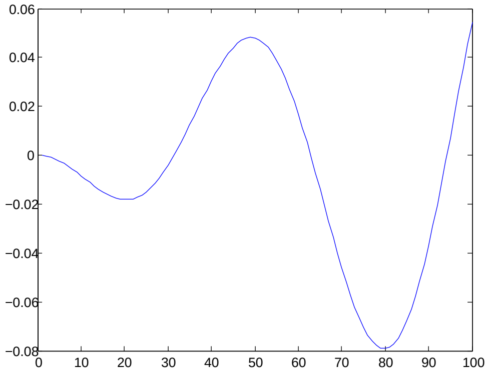

Figure 2: Local minimum $x = 20$ vs global $x = 80$.

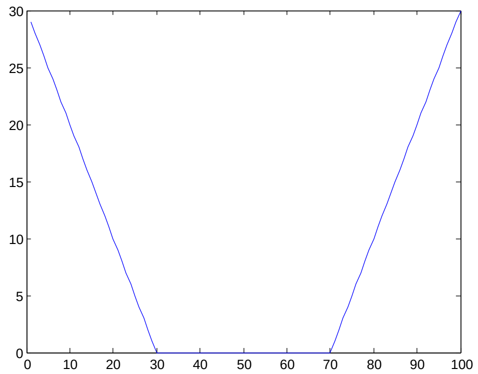

Figure 3: Any $x \in [30, 70]$ is a global minimum.

Note, first of all, that

$$F(t) = \min\{g_1, \dots, g_n\} = -\max\{-g_1, \dots, -g_n\},$$

where the functions $h_i(t) = -g_i(t),\ i = 1, \dots, n$ *are convex. Since maximum of convex functions is convex, function*

$$-F(t) = \max\{h_1(t), \dots, h_n(t)\}$$

is convex and therefore function

$$F(t) = \min\{g_1(t), \dots, g_n(t)\}$$

is concave.

Convex sets

Convex sets play a very important role in optimisation.

Definition 3.6 *A set* $S \subset \mathbb{R}^n$ *is* convex if and only if for any pair $\mathbf{x},\ \mathbf{y} \in S$ and any $\lambda \in [0,1]$ (that is, $0 \leq \lambda \leq 1$), we have

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S.$$

In other words, a set S *is convex if and only if for any pair* $\mathbf{x}, \mathbf{y} \in S$*, the whole line segment joining these two vectors belongs to* S*.*

Example 3.4 *Figure 4 depicts two sets. The first set (left), a polygon is convex, since for any pair of point from this set, the whole segment, joining these points together, also belongs to this set. The second set (right) is not convex, since points* X *and* Y *belong to this set, while point* Z*, lying on the segment* XY *does not belong to this set.*

Definition 3.7 *If* $\lambda \in [0,1]$*, then* $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ *is a* weighted average of the vectors $\mathbf{x}$ and $\mathbf{y}$.

Similarly, one can define a weighted average for more than two vectors (finite number).

Definition 3.8 *Let* $\mathbf{x}^1, \dots, \mathbf{x}^k \in \mathbb{R}^n$ *and* $\lambda_1, \dots, \lambda_k$ *be nonnegative scalars whose sum is equal to one (that is,* $\sum_{i=1}^k \lambda_i = 1$*).*

- *The vector* $\sum_{i=1}^k \lambda_i \mathbf{x}^i$ *is called a* convex combination (weighted average) of the vectors $\mathbf{x}^1, \dots, \mathbf{x}^k$.
- The set of all convex combinations of the vectors $\mathbf{x}^1, \dots, \mathbf{x}^k$ is called the *convex hull of these vectors (see Figure 5).*

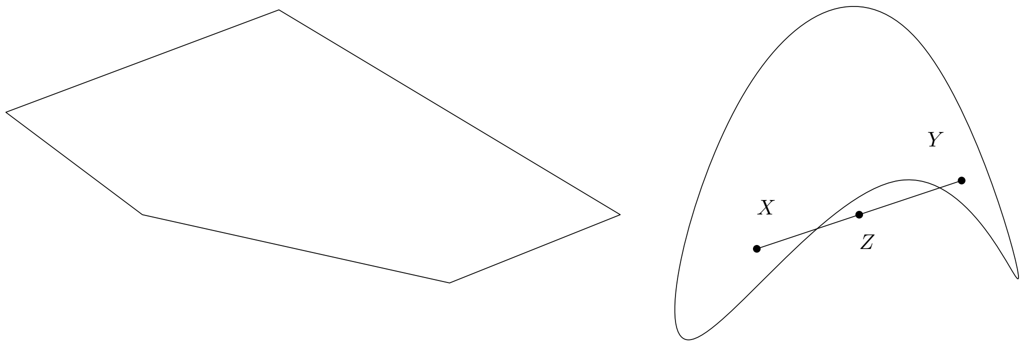

Figure 4: Convex and nonconvex sets.

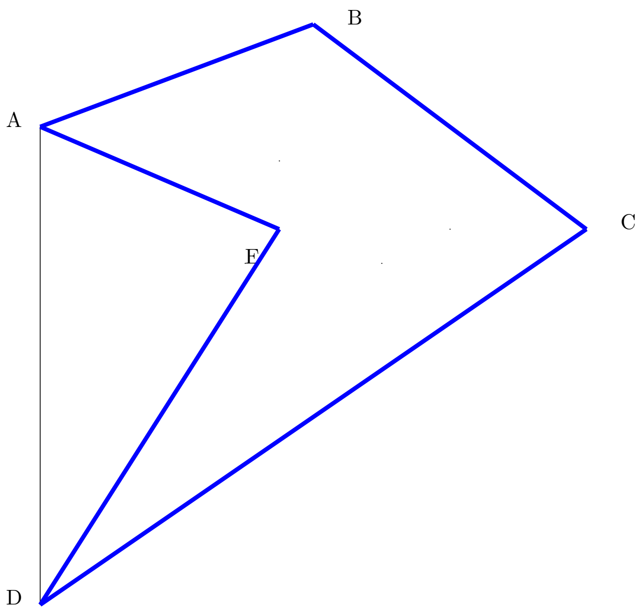

Figure 5: Convex hull: ABCD is the convex hull of A, B, C, D, F.

It can be demonstrated that the convex hull of a finite number of vectors is a convex set. Also, a convex combination of a finite number of elements of a convex set belongs to that set as well.

Theorem 3.4 *Hyperplanes and halfspaces are convex sets.*

Proof:
Hyperplanes: *Consider two vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *from* $\mathbb{R}^n$*, such that*

$$\mathbf{a}^T\mathbf{x} = b \text{ and } \mathbf{a}^T\mathbf{y} = b.$$

Consider now $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$*, where* $\lambda \in [0,1]$.

$$\mathbf{a}^T\mathbf{z} = \mathbf{a}^T\lambda\mathbf{x} + \mathbf{a}^T(1-\lambda)\mathbf{y} = \lambda b + (1-\lambda)b = b.$$

Halfspaces: *Consider two vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *from* $\mathbb{R}^n$*, such that*

$$\mathbf{a}^T\mathbf{x} \geq b \text{ and } \mathbf{a}^T\mathbf{y} \geq b.$$

Consider now $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$*, where* $\lambda \in [0,1]$.

$$\mathbf{a}^T\mathbf{z} = \mathbf{a}^T\lambda\mathbf{x} + \mathbf{a}^T(1-\lambda)\mathbf{y} \geq \lambda b + (1-\lambda)b = b.$$

□

Theorem 3.5 *The intersection of convex sets is convex.*

Proof: *Let* $S_i,\ i \in I$ *are convex sets, where* I *is an index set (set of indices). Assume that* $\mathbf{x}$ *and* $\mathbf{y}$ *belong to the intersection* $\cap_{i\in I}S_i$ *and* $\lambda \in [0,1]$*. Since* S_i *is convex, vector*

$$\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S_i.$$

Therefore, $\mathbf{z}$ *belongs to every* $S_i,\ i \in I$ *and therefore* $\mathbf{z}$ *belongs to the intersection of the sets* $S_i,\ i \in I$*. Therefore, the intersection of convex sets is convex.*

□

Combining Theorems 3.4 and 3.5, obtain that every convex polyhedron is a convex set.

Definition 3.9 *Let* P *be a convex polyhedron. A vector* $\mathbf{x} \in P$ *is an* extreme point of P if we can not find two vectors $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{z} \neq \mathbf{x}$ that belong to this polyhedron and $\mathbf{x} = \lambda\mathbf{y} + (1-\lambda)\mathbf{z}$, where $\lambda \in [0,1]$.

Essentially, this means that an extreme point can not be presented as a convex combination of two distinct points from the same polyhedron (that is, can not lie on a segment line, connecting two points from this polyhedron).

Definition 3.10 *Let P be a convex polyhedron. A vector $\mathbf{x} \in P$ is a* vertex of P if there exists a vector $\mathbf{c}$ such that $\mathbf{c}^T\mathbf{x} < \mathbf{c}^T\mathbf{y}$ for all $\mathbf{y}$, such that $\mathbf{y} \in P$ and $\mathbf{y} \neq \mathbf{x}$.

Therefore, for any vertex there exists a hyperplane, which is passing through this vertex and all the other points of this polyhedron (distinct from the vertex) lie in the same halfspace (that is, lie from one side of the hyperplane).

Theorem 3.6 *Let P be a nonempty polyhedron and $\mathbf{x} \in P$. $\mathbf{x}$ is a vertex if and only if it is an extreme point (that is, all extreme points are also vertices and vice versa).*

Remark 3.1 *A continuous, twice differentiable function of several variables is convex on a convex set if and only if its Hessian matrix of second partial derivatives is positive semidefinite on the interior of the convex set. In the case of univariate functions, the Hessian matrices are simply their second derivatives.*

Remark 3.2 *The Hessian matrix of a twice differential function of n variable $f(x_1, \ldots, x_n)$ is the matrix whose elements*

$$h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

To check if H is positive semidefinite, one can check all the principal minors of the matrix. If all of them are nonnegative, the matrix is positive semidefinite. There are other ways to check this property. We will talk about it later in this unit.

4 Local optimality for twice differentiable functions

In this section we review local optimality conditions for twice differential functions. This assumption is very strong and can not be removed very easily. We will only talk about functions of one variable (also known as univariate functions) and functions of two variables. In general, functions of two or more variables are called *multivariate functions.*

Recall that a point $a \in X \subset \mathbb{R}$ is said to be a stationary point of a univariate function $f(x)$ differentiable in X if its first derivative $f'(x) = 0$. In the case of multivariate functions, a point is stationary if all the partial derivative vanish (that is they are equal to zero) at this point.

In general, there are two types of points at which a function may attain its (local) maximum or minimum. These points are *stationary points and points where the function is not differentiable. These two types of points* *are also called* critical points. In this section we are only concentrating on twice differentiable functions and therefore we exclude the points where the derivative does not exist.

In the case when a function is defined on a closed set (for example, an interval or closed *hyperbox (a multidimensional analogue of a rectangle), it is also important to investigate the end-points.*

Univariate functions

Theorem 4.1 *Consider a twice differentiable function $f : X \to \mathbb{R}$. If a is a stationary points and $f''(x) > 0$ then this stationary point is a local minimum. If a is a stationary points and $f''(x) < 0$ then this stationary point is a local maximum.*

Note that if the second derivative is zero, a more detailed analysis is required to identify the nature of this stationary point.

Example 4.1 *Consider $f(x) = x^4$, $g(x) = -x^4$ and $h(x) = x^3$. Then $x = 0$ is a stationary point for these functions, since*

$$f'(0) = g'(0) = h'(0) = 0.$$

The second derivatives f'', g'' and h'' are also zero at $x = 0$. For $f(x)$ this stationary point is a local minimum, for $g(x)$ it is a local maximum, while for $h(x)$ it is neither (point of inflection).

Multivariate functions

Theorem 4.2 *Consider a twice differentiable function $f : X \times Y \to \mathbb{R}$. If (a, b) is a stationary points and*

- $H = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - (\frac{\partial^2 f}{\partial x \partial y})^2 < 0$ *then this stationary point is a saddle point;*
- $H = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - (\frac{\partial^2 f}{\partial x \partial y})^2 > 0$ *and* $\frac{\partial^2 f}{\partial x^2} > 0$ *then this stationary point is a local minimum;*
- $H = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - (\frac{\partial^2 f}{\partial x \partial y})^2 > 0$ *and* $\frac{\partial^2 f}{\partial x^2} < 0$ *then this stationary point is a local maximum.*

Note that if $H = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - (\frac{\partial^2 f}{\partial x \partial y})^2 = 0$ then more advanced analysis is required to identify the nature of the stationary point.

Example 4.2 *For the function*

$$f(x, y) = xy - (x + y)^2$$

find all stationary points and identify their nature.

Note, first of all, that

$$f(x,y) = xy - (x+y)^2 = -x^2 - y^2 - xy.$$

First order partial derivatives are:

$$\frac{\partial f}{\partial x} = -2x - y;\ \frac{\partial f}{\partial y} = -2y - x.$$

Second order partial derivatives are:

$$\frac{\partial^2 f}{\partial x^2} = -2;\ \frac{\partial^2 f}{\partial y^2} = -2;\ \frac{\partial^2 f}{\partial x \partial y} = -1.$$

All the stationary points are solutions of the system

$$\begin{cases} 2x + y = 0; \\ 2y + x = 0. \end{cases}$$

Therefore, the only stationary point is $(0,0)$*. At this point*

$$H = (-2)(-2) - (-1)^2 = 3 > 0.$$

Since $\frac{\partial^2 f}{\partial x^2} = -1 < 0$*, point* $(0,0)$ *is a local maximum.*

5 Linear programming

Problem formulation

A general linear programming problem or *LPP can be formulated as follows:*

$$\textit{minimise } \mathbf{c}^T\mathbf{x} \tag{7}$$

$$\begin{aligned}
\textit{subject to } \mathbf{a}_i^T\mathbf{x} &\geq b_i, & i \in M_1; && (8)\\
\mathbf{a}_i^T\mathbf{x} &\leq b_i, & i \in M_2; && (9)\\
\mathbf{a}_i^T\mathbf{x} &= b_i, & i \in M_3; && (10)\\
x_j &\geq 0, & j \in N_1; && (11)\\
x_j &\leq 0, & j \in N_2. && (12)
\end{aligned}$$

Now we need to explain all the components of this problem.

- *The variables* $x_1, \ldots, x_n$ *are called* decision variables.
- $\mathbf{c} = (c_1 \ldots, c_n)$ is the *cost vector.*
- *The function* $\mathbf{c}^T\mathbf{x}$ *in (7) is called the* objective function or *cost function.*
- *The set of equalities and inequalities (8)-(12) are called the* constraints.

Definition 5.1 *A vector* $\mathbf{x}$ *satisfying all the constraints is called a* feasible solution. The set of all feasible solutions is called the *feasible set or* feasible region.

If index j *is in neither* N_1 *nor* N_2 *then there are no restriction on the sign of* x_j*. In this case, we say that* x_j *is a* free variable or *unrestricted variable.*

Definition 5.2 *A feasible solution* $\mathbf{x}^*$ *that minimises the objective function is called an* optimal feasible solution or an *optimal solution.*

Definition 5.3 *If the optimal cost is* $-\infty$ *(minimisation), we say that the cost is* unbounded below and the problem itself is *unbounded problem.*

Note that any maximisation problem (maximising $\mathbf{c}^T\mathbf{x}$*) can be formulated as minimising* $-\mathbf{c}^T\mathbf{x}$*, see next section for details. Therefore, we can also define unbounded problems in the case of maximisation.*

Definition 5.4 *If the optimal cost is* $+\infty$ *(maximisation), we say that the cost is* unbounded above and the problem itself is *unbounded problem.*

Definition 5.5 *A LPP is called infeasible if the corresponding feasible set does not contain any point (that is, the corresponding system of linear equalities and inequalities does not have any solution).*

Example 5.1 *Consider the following linear programming problem:*

$$\begin{aligned} \textit{minimise } & 3x_1 - 2x_2 + 7x_3 && (13)\\ \textit{subject to } & x_1 + 5x_2 + x_4 \leq 2 && (14)\\ & 3x_1 - x_2 = 5 && (15)\\ & x_2 - x_3 \geq 3 && (16)\\ & x_1 \geq 0 && (17)\\ & x_2 \leq 0 && (18) \end{aligned}$$

- x_1, x_2, x_3 *and* x_4 *are the decision variables, whose values are to be chosen to minimise the objective function.* $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ *is the decision vector.*
- *The objective (cost) function is in (13) and the corresponding cost vector* $\mathbf{c}^T = (3, -2, 7, 0)$*.*
- *The constraints (14), (16), (18) and (17) are linear inequality constraints, while the constraint (15) is a linear equality constraint. The constraint (15) can be expressed as* $\mathbf{a}^T\mathbf{x} = 5$*, where* $\mathbf{a} = (3, -1, 0, 0)$*.*
- *The sign of the variables* x_3 *and* x_4 *is unrestricted, while* x_1 *is nonnegative and* x_2 *is nonpositive.*

Definition 5.6 *If a vector* $\mathbf{x}^*$ *satisfies a linear constraint as equality (equality or inequality constraints) then this constraint is* active or *bounding at* $\mathbf{x}^*$*.*

Consider the following LPP

$$\begin{aligned} \textit{minimise } & 3x_1 - 2x_2 + 7x_3\\ \textit{subject to } & x_1 + 5x_2 + x_3 \leq 2\\ & 3x_1 - x_2 \leq 5\\ & x_2 - x_3 \leq 3\\ & x_1 \geq 0\\ & x_2 \geq 0. \end{aligned}$$

*This problem can be expressed in the following way (*matrix form)

$$\begin{aligned} \text{minimise } & \mathbf{c}^T\mathbf{x}\\ \text{subject to } & \mathbf{A}\mathbf{x} \leq b\\ & x_1, x_2 \geq 0, \end{aligned}$$

where $\mathbf{c}^T = (3, -2, 7)$, $\mathbf{b} = (2, 5, 3)^T$ and

$$\mathbf{A} = \begin{bmatrix} 1 & 5 & 1 \\ 3 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Types of LPPs

Note, first of all, that any maximisation LLP can be reduced to minimisation. To achieve this, the cost vector $\mathbf{c}$ of the maximisation problem should be replaced by $-\mathbf{c}$ (and maximisation by minimisation).

An LPP of the form

$$\begin{aligned} \text{minimise } & \mathbf{c}^T\mathbf{x} \\ \text{subject to } & \mathbf{A}\mathbf{x} = b \\ & \mathbf{x} \geq 0 \end{aligned}$$

is said to be in *standard form. It can be shown that any LPP can be reduced to its standard form.*

- *Elimination of inequality constraints by introducing an additional* surplus variable or *slack variable s_i.*

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_j \Rightarrow \sum_{j=1}^{n} a_{ij}x_j + s_i = b_j.$$

- *Elimination of unrestricted (free) variables. Any real number can be presented as a difference of two non-negative numbers (there are several possibilities to construct this difference). Since any unrestricted variable x_j can be presented as $x_j^+ - x_j^-$, where $x_j^+, x_j^- \geq 0$.*

Therefore, any LPP can be expressed in its standard form, we only need to develop an algorithm to optimise LPPs in their standard forms.

Note that it is also possible to convert

- *equalities to inequalities:*

$$\sum_{j=1}^{n} a_{ij}x_j = b_j \Leftrightarrow \begin{cases} \sum_{j=1}^{n} a_{ij}x_j \leq b_j \\ \sum_{j=1}^{n} a_{ij}x_j \geq b_j \end{cases}$$

- *convert from* min *to* max *in the cost (objective) function and vice versa by multiplying it by -1:*

$$\textit{minimise} \sum_{j=1}^{n} c_j x_j \Leftrightarrow \textit{maximise} \sum_{j=1}^{n} -c_j x_j$$

and leaving the constraint unchanged. In this case, the original and obtained problems have the same optimal solutions; however, the their objective function optimal values have opposite signs.

Example 5.2 *[Example 1.3, Bertsimas and Tsitsiklis] The problem*

$$\text{minimise } 2x_1 + 4x_2,$$

$$\begin{aligned} \text{subject to } x_1 + x_2 &\geq 3 \\ 3x_1 + 2x_2 &= 14 \\ x_1 &\geq 0 \end{aligned}$$

can be expressed as

$$\text{minimise } 2x_1 + 4(x_2^+ - 4x_2^-),$$

$$\begin{aligned} \text{subject to } x_1 + x_2^+ - x_2^- + x_3 &= 3 \\ 3x_1 + 2(x_2^+ - x_2^-) &= 14 \\ x_1, x_2^+, x_2^-, x_3 &\geq 0 \end{aligned}$$

For example, given the feasible solution $(x_1, x_2) = (6, -2)$ to the original problem, we obtain the feasible solution

$$(x_1, x_2^+, x_2^-, x_3) = (6, 0, 2, 1)$$

to the standard form problem, which has the same cost. Conversely, given the feasible solution

$$(x_1, x_2^+, x_2^-, x_3) = (8, 1, 6, 0)$$

to the standard form problem, we obtain the feasible solution

$$(x_1, x_2) = (8, -5)$$

to the original problem with the same cost.

How to solve LPPs

Geometric method *This method is based on the following theorem.*

Theorem 5.1 *The maximum or minimum of a linear program, if it exists, will necessarily occur on a vertex (extreme point, corner point) of the feasible set.*

Proof: *Assume that in a linear program, an optimal solution* $\mathbf{x}$ *is not a vertex. In this case, for any vector* $\mathbf{c}$ *there are feasible points* $\mathbf{y}$ *and* $\mathbf{z}$*, such that* $\mathbf{cx} < \mathbf{cy}$ *and* $\mathbf{cx} > \mathbf{cz}$*. Therefore, since these conditions hold for any* $\mathbf{c}$*, it is also true for the cost vector of the linear program. Hence, indeed,* $\mathbf{x}$ *can not be a maximum or a minimum of this linear program.*

□

Definition 5.7 *A set, where the objective function value remains constant, is called* level set.

Note that in the case of linear objective function (LPPs) level sets are parallel lines.

Example 5.3 *Consider the following linear program*

$$\min 3x + y$$

subject to

$$x + y \leq 4$$
$$x, y \geq 0$$

The feasible region is a triangle with vertices

$$(x = 0, y = 0),\ (x = 4, y = 0),\ (x = 0, y = 4),$$

see Figiure 6. The green line (they are parallel!!!) represent level sets. The structure of these sets demonstrate that the optimal objective function value is 12, reached at a corner point $(x = 4, y = 0)$ *of the feasible region.*

Note that in Theorem 5.1 it is assumed that the minimum or maximum should exist. There are two main types of LPPs where optimal solutions do not exist.

1. *Infeasible problems: the corresponding feasible sets contain no points, that is the corresponding systems of linear inequalities does not have any solution.*

2. *Unbounded problems: the corresponding feasible set is unbounded and the cost function value is unbounded (positive for maximisation problems and and negative for minimisation problems).*

Example 5.4 *Consider the following LPP*

$$\min x + 2y$$

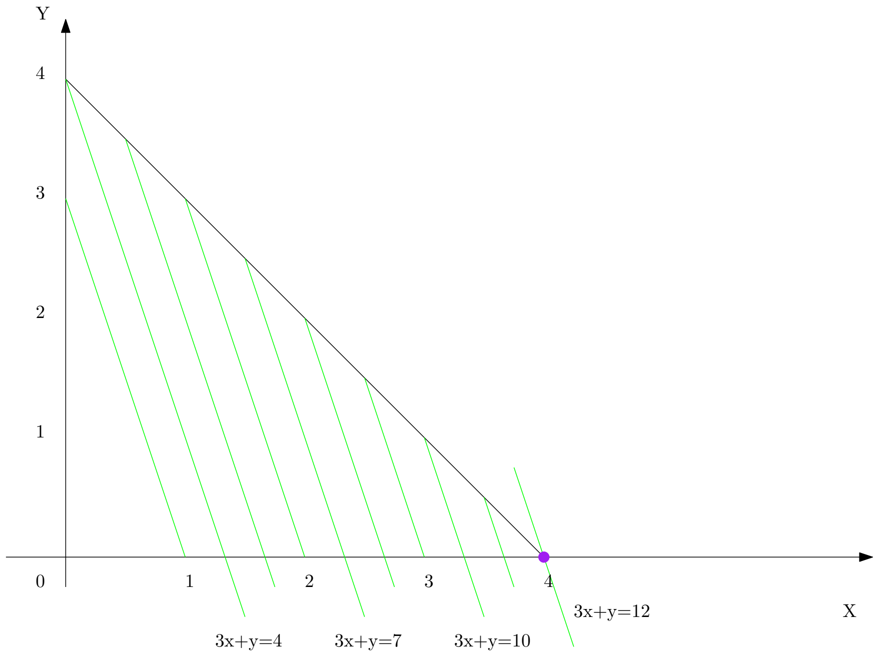

Figure 6: Corner point Theorem.

subject to

$$x + y \leq 1$$
$$x, y \geq 2$$

In this case, the region where the first inequalities are satisfied is covered with blue lines, while the region where the last two inequalities are satisfied is covered with red lines (see Figiure 7). These regions do not intersect (that is, they do not have any common points). Therefore, there is no point where all three inequalities are satisfied.

Example 5.5 *Consider the following LPP*

$$\max x + y$$

subject to

$$x + y \geq 3$$
$$x, y \geq 1$$

The feasible region (see Figiure 8) is covered with blue lines. From the picture one can see that the optimal cost function value is unbounded (infinitely large).

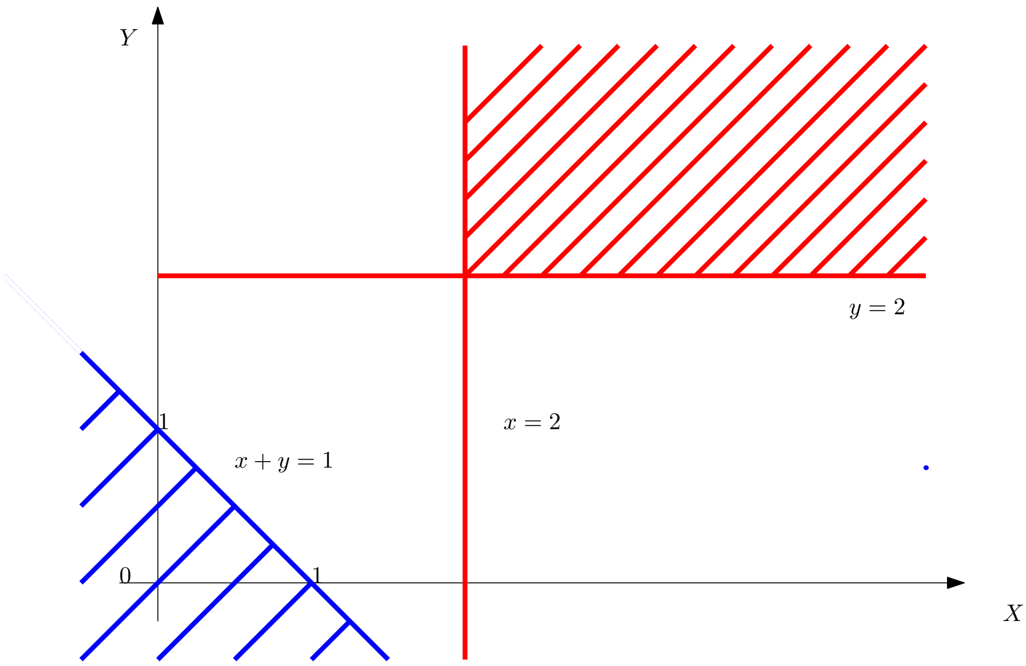

Figure 7: Infeasible problem.

Note that

- *in some LPPs the corresponding feasible sets may be unbounded, but these problems have optimal solutions.*
- *some LPPs may have more than one solution, that is the optimal value of the cost function is reached at more that one point.*

According to Theorem 5.1, at least one of the optimal solutions is a vertex. It can be demonstrated that the set of optimal solution is convex, that is, for any two optimal solution, all the points from the segment that joins these solutions are optimal solutions as well.

Example 5.6 *Consider the following LPP*

$$\min x + y$$

subject to

$$\begin{aligned} x + y &\geq 3 \\ x, y &\geq 1 \end{aligned}$$

This problem is similar to the one from Example 5.5, but the maximisation of the cost function is changed to minimisation. Then the optimal value of

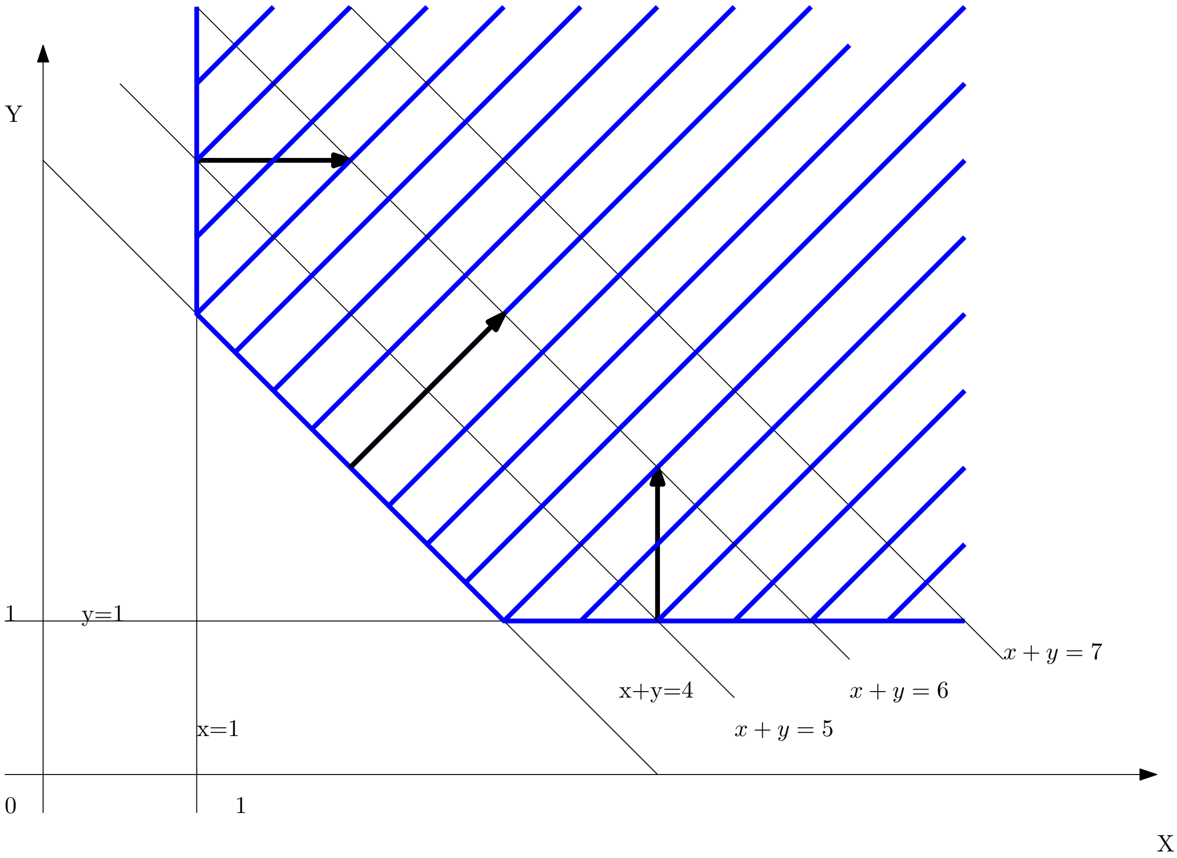

Figure 8: Unbounded problem.

the objective function is 4. This value is reached at two distinct vertices ($x = 1, y = 3$) *and* ($x = 3, y = 1$). *In fact, any point from the segment that joins these vertices gives the same cost function value. Therefore, all the points from this segment are optimal solutions (see Figiure 8).*

Simplex method and Interior points method *These two techniques are powerful optimisation tools, that are part of most optimisation software packages, working with LPPs.*

The simplex method was invented by Dantzig in 1947. A comprehensive description of this method was published in 1963 (same author). The idea behind this method is that an optimal solution (if exists) can not appear in the interior of the feasible set (polyhedron), but at a vertex or on an edge. The simplex method solves LPPs by visiting extreme points (vertices) of the boundary of the feasible set, improving the objective function value at every iteration. This method has been applied successfully to a wide range of LPPs. However, when the dimension is increasing, the number of vertices becomes very large.

In 1984 Karmarkar developed a new method, that was moving from one *interior point (not a vertex!) to another, improving the objective function value at each iteration. This method (known as the* interior point method or *IPM) has better theoretical properties and performs better than the simplex method on very large problems.*

These methods are included in most optimisation software. You will work with these methods in your computer laboratory classes.

6 More examples and applications

Production models

A company produces n different products using m different raw materials. Let b_i, $i = 1, \ldots, m$, be the available amount of the ith raw material. The jth product ($j = 1, \ldots n$), requires a_{ij} units of the ith material and can be sold at the price c_j dollars per unit produced. The company has to decide how much of each product to produce in order to maximize its total revenue.

Let x_j, $j = 1, \ldots, n$, be the amount of the jth product. Then, the problem can be formulated as follows:

$$\text{maximize} \sum_{j=1}^{n} c_j x_j \tag{19}$$

$$\text{subject to} \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \; i = 1, \ldots, m, \tag{20}$$

$$x_j \geq 0, \; j = 1, \ldots, n. \tag{21}$$

$$\tag{22}$$

Non-linear problems involving absolute values

Consider the following problem:

$$\text{minimise} \sum_{i=1}^{n} c_i |x_i| \tag{23}$$

$$\text{subject to } \mathbf{Ax} \geq \mathbf{b}. \tag{24}$$

$$\tag{25}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and the cost coefficients c_i, $i = 1, \ldots n$ are assumed to be nonnegative. The objective function is convex (as a sum of convex functions). Note that $|x_i|$ is the smallest number z_i, such that $x_i \leq z_i$ and $-x_i \leq z_i$, then we obtain an equivalent linear programming formulation

$$\text{minimise} \sum_{i=1}^{n} c_i z_i \tag{26}$$

$$\text{subject to } \mathbf{Ax} \geq \mathbf{b} \tag{27}$$

$$x_i \leq z_i, \; i = 1, \ldots, n, \tag{28}$$

$$-x_i \leq z_i, \; i = 1, \ldots, n. \tag{29}$$

Now consider another problem:

$$\text{minimise} \max_{j=1,\ldots,m} |\mathbf{c}_j \mathbf{x} + d_j|. \tag{30}$$

$$\tag{31}$$

Define a new variable

$$z = \max_{j=1,\dots,m} |\mathbf{c}_j \mathbf{x} + d_j|,$$

then

$$\mathbf{c}_j^T \mathbf{x} + d_j \le z \text{ and } -\mathbf{c}_j^T \mathbf{x} - d_j \le z, \; j = 1, \dots, m.$$

Then the corresponding linear programming formulation is

$$\min \; z \tag{32}$$

$$\text{subject to } \mathbf{c}_j^T \mathbf{x} + d_j \le z, \; j = 1, \dots, m \tag{33}$$

$$-(\mathbf{c}_j^T \mathbf{x} + d_j) \le z, \; j = 1, \dots, m \tag{34}$$

$$\mathbf{x} \in \mathbb{R}^n, \; z \in \mathbb{R}, \tag{35}$$

where z *and* $x_i, \; i = 1, \dots, n$ *are the decision variables.*

Data fitting

We are given n *data points of the form* $(\mathbf{a}_i^T, b_i), \; i = 1, \dots, n,$ *where* $\mathbf{a}^T \in \mathbb{R}^n$ *and* $b_i \in \mathbb{R}$*. The goal is to build a model that predicts the value of the variable* b *for any given vector* $\mathbf{a}^T$ *through a parameter vector* $\mathbf{x}$*. For a given parameter vector* $\mathbf{x}$*, the* residual, or *approximation error, at the* i*-th data point is* $|b_i - \mathbf{a}_i^T \mathbf{x}|$*.*

Assume now that we can choose $\mathbf{x}$ *(parameter vector) is such a way, that the corresponding model is as accurate as possible, that is, the model that results in small residuals (on available data).*

Example 6.1 *Consider polynomial approximation, that is, approximation by polynomials. A polynomial of degree* m *can be presented as*

$$p(t) = \sum_{j=0}^{m} a_j t^j = \mathbf{a}^T \mathbf{T},$$

where $\mathbf{a}^T = (a_0, a_1, \dots, a_m)$ *and* $\mathbf{T} = (1, t, \dots, t^m)^T$*.*

Therefore, if a scalar b_i *is assigned to* t_i*,* $i = 1, \dots, n$*, then the residual*

$$b_i - p(t_i) = b_i - \mathbf{a}_i^T \mathbf{T}_i,$$

where $\mathbf{a}_i^T = (a_{i0}, a_{i1}, \dots, a_{im})$ *and* $\mathbf{T}_i = (1, t_i, \dots, t_i^m)^T$*.*

Chebyshev approximation *One possibility is to minimise the largest residual (*Chebyshev approximation problem):

$$\text{minimise} \max_{i=1,\dots,m} |b_i - \mathbf{a}_i^T \mathbf{x}|$$

subject to

$$\mathbf{x} \in \mathbb{R}.$$

This problem can be also formulated as a linear programming problem:

$$\min \ z \tag{36}$$

$$\text{subject to } \mathbf{a}_j^T\mathbf{x} - b_j \leq z, \ j = 1, \ldots, m \tag{37}$$

$$-\mathbf{a}_j^T\mathbf{x} + b_j \leq z, \ j = 1, \ldots, m \tag{38}$$

$$\mathbf{x} \in \mathbb{R}^n, \ z \in \mathbb{R}, \tag{39}$$

where z and $x_i, \ i = 1, \ldots, n$ are the decision variables.

Sum of deviations Consider an alternative formulation, where the objective function is

$$\sum_{i=1}^{m} |b_i - \mathbf{a}_i^T\mathbf{x}| \tag{40}$$

Consider new variables $z_i = |b_i - \mathbf{a}_i^T\mathbf{x}|, \ i = 1, \ldots, m$, that is

$$b_i - \mathbf{a}_i^T\mathbf{x} \leq z_i, \ i = 1, \ldots, m, \tag{41}$$

$$-(b_i - \mathbf{a}_i\mathbf{x}) \leq z_i, \ i = 1, \ldots, m. \tag{42}$$

This is a linear programming problem, $z_1, \ldots, z_m$ and $\mathbf{x}$ are the decision variables: programming problem:

$$\min \ \sum_{i=1}^{m} z_i \tag{43}$$

$$\text{subject to } b_i - \mathbf{a}_i^T\mathbf{x} \leq z_i, \ i = 1, \ldots, m, \tag{44}$$

$$-(b_i - \mathbf{a}_i\mathbf{x}) \leq z_i, \ i = 1, \ldots, m. \tag{45}$$

Least squares In some practical applications, it is possible to use a quadratic based cost function:

$$\sum (b_i - \mathbf{a}_i^T x)^2.$$

This cost function is often called a *least squares fit. In most cases, this problem is easier than linear programming (can be solved using calculus methods). We will study this problem in details in Convex Optimisation.*