Linear Programming

Alberto Santini

Fall 2024

Many examples in these lecture notes are adapted from popular books:

- Alexander Schrijver (1998). *Theory of linear and integer programming.* Wiley. ISBN: 0-471-98232-6.
- Vašek Chvátal (1983). *Linear Programming.* W.H. Freeman and Company. ISBN: 0-716-71195-8.
- Laurence Wolsey (2020). *Integer Programming.* 2nd Edition. Wiley. ISBN: 978-1-119-60653-6.
- Silvano Martello and Paolo Toth (1990). *Knapsack Problems: algorithms and computer implementations.* Wiley. ISBN: 978-0-471-92420-3.

1 Linear programming

A **Linear Programme** (LP) is a mathematical programming model with a finite number of continuous variables, a linear objective function and a finite set of linear constraints. In particular, we start by considering LPs written in a special way, called the **standard form**:

$$\max \quad \sum_{j=1}^{n} c_j x_j \tag{1}$$

$$\text{subject to} \quad \sum_{j=1}^{n} a_{ij} x_j \leq b_i \qquad \forall i \in \{1, \dots, m\} \tag{2}$$

$$x_j \geq 0 \qquad \forall j \in \{1, \dots, n\}. \tag{3}$$

This model has n variables (indexed by j) and m constraints (indexed by i). The variables are denoted with $x_1, \dots, x_n$ and are all continuous and non-negative. Constants $c_1, \dots, c_n \in \mathbb{R}$ are the **objective coefficients** of the variables. Constants $a_{ij} \in \mathbb{R}$ are the **constraint matrix coefficients** because it is often helpful to see them as entries of a matrix $A = (a_{ij})_{i=1,\dots,m}^{j=1,\dots,n}$. Finally, constants $b_1, \dots, b_m \in \mathbb{R}$ are the **RHS coefficients** and owe their name to being on the right-hand side of the constraints. Formulation (1)–(3) has a few defining characteristics:

1. It is a *maximisation* problem, i.e., we are momentarily disregarding minimisation problems.
2. All its constraints are "$\leq$"-inequalities, i.e., we are momentarily disregarding equalities and "$\geq$"-inequalities.
3. All variables have the same definition domain, the non-negative real numbers $\mathbb{R}_0^+$. Constraints (3) defining these domains are called the **non-negativity constraints**.

A list of numbers $x_1, \dots, x_n \in \mathbb{R}_0^+$ which, when substituted to the corresponding variables in (1)–(3), satisfies all inequalities is called a **feasible solution** of the model. Such a solution has a corresponding **objective value** $\sum_{j=1}^{n} c_j x_j$. A feasible solution which attains the highest possible objective value is called an **optimal solution**.

Remark 1. Not all LPs have feasible solutions. Trivially, the following LP does not have any feasible solution:

$$\begin{aligned} \max \quad & x && (4)\\ \text{subject to} \quad & x \leq 2 && (5)\\ & -x \leq -3 && (6)\\ & x \geq 0. && (7) \end{aligned}$$

Rewriting (6) as $x \geq 3$, it is immediately clear that no number x can simultaneously satisfy (5) and (6). An LP which admits a feasible solution is called a **feasible problem**. Otherwise, it is called **infeasible**. The set of all feasible solutions is called the **feasible region** (or the feasible set).

Remark 2. Not all feasible problems have an optimal solution. Again trivially, consider the following LP:

$$\begin{aligned} \max \quad & x && (8)\\ \text{subject to} \quad & x \geq 0. && (9) \end{aligned}$$

The objective function is unbounded, in the sense that for any value Z attained by a feasible solution (in our case by solution $x = Z$), there is always another feasible solution which yields a better objective value (for example, $x = Z + 1$). A feasible problem without an optimal solution is called an **unbonded problem**. Otherwise, it is called a **bounded problem**.

Remark 3. Finally, let us note that the optimal solution of a feasible bounded problem is not necessarily unique. Consider, for example, the following LP:

$$\begin{aligned} \max \quad & 2x_1 + 2x_2 && (10)\\ \text{subject to} \quad & x_1 + x_2 \leq 1 && (11)\\ & x_1 \leq \frac{2}{3} && (12)\\ & x_2 \leq \frac{2}{3} && (13)\\ & x_1 \geq 0 && (14)\\ & x_2 \geq 0. && (15) \end{aligned}$$

It is easy to see that, for any $x_1 \in \left[\frac{1}{3}, \frac{2}{3}\right]$, choosing $x_2 = 1 - x_1$ gives an optimal solution with objective value 2.

Exercise 1. Consider two vectors of the "non-negative orthant,"

$$\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}_0^{+n},$$
$$\vec{x}' = (x'_1, \ldots, x'_n) \in \mathbb{R}_0^{+n}.$$

Note that we can naturally define sum and multiplication-by-scalar in $\mathbb{R}_0^{+n}$:

$$\vec{x} + \vec{x}' = (x_1 + x'_1, \ldots, x_n + x'_n),$$
$$\lambda \cdot \vec{x} = (\lambda x_1, \ldots, \lambda x_n), \quad \lambda \geq 0.$$

Vectors in the form $\lambda\vec{x} + (1 - \lambda)\vec{x}'$, with $\lambda \in [0, 1]$, are called **convex combinations** of $\vec{x}$ and $\vec{x}'$.

Show that if $\vec{x}$ and $\vec{x}'$ are two optimal solutions of an LP in standard form, then all their convex combinations are also optimal solutions.

Exercise 2. Continuing Exercise 1, show that if $\vec{x}$ and $\vec{x}'$ are two optimal solutions of an LP in standard form, then vectors $\vec{c} = (c_1, \dots, c_n)$ and $\vec{x} - \vec{x}'$ are orthogonal in $\mathbb{R}^n$. Recall that $c_1, \dots, c_n$ are the objective coefficients of the variables.

2 LPs in non-standard form

We have introduced LPs in their standard form (1)–(3). But, surely, there are interesting models involving minimisation, or "$\geq$"-inequalities, equalities, unconstrained variables, etc. Do we have to disregard all these models? Of course not! All LPs can be turned into standard form LPs, and therefore, restricting ourselves to this form is *not* restrictive.

We now list some standard techniques to change LPs in general form to LPs in standard form. We start from the optimisation sense noticing that minimising a linear function $\sum_{j=1}^n c'_j x_j$ is equivalent to maximising function $\sum_{j=1}^n -c'_j x_j$. Therefore, any minimisation problem can be turned into a maximisation problem by setting $c_j = -c'_j$ in (1).

Analogously, if we need a non-positive variable $x'_j \leq 0$, we can define a new variable $x_j \geq 0$ and replace $x'_j = -x_j$ throughout the model.

A more interesting case is that of an unrestricted variable $x'_j \in \mathbb{R}$. In this case, we introduce *two* new variables, say $x_j^+, x_j^- \geq 0$, and we replace $x'_j = x_j^+ - x_j^-$. Note that any non-negative value $v \geq 0$ attained by x'_j can be obtained by x_j^+ and x_j^-: for example, by $x_j^+ = v$ and $x_j^- = 0$. But every non-positive value $w \leq 0$ attained by x'_j can also be obtained, for example with $x_j^+ = 0$ and $x_j^- = -w$. Therefore, any solution which involves the unrestricted x'_j has at least one corresponding solution in the modified problem involving the non-negative x_j^+ and x_j^-.

Finally, we will deal with "$\geq$" and "$=$" constraints. The first case is immediately solved because any inequality of the form

$$\sum_{j=1}^n a'_{ij} x_j \geq b'_i$$

can be rewritten as

$$-\sum_{j=1}^n a'_{ij} x_j \leq -b'_i.$$

Therefore, we can replace the "$\geq$"-inequality with a "$\leq$"-inequality in standard form, with coefficients $a_{ij} = -a'_{ij}$ and $b_i = -b'_i$.

In the second case, we have an equality of the form

$$\sum_{j=1}^n a_{ij} x_j = b_i.$$

Such an equality can be written as a pair of inequalities:

$$\sum_{j=1}^n a_{ij} x_j \leq b_i,$$
$$\sum_{j=1}^n a_{ij} x_j \geq b_i.$$

And, using the "≥"-to-"≤" transformation that we already know, we are finally left with two standard-form inequalities:

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i,$$
$$\sum_{j=1}^{n} -a_{ij} x_j \leq -b_i.$$

Exercise 3. The following exercise uses a bit of an obsolete and unpleasant setting: oil production. Indeed, it is taken from Chvatal's book, which was written more than forty years ago.

An oil refinery produces four types of raw petrol: RAWA, RAWB, RAWC, and RAWD. Two crucial characteristics of each raw petrol type are its performance number PN and vapour pressure RVP. The following table describes the characteristics of each petrol, together with daily production levels:

Petrol	**PN**	**RVP**	**Daily production** (barrel)
RAWA	107	5	3814
RAWB	93	8	2666
RAWC	87	4	4016
RAWD	108	21	1300

Each kind of petrol can be sold raw for a profit of 4.83 €/barrel independent of its type. Petrols can also be blended to produce jet fuel. There are two types of jet fuel:

- JETA must have PN of at least 100 and RVP of at most 7. It sells at 6.45 €/barrel.
- JETB must have PN of at least 91 and RVP of at most 7. It sells at 5.91 €/barrel.

The PN and RVP of a blend is the weighted average of their PN and RVP. For example, a blend of 60% RAWA and 40% RAWB will have

$$\text{PN} = 0.6 \cdot 107 + 0.4 \cdot 93 = 101.4, \quad \text{RVP} = 0.6 \cdot 5 + 0.4 \cdot 8 = 6.2.$$

Formulate an LP (not necessarily in standard form) for the refinery's production plan. The objective is to earn the maximum profit.

Identify which symbols are variables and which (if any) represent the problem's parameters. Do not forget to add the variable definition domains. Also, add an explanation of the variables' meaning in plain terms.

Exercise 4. Given a sample of real-valued observations $S = \{y_1, \ldots, y_n\}$, a **median** of S is any value $m \in \mathbb{R}$ such that the number of observations smaller and larger than m is the same. Assume that S contains an odd number of observations ($n = 2k + 1$) and, without loss of generality, that they are sorted ($y_1 \leq y_2 \leq \ldots \leq y_n$). Then the unique median is $m = y_{k+1}$:

$$\underbrace{y_1 \leq \ldots \leq y_k}_{k \text{ observations}} \leq y_{k+1} \leq \underbrace{y_{k+2} \leq \ldots \leq y_{2k+1}}_{k \text{ observations}}.$$

If S contains an even number of observations ($n = 2k$), again sorted by index, then any number m between y_k and y_{k+1} is a median:

$$\underbrace{y_1 \leq \ldots \leq y_k}_{k \text{ observations}} \leq m \leq \underbrace{y_{k+1} \leq \ldots \leq y_{2k}}_{k \text{ observations}}.$$

Write a linear programme to find the median of a given sample. *Hint:* you only need one variable, m. *Hint:* first, try to write a non-linear programme using the absolute value in its objective function; then, linearise it.

Exercise 5. A **quantile** is a generalisation of the median. In the same setting as Exercise 4, an α-quantile is a number q_α such that a fraction $\alpha \in [0, 1]$ of the observations is below it. For example, when $\alpha = 0.4$, 40% of the observations are not larger than q_α and the remaining 60% are not smaller than q_α. Write a linear programme to find the α-quantile of a given sample. *Hint:* a good keyword to search is "importance sampling"; in fact, it turns out that you can estimate a quantile by minimising a weighted sum of absolute deviations.

Exercise 6. Quantile regression is a type of regression used to estimate the conditional quantile of the dependent variable (compare this with OLS, which estimates the conditional mean). It assumes that the desired α-quantile is a linear function of the independent variables $X_1, \ldots, X_p$. Using the same notation of the Lasso regression example in the "Introduction to Optimisation" lecture notes, let x_{ij} be the value of the j-the independent variable of the i-th observation and y_i be the value of the dependent variable. Denote with $\beta_0, \beta_1, \ldots, \beta_p$ the model parameters so that the α-quantile is expressed as

$$q_\alpha = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p.$$

To perform quantile regression, we are looking for the model parameters minimising the following objective function:

$$\sum_{i \,:\, y_i < \vec{x}_i^\top \vec{\beta}} (1 - \alpha) \big| y_i - \vec{x}_i^\top \vec{\beta} \big| + \sum_{i \,:\, y_i \geq \vec{x}_i^\top \vec{\beta}} \alpha \big| y_i - \vec{x}_i^\top \vec{\beta} \big|, \tag{16}$$

where $\vec{x}_i = (1, x_{i1}, \ldots, x_{ip})$ and $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$. Equation (16) wants to minimise the absolute deviation between the dependent variables y_i and the estimates $q_{\alpha,i} = \vec{x}_i^\top \vec{\beta}$, but giving different weights to the estimates relative to points i, based on their dependent variable being smaller (weight $1 - \alpha$) or larger (weight α) than the estimate. Write a linear programme to minimise objective function (16) and find the optimal parameters $\beta_0, \ldots, \beta_p$.

3 Graphical approach to simple LPs

Given an LP with two variables, it is easy to draw its **feasible region**, i.e., the region defined by inequalities (2) and (3). Consider, for example, the LP presented in Remark 3. Figure 1 represents this LP in graphical form.

Each constraint corresponds to a thick black line, and feasible solutions lie on one side of the line (the one indicated by the black arrow). The shaded area denotes the feasible region and is called a **polytope**. A polytope is a set of points defined by a finite number of linear inequalities. (This definition is valid in any dimension, not just the 2-dimensional figure presented above.) The large blue arrow shows the direction of the increase of the objective function. As we have seen in Exercise 2, this arrow is orthogonal to the segment BC containing all the optimal points.

Points O, A, B, C, and D are the **vertices** of the polytope. In the general n-dimensional case, i.e., when we have an LP in standard form with n variables, a vertex is a point which satisfies with equality at least n of the linear inequalities which define the polytope. Recall that a a problem with n variables has at least n non-negativity constraints and, therefore, the number of constraints is never smaller than the number of variables.

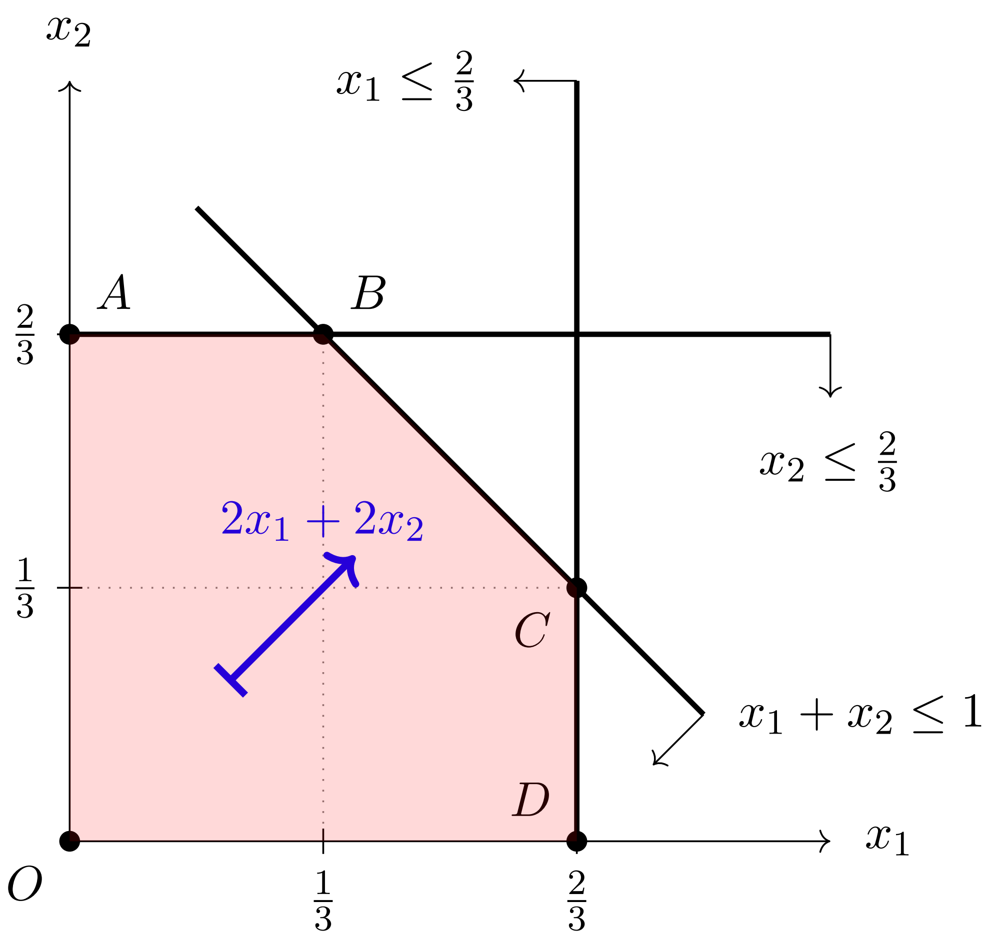

Figure 1: Graphical representation of the LP introduced in Remark 3

Exercise 7. Given a bounded feasible LP, show that there is at least one optimal solution, which is a vertex. *Hint:* you can use Caratheodory's theorem.

4 A brief overview of solution algorithms for LPs

The first proposed algorithm for solving LPs is the **simplex method**. This algorithm has both an algebraic and a geometric interpretation. We explain the algebraic interpretation in the extra material "Simplex" and we focus here on the more intuitive geometric interpretation. The key observation underlying this algorithm is that it is possible to restrict the search for the optimal solution of an LP to the solve vertices of the feasible region polytope, as proved in Exercise 7. Therefore, the algorithm starts from any feasible vertex and, at each iteration, moves to an adjacent vertex with a better objective value. When no improvement is possible any more, the last visited vertex will be an optimum.

This description is a bit hand-wavy. For example, how do we find an initial vertex? How to enumerate all neighbouring vertices (i.e., the vertices linked by an edge to the current one)? We refer to the "Simplex" extra lecture notes and to Chvatal's book to answer these questions. Here, we only mention that the simplex algorithm does not enjoy very strong theoretical guarantees on its runtime but is usually fast in practice. In particular, this algorithm can run for a number of iterations that is exponential in the number of variables. However, it takes specially constructed LPs for this worst-case to happen and, for practically relevant LPs, the average-case runtime of the simplex is much lower.

Another algorithm is the **Ellipsoid method**, which keeps track of an ellipsoid containing the feasible region and makes it increasingly smaller (in volume) at each iteration. This algorithm is slow in practice, but it was the first polynomial-time algorithm to solve an LP. Therefore, it established the theoretical complexity of linear programming as a polynomially-solvable problem.

Many modern LP solvers implement both (a variation of) the simplex method and **interior point** algorithms such as the **barrier method**. These algorithms search for the optimum of an LP by sequentially moving within points in the interior of the feasible region (as opposed to the simplex method which moves from vertex to vertex).

5 Solutions

Exercise 1

First note that a convex combination of two feasible solutions is feasible. For each constraint i, in fact, because both x and x' are feasible, we have

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \tag{17}$$

$$\sum_{j=1}^{n} a_{ij} x'_j \leq b_i. \tag{18}$$

Taking a linear combination of (17) and (18) with coefficients λ and $1 - \lambda$, one gets:

$$\sum_{j=1}^{n} \lambda a_{ij} x_j + \sum_{j=1}^{n} (1-\lambda) a_{ij} x'_j \leq \lambda b_i + (1-\lambda) b_i. \tag{19}$$

Rewriting the above inequality as

$$\sum_{j=1}^{n} a_{ij}\big(\lambda x_j + (1-\lambda)x'_j\big) \leq b_i, \tag{20}$$

we prove the feasibility of $\lambda x + (1-\lambda)x'$. Finally, we prove that the convex combination is optimal. Denote with z^* the optimal objective value, i.e.,

$$z^* = \sum_{j=1}^{n} c_j x_j = \sum_{j=1}^{n} c_j x'_j. \tag{21}$$

Then the objective value of the convex combination is

$$\sum_{j=1^n} c_j\big(\lambda x_j + (1-\lambda)x'_j\big) = \lambda z^* + (1-\lambda)z^* = z^*. \tag{22}$$

Exercise 2

Let z^* be the optimal solution value, i.e., $c^\top x = c^\top x' = z^*$. Then c and $x - x'$ are orthogonal because:

$$c^\top(x - x') = c^\top x - c^\top x' = z^* - z^* = 0. \tag{23}$$

Exercise 3

Let $\mathcal{P}$ be the set of petrols and $\mathcal{F}$ the set of fuels. Let $p_i, r_i, d_i \in \mathbb{R}^+$ be, respectively, the PN, the RVP, and the daily production of petrol $i \in \mathcal{P}$. Let $P_j, R_j, \pi_j \in \mathbb{R}^+$ be, respectively, the minimum PN, the maximum RVP, and the unit selling price of fuel $j \in \mathcal{F}$. Finally, let π^{raw} be the selling price of raw fuel. A possible model for this problem uses variables $x_{ij} > 0$, representing the amount of petrol $i \in \mathcal{P}$ used to make fuel $j \in \mathcal{F}$. A few things to note:

- The total amount of fuel j produced is given by the sum of the petrols blended to make the fuel:

$$\text{amount of fuel } j \text{ produced } = \sum_{i\in\mathcal{P}} x_{ij}. \tag{24}$$

- The fraction of petrol i in the blend of fuel j is given by:

$$\frac{x_{ij}}{\sum_{k\in\mathcal{P}} x_{kj}}. \tag{25}$$

- The amount of petrol i which is not used to make any fuel and, therefore, is sold raw is:

$$d_i - \sum_{j\in\mathcal{F}} x_{ij}. \tag{26}$$

A natural way of writing down the objective function of the problem is by considering its two components: the profit obtained from making fuel, and the profit obtained by selling raw petrol:

$$\max \quad \sum_{j\in\mathcal{F}} \pi_j \sum_{i\in\mathcal{P}} x_{ij} + \pi^{\text{raw}} \sum_{i\in\mathcal{P}} \Bigg(d_i - \sum_{j\in\mathcal{F}} x_{ij}\Bigg), \tag{27}$$

in which we use (24) and (26). After making some algebraic manipulations and realising that term $\pi^{\text{raw}} \sum_{i \in \mathcal{P}} d_i$ is a constant, an equivalent objective function is:

$$\max \quad \sum_{i \in \mathcal{P}} \sum_{j \in \mathcal{F}} (\pi_j - \pi^{\text{raw}}) x_{ij}. \tag{28}$$

Any solution optimal according to (27) is also optimal according to (28), but their objective values will differ by the above-mentioned constant.

The constraints of the LP are the following:

- To ensure that no more petrol is used than it is produced daily:

$$\sum_{j \in \mathcal{F}} x_{ij} \leq d_i \quad \forall i \in \mathcal{P}. \tag{29}$$

- To ensure that the minimum PN requirements are met, using (25), one can write:

$$\sum_{i \in \mathcal{P}} p_i x_{ij} \geq P_j \sum_{i \in \mathcal{P}} x_{ij} \quad \forall j \in \mathcal{F}. \tag{30}$$

- Analogously, to ensure that the maximum RVP requirements are met:

$$\sum_{i \in \mathcal{P}} r_i x_{ij} \leq R_j \sum_{i \in \mathcal{P}} x_{ij} \quad \forall j \in \mathcal{F}. \tag{31}$$

Exercise 4

The problem of finding the median can be written as a non-linear unconstrained minimisation problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} |y_i - m| \\ \text{s.t.} \quad & m \in \mathbb{R}. \end{aligned}$$

To transform this problem into a Linear Programme, we must linearise the absolute value in the objective function. There are at least two ways of doing so.

- Introduce non-negative variables $z_i \geq 0$. If we manage to devise some constraints such that each z_i holds the value of $|y_i - m|$, the objective function will be $\sum_{i=1}^{n} z_i$. Let us, then, construct the required constraints.

  Observe that, for any number $\gamma \in \mathbb{R}$, it holds both that $|\gamma| \geq \gamma$ and $|\gamma| \geq -\gamma$. Therefore, if we want z_i to have the value of $|y_i - m|$, the following two inequalities will be valid: $z_i \geq y_i - m$ and $z_i \geq m - y_i$ for all $i \in \{1, \ldots, n\}$.

  One of these constraints will be moot and the other one will be active, depending on the sign of $y_i - m$. For example, if $y_i - m \leq 0$ then the first constraint is moot because its right-hand side is non-positive and the constraint is not binding z_i any more than its non-negativity does. Conversely, the second constraint is active because the right-hand side is non-negative and, unless we are in the special case $y_i = m$, it is even strictly positive. Therefore, this constraint gives a higher lower bound for z_i compared with the non-negativity lower bound $z_i \geq 0$.

We are penalising the z_i's in the objective function. Therefore, for each z_i, the active constraint will be tight.

Bringing together the above observations we obtain the following Linear Programme:

$$\begin{aligned}
\min \quad & \sum_{i=1}^{n} z_i \\
\text{s.t.} \quad & z_i \geq y_i - m && \forall i \in \{1, \ldots, n\} \\
& z_i \geq m - y_i && \forall i \in \{1, \ldots, n\} \\
& z_i \geq 0 && \forall i \in \{1, \ldots, n\} \\
& m \in \mathbb{R}.
\end{aligned}$$

- The second way to linearise the absolute value appears slightly more complicated, but will be useful when solving the next exercises. It involves the creation of two new sets of non-negative variables, z_i^+, z_i^- for each $i \in \{1, \ldots, n\}$. The idea is that, if $y_i - m \geq 0$, then $z_i^+ = y_i - m$ and $z_i^- = 0$. Conversely, if $y_i - m \leq 0$, then $z_i^- = m - y_i$ and $z_i^+ = 0$.

  If we can devise constraints to ensure the above relations, the objective function becomes $\sum_{i=1}^{n}(z_i^+ + z_i^-)$. Because, for each i, only one between $y_i - m$ and $m - y_i$ will be non-negative (while both z_i^+ and z_i^- are non-negative by definition), it is sufficient to impose the following constraint: $z_i^+ - z_i^- = y_i - m$ for all $i \in \{1, \ldots, m\}$.

  Therefore, the Linear Programme we are looking for is:

$$\begin{aligned}
\min \quad & \sum_{i=1}^{n} (z_i^+ + z_i^-) \\
\text{s.t.} \quad & z_i^+ - z_i^- = y_i - m && \forall i \in \{1, \ldots, n\} \\
& z_i^+ \geq 0 && \forall i \in \{1, \ldots, n\} \\
& z_i^- \geq 0 && \forall i \in \{1, \ldots, n\} \\
& m \in \mathbb{R}.
\end{aligned}$$

Exercise 5

The non-linear programme to solve is:

$$\begin{aligned}
\min \quad & \sum_{i \,:\, y_i < q_\alpha} (1 - \alpha)\big|y_i - q_\alpha\big| + \sum_{i \,:\, y_i \geq q_\alpha} \alpha\big|y_i - q_\alpha\big| \\
\text{s.t.} \quad & q_a \in \mathbb{R}.
\end{aligned}$$

The linearisation is trivial if we start from the solution of Exercise 4:

$$\begin{aligned}
\min \quad & \sum_{i=1}^{n} \big((1 - \alpha) z_i^- + \alpha z_i^+\big) \\
\text{s.t.} \quad & z_i^+ - z_i^- = y_i - q_\alpha && \forall i \in \{1, \ldots, n\} \\
& z_i^+ \geq 0 && \forall i \in \{1, \ldots, n\} \\
& z_i^- \geq 0 && \forall i \in \{1, \ldots, n\} \\
& q_\alpha \in \mathbb{R}.
\end{aligned}$$

Exercise 6

Equipped with the solution of Exercise 5, we just have to replace q_α with a different $q_{\alpha,i}$ for each i; $q_{\alpha,i}$ is defined in the text of the exercise.

$$\begin{aligned}
\min \quad & \sum_{i=1}^{n} \big((1-\alpha) z_i^- + \alpha z_i^+\big) && \\
\text{s.t.} \quad & z_i^+ - z_i^- = y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j && \forall i \in \{1,\dots,n\} \\
& z_i^+ \geq 0 && \forall i \in \{1,\dots,n\} \\
& z_i^- \geq 0 && \forall i \in \{1,\dots,n\} \\
& \beta_j \in \mathbb{R} && \forall j \in \{0,\dots,p\}.
\end{aligned}$$

Exercise 7

Assume, by contradiction, that no optimal solution is a vertex. Let x^* be an optimal solution. Because x^* is feasible, it can be written as a convex combination of the vertices of the polytope defining the feasible region. Let $x_1,\dots,x_p$ be such vertices and $\lambda_1,\dots,\lambda_p$ the coefficients of the convex combination, i.e.,

$$x^* = \sum_{i=1}^{p} \lambda_i x_i, \quad \lambda_i \in [0,1]\ \forall i \in \{1,\dots,p\}, \quad \sum_{i=1}^{p} \lambda_i = 1. \tag{32}$$

By hypothesis, $c^\top x_i < c^\top x^*$, i.e., none of the vertices is optimal. But then

$$c^\top x^* = c^\top \left(\sum_{i=1}^{p} \lambda_i x_i \right) = \sum_{i=1}^{p} \lambda_i c^\top x_i < \sum_{i=1}^{p} \lambda_i c^\top x^* = (c^\top x^*) \sum_{i=1}^{p} \lambda_i = c^\top x^*, \tag{33}$$

which is a contradiction.