Chapter 4
Strong Duality

Via the termination argument by perturbation, we we can now prove the duality theorem. We are given a linear program

$$\max\{c^T x\colon\, x \in \mathbb{R}^n,\, Ax \leqslant b\}, \tag{4.1}$$

called the *primal* and its *dual*

$$\min\{b^T y\colon\, y \in \mathbb{R}^m,\, A^T y = c,\, y \geqslant 0\}. \tag{4.2}$$

We again formulate the theorem of weak duality in this setting.

Theorem 4.1 (Weak duality). *If x^* and y^* are primal and dual feasible respectively, then $c^T x^* \leqslant b^T y^*$.*

Proof. We have $c^T x^* = {y^*}^T A x^* \leqslant {y^*}^T b$. □

The strong duality theorem tells us that if there exist feasible primal and dual solutions, then there exist feasible primal and dual solutions which have the same objective value. We can prove it with the simplex algorithm.

Theorem 4.2. *If the primal linear program is feasible and bounded, then so is the dual linear program. Furthermore in this case, both linear programs have an optimal solution and the optimal values coincide.*

Proof. Suppose first that A has full column rank. The simplex method finds a roof B of (4.1) with x_B^* being an optimal feasible solution. At the same time, the roof is a conic combination of the normal-vectors in the roof. This means that there exists a $y^* \in \mathbb{R}^m_{\geqslant 0}$ with $y^*(i) = 0$ for all $i \notin B$ such that ${y^*}^T A = c$. But $b^T y^* = \sum_{i \in B} b(i) y^*(i) = \sum_{i \in B} a_i^T x_B^* y^*(i) = (\sum_{i \in B} y^*(i) a_i)^T x_B^* = c^T x_B^*$. Thus y^* is an optimal solution of the dual and the objective function values coincide.

Suppose now that A does not have full column rank and we can write $A = [A_1 \mid A_2]$ where A_1 has full column rank and $A_2 = A_1 \cdot U$ with some matrix U as in Section 3.4. Again, as in Section 3.4 we define the linear program

$$\max\{c_1^T x_1 \colon x_1 \in \mathbb{R}^k,\ A_1 x_1 \leqslant b\}. \tag{4.3}$$

This linear program has an optimal solution x_1^* and $c_2^T = c_1^T \cdot U$. By what we have proved above, there exists a $y^* \in \mathbb{R}^m_{\geqslant 0}$ such that $b^T y^* = c_1^T x_1^*$ and $y^{*T} A_1 = c_1^T$ and since $c_2^T = c_1^T \cdot U$ and $A_2 = A_1 \cdot U$ we have $y^{*T}[A_1 \mid A_2] = [c_1^T \mid c_2^T]$. □

We can formulate dual linear programs also if the linear program is not in inequaliy standard form. The procedure above can be described as follows. We transform a linear program into a linear program in inequality standard form and construct its dual linear program. This dual is then transformed into an equivalent linear program again which is conveniently described.

Let us perform such operations on the dual linear program

$$\min\{b^T y \colon y \in \mathbb{R}^m,\ A^T y = c,\ y \geqslant 0\}$$

of the primal $\max\{c^T x \colon x \in \mathbb{R}^n,\ Ax \leqslant b\}$. We transform it into inequality standard form

$$\begin{aligned} \max -b^T y & \\ A^T y &\leqslant c \\ -A^T y &\leqslant -c \\ -I y &\leqslant 0. \end{aligned}$$

The dual linear program of this is

$$\begin{aligned} \min c^T x_1 - c^T x_2 & \\ A x_1 - A x_2 - x_3 &= -b \\ x_1, x_2, x_3 &\geqslant 0 \end{aligned}$$

This is equivalent to

$$\begin{aligned} \max c^T (x_2 - x_1) & \\ A(x_2 - x_1) + x_3 &= b \\ x_1, x_2, x_3 &\geqslant 0 \end{aligned}$$

which is equivalent to the primal linear program

$$\begin{aligned} \max c^T x & \\ Ax &\leqslant b. \end{aligned}$$

Loosely formulated one could say that "The dual of the dual is the primal". But this, of course, is not to be understood as a mathematical statement. In any case we can state the following corollary.

Corollary 4.1. *If the dual linear program has an optimal solution, then so does the primal linear program and the objective values coincide.*

We present another example of duality that we will need later on. Consider a linear program

$$\begin{aligned} \max\ & c^T x \\ Bx &= b \\ Cx &\leqslant d. \end{aligned} \tag{4.4}$$

After re-formulation, we obtain

$$\begin{aligned} \max\ & c^T x \\ Bx &\leqslant b \\ -Bx &\leqslant -b \\ Cx &\leqslant d \end{aligned}$$

We can form the dual of the latter problem and obtain

$$\begin{aligned} \min\ & b^T y_1 - b^T y_2 + d^T y_3 \\ B^T y_1 - B^T y_2 + C^T y_3 &= c \\ y_1, y_2, y_3 &\geqslant 0. \end{aligned}$$

But this linear program is equivalent to the linear program

$$\begin{aligned} \min\ & b^T y_1 + d^T y_2 \\ B^T y_1 + C^T y_2 &= c \\ y_2 &\geqslant 0. \end{aligned} \tag{4.5}$$

This justifies to say that (4.5) is the dual of (4.4).

4.1 Zero sum games

Consider the following two-player game defined by a matrix $A \in \mathbb{R}^{m\times n}$. The *row-player* chooses a row $i \in \{1,\dots,m\}$ and the column-player chooses a column $j \in \{1,\dots,n\}$. Both players make this choice at the same time. The *payoff* for the row-player is then the matrix-element $A(i,j)$ whereas $A(i,j)$ also determines the *loss* of the column player. In other words, the column player pays $A(i,j)$ to the row-player. If this number is negative, then the row-player actually pays the absolute value of $A(i,j)$ to the column player.

Consider for example the matrix

$$A = \begin{pmatrix} 5 & 1 & 3 \\ 3 & 2 & 4 \\ -3 & 0 & 1 \end{pmatrix}. \tag{4.6}$$

If the row-player chooses the second row and the column player chooses the second-column, then the payoff for the row-player is 2, whereas this is the loss of the column player.

The row-player is now interested in finding a strategy that maximizes his *guaranteed* payoff. For example, if he chooses row 1, then the best choice of the column player would be column 2, since the second element of the first row is the smallest element of that row. Thus the strategy that maximizes the minimal possible payoff would be to choose row 2. In other words

$$\max_i \min_j A(i,j) = 2.$$

What would be the column-player's best hedging strategy? He wants to choose a column such that the largest element in this column is minimized. This column would be the second one. In other words

$$\min_j \max_i A(i,j) = 2.$$

Is it always the case that $\max_i \min_j A(i,j) = \min_j \max_i A(i,j)$? The next example shows that the answer is no:

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{4.7}$$

Here we have $\max_i \min_j A(i,j) = -1$ and $\min_j \max_i A(i,j) = 1$. This can be interpreted as follows. If the column player knows beforehand, the row to be chosen by the row-player, then he would choose a column that results in a gain for him. Similarly, if the row-player knows beforehand the column to be chosen by the column-player, then he can guarantee him a gain of one.

The idea is thus not to stick with a *pure* strategy, but to play with a *random* or *mixed* strategy. If the row-player chooses each of the two rows above uniformly at random, then his expected payoff is zero. Similarly, if the column player chooses each of his two columns with probability 1/2, then his expected payoff is zero as well.

Definition 4.1 (Mixed strategy). Let $A \in \mathbb{R}^{m \times n}$ define a two-player matrix game. A mixed strategy for the row-player is a vector $x \in \mathbb{R}^m_{\geqslant 0}$ with $\sum_{i=1}^m x(i) = 1$. A mixed strategy for the column player is a vector $y \in \mathbb{R}^n_{\geqslant 0}$ with $\sum_{j=1}^n y(i) = 1$.

Such mixed strategies define a probability distribution on the row and column indices respectively. If the row-player and column-player choose a row and column according to this distribution respectively, then the *expected payoff* for the row-player is

$$E[\text{Payoff}] = x^T A y. \tag{4.8}$$

For the game defined by (4.7) and $x^T = (1/2, 1/2)$ and $y^T = (1/2, 1/2)$ the expected payoff is 0.

Lemma 4.1. *Let $A \in \mathbb{R}^{m \times n}$, then*

$$\max_{x \in X} \min_{y \in Y} x^T A y \leqslant \min_{y \in Y} \max_{x \in X} x^T A y,$$

where X and Y denote the set of mixed row and column-strategies respectively.

Proof. Let x' and y' be some fixed mixed strategies of the row and column-player respectively. Clearly

$$\min_y x'^T Ay \leqslant x'^T Ay' \leqslant \max_x x^T Ay',$$

which implies the assertion. □

The next theorem is one of the best-known results in the field of *game theory*. It states that there are mixed strategies x' and y' from above such that equality holds. It is proved with the theorem of strong duality.

Theorem 4.3 (Minimax-Theorem).

$$\max_{x\in X}\min_{y\in Y} x^T Ay = \min_{y\in Y}\max_{x\in X} x^T Ay,$$

where X and Y denote the set of mixed row and column-strategies respectively.

Proof. Let us inspect the value $\max_{x\in X}\min_{y\in Y} x^T Ay$. This can be understood as to maximize the function

$$f(x) = \min\{(x^T A)\cdot y\colon \sum_{j=1}^{n} y_j = 1,\ y \geqslant 0\}.$$

Thus the value $f(x)$ is the optimal solution of a bounded and feasible linear program. The dual of this linear program (for fixed x) has only one variable x_0 and reads

$$\max\{x_0\colon x_0 \in \mathbb{R},\ \mathbb{1}x_0 \leqslant A^T x\}.$$

But this shows that the maximum value of $f(x)$, where x ranges over all mixed row-strategies is the linear program

$$\begin{aligned} \max x_0 & \\ \mathbb{1}x_0 - A^T x &\leqslant 0 \\ \textstyle\sum_{i=1}^{m} x_i &= 1 \\ x &\geqslant 0. \end{aligned} \tag{4.9}$$

Let us now inspect the value $\min_{y\in Y}\max_{x\in X} x^T Ay$. Again, by applying duality this can be computed with the linear program

$$\begin{aligned} \min y_0 & \\ \mathbb{1}y_0 - Ay &\leqslant 0 \\ \textstyle\sum_{j=1}^{n} y_j &= 1 \\ y &\geqslant 0. \end{aligned} \tag{4.10}$$

It follows from the duality of (4.5) and (4.4) that the linear programs (4.9) and (4.10) are duals of each other. This proves the Minimax-Theorem. □

4.2 The classical simplex algorithm

I have chosen to explain the simplex algorithm as a method to obtain primal feasibility of the linear program

$$\max\{c^T x\colon x \in \mathbb{R}^n,\ Ax \leqslant b\} \tag{4.11}$$

via improving the roofs. By duality, our description can be re-interpreted as solving the linear program

$$\min\{b^T y\colon y \in \mathbb{R}^m,\ A^T y = c,\ y \geqslant 0\} \tag{4.12}$$

where $A^T \in \mathbb{R}^{n\times m}$ has now full row-rank. Recall that a roof is set of row-indices B of A such that the corresponding rows are a basis of $\mathbb{R}^n$ and c is a conic combination of these rows. The translation of a roof to the classical setting (4.12) is a set of n linearly independent columns such that c is a conic combination of these columns. The corresponding *basic feasible solution* is the vector x^*, where $x^*_B = A_B^{T^{-1}} c$ and $x^*_{\overline{B}} = 0$. The simplex method can now be interpreted as a method that improves the basic-feasible solution by letting a new index enter the basis while another index leaves the basis. More on this classical simplex versus the roof-interpretation can be found in exercise 4

4.3 A proof of the duality theorem via Farkas' lemma

Remember Farkas' lemma (Theorem 2.9) which states that $Ax = b, x \geqslant 0$ has a solution if and only if for all $\lambda \in \mathbb{R}^m$ with $\lambda^T A \geqslant 0$ one also has $\lambda^T b \geqslant 0$. In fact the duality theorem follows from this. First, we derive another variant of Farkas' lemma.

Theorem 4.4 (Second variant of Farkas' lemma). *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The system $Ax \leqslant b$ has a solution if and only if for all $\lambda \geqslant 0$ with $\lambda^T A = 0$ one has $\lambda^T b \geqslant 0$.*

Proof. Necessity is clear: If x^* is a feasible solution, $\lambda \geqslant 0$ and $\lambda^T A = 0$, then $\lambda^T A x^* \leqslant \lambda^T b$ implies $0 \leqslant \lambda^T b$.

On the other hand, $Ax \leqslant b$ has a solution if and only if

$$Ax^+ - Ax^- + z = b,\ x^+, x^-, z \geqslant 0 \tag{4.13}$$

has a solution. So, if $Ax \leqslant b$ does not have a solution, then also (4.13) does not have a solution. By Farkas' lemma, there exists a $\lambda \in \mathbb{R}^m$ with $\lambda^T [A \mid -A \mid I_m] \geqslant 0$ and $\lambda^T b < 0$. For this λ one also has $\lambda^T A = 0$ and $\lambda \geqslant 0$. □

We are now ready to prove the theorem of strong duality via the second variant of Farkas' lemma.

Proof (of strong duality via Farkas' lemma). Let δ be the objective function value of an optimal solution of the dual $\max\{b^T y \colon y \in \mathbb{R}^m, A^T y \leqslant c\}$. For all $\varepsilon > 0$, the system $A^T y \leqslant c, -b^T y \leqslant -\delta - \varepsilon$ does not have a solution. By the second variant of Farkas' lemma, there exists a $\lambda \geqslant 0$ with $\lambda^T \binom{-b^T}{A^T} = 0$ and $\lambda^T \binom{-\delta-\varepsilon}{c} < 0$. Write λ as $\lambda = \binom{\lambda_1}{\lambda'}$ with $\lambda' \in \mathbb{R}^n$. If λ_1 were zero, we could apply the second variant of Farkas' lemma to the system $A^T y \leqslant c$ and λ', since we know that $A^T y \leqslant c$ has a solution. Therefore, we can conclude $\lambda_1 > 0$. Furthermore, by scaling, we can assume $\lambda_1 = 1$. One has $\lambda'^T A^T = b^T$ and $\lambda'^T c < \delta + \varepsilon$. The first equation implies that λ' is a feasible solution of the primal (recall $\lambda' \geqslant 0$). The second equation shows that the objective function value of λ' is less than $\delta + \varepsilon$. This means that the optimum value of the primal linear program is also δ, since the primal has an optimal solution and ε can be chosen arbitrarily small. $\square$

Exercises

1. Formulate the dual linear program of

$$\begin{aligned} \max 2x_1 + 3x_2 - 7x_3 & \\ x_1 + 3x_2 + 2x_3 &= 4 \\ x_1 + x_2 &\leqslant 8 \\ x_1 - x_3 &\geqslant -15 \\ x_1, x_2 &\geqslant 0 \end{aligned}$$

2. Consider the following linear program

$$\begin{aligned} \max x_1 + x_2 & \\ 2x_1 + x_2 &\leqslant 6 \\ x_1 + 2x_2 &\leqslant 8 \\ 3x_1 + 4x_2 &\leqslant 22 \\ x_1 + 5x_2 &\leqslant 23 \end{aligned}$$

   Show that $(4/3, 10/3)$ is an optimal solution by providing a suitable feasible dual solution.
3. Show that for $A \in \mathbb{R}^{m \times n}$, one has

$$\max_i \min_j A(i,j) \leqslant \min_j \max_i A(i,j).$$

4. In the lecture you have seen the simplex algorithm for linear programs of the form

$$\max\{c^T x \,:\, Ax \leqslant b\}.$$

   We will now derive a simplex algorithm for linear programs of the form

$$\min\{c^T x \;:\; Ax = b,\; x \geqslant 0\} \tag{4.14}$$

with $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Throughout the exercise we assume that (4.14) is feasible and bounded, and that A has full row rank.
For $i \in \{1,\dots,n\}$ we define A_i as the i-th column of A. Moreover, for some subset $B \subseteq \{1,\dots,n\}$, A_B is the matrix A restricted to the columns corresponding to elements of B.
A subset $B \subseteq \{1,\dots,n\}$ with $|B| = m$ such that A_B has full rank is called a *basis*. The vector $x \in \mathbb{R}^n$ defined as $x_i := 0$ for all $i \notin B$ and $x_B := A_B^{-1} b$ is called the *basic solution* associated to B. Note that x is a feasible solution to (4.14) if and only if $x \geqslant 0$.
Given a basis B and let $j \in \{1,\dots,n\}$, $j \notin B$. The vector $d \in \mathbb{R}^n$ defined as $d_j = 1$, $d_i = 0$ for all $i \notin B$ and $d_B := -A_B^{-1} A_j$ is called the *j-th basic direction*.
Assume that the solution x associated to B is feasible. Moreover assume that $x_B > 0$.

a. Show that there is a $\theta > 0$ such that $x + \theta d$ is a feasible solution. Give a formula to compute the largest θ such that $x + \theta d$ is feasible.
b. Let θ^* be maximal. Show that there is a basis B' such that $x + \theta^* d$ is the basic solution associated to B'.
c. Let $x' = x + \theta d$. Show that the objective value of x' changes by $\theta\left(c_j - c_B^T A_B^{-1} A_j\right)$.
d. Consider a basis B with basic feasible solution x. Show that if $c - c_B^T A_B^{-1} A \geqslant 0$, then x is an optimal solution to (4.14).

This suggests the following algorithm: Start with some basis B whose associated basic solution is feasible. Compute $\bar{c} := c - c_B^T A_B^{-1} A$. If $\bar{c} \geqslant 0$, we have an optimal solution (see 4d). Otherwise, let j be such that $\bar{c}_j < 0$. Part 4b and 4c show that if we change the basis, we find a feasible solution with an improved objective value. We repeat these steps until the vector $\bar{c}$ is nonnegative.
This is the way the simplex algorithm usually is introduced in the literature. This algorithm is exactly the same as the one you learned in the lecture. To get an intuition why this is true, show the following:

a. Given a basis B, show that its associated basic solution is feasible if and only if B is a *roof* of the LP dual to (4.14).
b. Consider a basis B and its associated feasible basic solution x. As seen before, B is also a roof in the dual LP. Let y be the vertex of that roof. Show that for any $j \in \{1,\dots,n\}$ we have $\bar{c}_j < 0$ if and only if $A_j^T y > c_j$.

References

1. J. Edmonds. Maximum matching and a polyhedron with 0,1-vertices. *Journal of Research of the National Bureau of Standards*, 69:125–130, 1965.
2. J. Edmonds. Paths, trees and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.
3. F. Eisenbrand, A. Karrenbauer, and C. Xu. Algorithms for longer oled lifetime. In *6th International Workshop on Experimental Algorithms, (WEA 07)*, pages 338–351, 2007.
4. M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer, 1988.
5. L. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademii Nauk SSSR*, 244:1093–1097, 1979.
6. T. Koch. *Rapid Mathematical Programming*. PhD thesis, Technische Universität Berlin, 2004. ZIB-Report 04-58.
7. L. Lovász. Graph theory and integer programming. *Annals of Discrete Mathematics*, 4:141–158, 1979.
8. J. E. Marsden and M. J. Hoffman. *Elementary Classical Analysis*. Freeman, 2 edition, 1993.
9. J. Matouek and B. Gärtner. *Understanding and Using Linear Programming (Universitext)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
10. A. S. Nemirovskiy and D. B. Yudin. Informational complexity of mathematical programming. *Izvestiya Akademii Nauk SSSR. Tekhnicheskaya Kibernetika*, (1):88–117, 1983.
11. A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley, 1986.
12. N. Z. Shor. Cut-off method with space extension in convex programming problems. *Cybernetics and systems analysis*, 13(1):94–96, 1977.