WEEK 3

Duality, Cones and Representation of polyhedra

1 Duality

A strong property of Linear Programming is the existence of a *min-max relation*: the minimum of some problem is equal to the maximum of some other problem. Existence of a min-max relation for an optimization problem preludes efficient solvability of that problem. For those who know complexity theory a bit more: problems in NP that have a min-max relation belong also to the class *co-NP*. Certainly P⊂NP∩co-NP and it could be that P=NP∩co-NP also if P≠NP.

Almost surely any of you who had a basic course in LP has been instructed on duality based on the Simplex method. Some of you may not have seen the proof of strong duality. I will just briefly sketch this view on duality, leaving the detailed reading to yourself. More time I will spend on deriving strong duality from geometric theorems.

1.1 Duality based on the Simplex method

In the book LP-duality is derived as a special case of Lagrange-relaxation. I will take the other approach for deriving lower bounds.

Let us take as the *primal* LP problem the standard problem

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array} \qquad (1)$$

Or written out:

$$\begin{array}{ll} \min & \sum_{j=1}^n c_j x_j \\ \text{s.t.} & \sum_{j=1}^n a_{ij} x_j = b_i, \ i = 1, \ldots, m \\ & x_j \geq 0, \ j = 1, \ldots, n. \end{array} \qquad (2)$$

For each restriction $\sum_{j=1}^n a_{ij} x_j = b_i$ introduce dual variable p_i, $i = 1, \ldots, m$. Clearly,

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

Any feasible solution of (1) satisfies this equality. If we select values for p_i, $i = 1, \ldots, m$ such that

$$\sum_{i=1}^m p_i a_{ij} \leq c_j, \ \forall \ j = 1 \ldots, n,$$

then, since $x_j \geq 0$, $j = 1, \ldots, n$,

$$\sum_{i=1}^{m} p_i b_i = \sum_{i=1}^{m} p_i \sum_{j=1}^{n} a_{ij} x_j = \sum_{j=1}^{n} \sum_{i=1}^{m} p_i a_{ij} x_j \leq \sum_{j=1}^{n} c_j x_j$$

and hence we have a lower bound on the objective value of any feasible solution for the primal problem. The best lower bound is given by the outcome of the LP problem

$$\begin{array}{rl} \max & \sum_{i=1}^{m} p_i b_i \\ \text{s.t.} & \sum_{i=1}^{m} p_i a_{ij} \leq c_j,\ j = 1 \ldots, n, \end{array} \tag{3}$$

which is the *dual* LP-problem. Rewriting the restrictions and bringing it in matrix notation the dual is

$$\begin{array}{rl} \max & p^T b \\ \text{s.t.} & c^T - p^T A \geq \underline{0}^T. \end{array} \tag{4}$$

I refer to the book to read how dualisation of all the different forms of inequalities and sign-constraints work.

On the way we have proven *weak duality*.

Theorem 1.1 (4.3 in [B&T]), Weak duality. *For any dual pair of feasible LP problems, where the primal problem is minimization, the objective value of any feasible solution for the dual problem is a lower bound on the objective value of any feasible solution for the primal problem.*

But a stronger result holds.

Theorem 1.2 (4.4 in [B&T]), Strong duality. *If a linear programming problem has an optimal solution then so does its dual and their objective values are equal.*

Proof. Starting from the dual pair (1) and (4). Let x be an optimal bfs to (1) and B the corresponding basis matrix. Thus the optimal solution value is $c_B^T x_B = c_B^T B^{-1} b$. Simplex terminates if $c^T - c_B^T B^{-1} A \geq \underline{0}^T$. This implies that the choice $p^T = c_B^T B^{-1}$ satisfies the constraints of (4). It has objective value $p^T b = c_B^T B^{-1} b$ equal to the primal optimal solution value, which together with weak duality proves the theorem. □

Another relation between primal and dual problem is the so-called complementary slack relations. I give them without the easy proof.

Theorem 1.3 *Let x and p be feasible solutions to the primal and the dual problem, respectively. Then x and p are optimal if and only if*

$$(c^T - p^T A)x = 0.$$

□

1.2 Duality based on geometry

The following is a basic theorem in mathematics, the proof of which you may know or not, but will not learn here.

Theorem 1.4 (4.10 in [B&T]), Weierstrass' theorem. *If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and $S \subset \mathbb{R}^n$ non-empty and compact, then f attains both a maximum and a minimum value on S.*

In a way the following theorem is the basic theorem from which the whole LP-duality theory can be derived.

Theorem 1.5 (4.11 in [B&T]), Separating Hyperplane theorem. *Let $S \subset \mathbb{R}^n$ be non-empty, closed and convex and $x^* \in \mathbb{R}^n \setminus S$, then there exist $c \in \mathbb{R}^n$ such that $c^T x^* < c^T x$, $\forall\ x \in S$.*

PROOF. Let y be the point in S nearest to x^* in Euclidean distance. Weierstrass' theorem can be used to show that y exists (the distance to x^* is a continuous function and we can choose a closed ball around x^* containing at least one point in S to have a compact subset of S).

We show that $c = y - x^*$ proves the theorem. Take any $x \in S$ and any $\lambda \in (0, 1]$, then $y + \lambda(x - y) \in S$, by convexity and

$$\|y - x^*\|^2 \leq \|y + \lambda(x - y) - x^*\|^2 = \|y - x^*\|^2 + 2\lambda(y - x^*)^T(x - y) + \lambda^2\|x - y\|^2,$$

yielding

$$2\lambda(y - x^*)^T(x - y) + \lambda^2\|x - y\|^2 \geq 0$$

Dividing by λ and taking $\lim_{\lambda \downarrow 0}$ gives

$$(y - x^*)^T(x - y) \geq 0.$$

Hence,

$$(y - x^*)^T x \geq (y - x^*)^T y = (y - x^*)^T x^* + (y - x^*)^T(y - x^*) > (y - x^*)^T x^*.$$

$\square$

A fundamental theorem in the theory of systems of linear equalities and inequalities is Farkas' Lemma, which is stated below in the form corresponding to the standard LP formulation.

Theorem 1.6 (4.6 in [B&T]) *Given $m \times n$ matrix A and $b \in \mathbb{R}^m$, exactly one of the following two alternatives holds:*
(i) $\exists\ x \geq 0 :\ Ax = b$;
(ii) $\exists\ p \in \mathbb{R}^m :\ p^T A \geq 0 \wedge p^T b < 0$.

PROOF. (ii) $\Rightarrow \neg$(i)). $x \geq 0 \;\Rightarrow p^T Ax \geq 0$, whereas $p^T b < 0$.

$\neg$(i) $\Rightarrow$ (ii). The set $S =: \{y \mid \exists x \geq 0 : y = Ax\}$ is convex (easily verified from the definition of convexity), non-empty (e.g. $\underline{0} \in S$) and closed (not so obvious to prove since we skipped Section 2.8). $\neg$(i) $\Rightarrow b \notin S$. Hence, by the Separating Hyperplane theorem $\exists p$ such that $p^T b < p^T y$, $\forall y \in S$. Thus, $\underline{0} \in S \Rightarrow p^T b < 0$. For every $i = 1, \ldots, n$, $\lambda A_i \in S$, $\forall \lambda \geq 0$. Hence if $p^T A_i < 0$ then $p^T(\lambda A_i) = \lambda p^T A_i < p^T b$ for λ sufficiently large, contradicting $p^T y > p^T b$, $\forall y \in S$. □

The following corollary of Farkas' Lemma states $\neg$(ii) $\Rightarrow$ (i) in words.

Corollary 1.1 (4.3 in [B&T]) *If $p^T b \geq 0$ for all $p \in \mathbb{R}^m$ with $p^T A_i \geq 0$, $i = 1, \ldots, n$, then b is a non-negative linear combination of $A_1, \ldots A_n$.*

There are many equivalent forms of Farkas' Lemma, which can all be proved by manipulating Farkas' Lemma in an appropriate way. For instance:

Theorem 1.7 *Given $m \times n$ matrix A and $b \in \mathbb{R}^m$, then $\exists\, x \geq 0 :\; Ax \leq b$ if and only if $p^T b \geq 0$ for all $p \geq 0$ with $p^T A \geq 0$.*

Which is equivalent to

Theorem 1.8 *Given $m \times n$ matrix A and $b \in \mathbb{R}^m$, exactly one of the following two alternatives holds:*
(i) $\exists\, x \geq 0 :\; Ax \leq b$;
(ii) $\exists\, p \geq 0 :\; p^T A \geq 0 \wedge p^T b < 0$.

Strong duality follows form Farkas' Lemma.

Theorem 1.9 (4.4 in [B&T]), Strong duality. $\min\{c^T x \mid Ax = b, x \geq 0\} = \max\{p^T b \mid p^T A \leq c^T\}$ *if both exist.*

PROOF. Let $\hat{p}$ be an optimal bfs of the dual problem. Let $J = \{j \mid \hat{p}^T A_j = c_j\}$. Take any vector d such that $d^T A_j \leq 0$, (i.e. $-d^T A_j \geq 0$), $\forall\, j \in J$. Then

$$(\hat{p} + \epsilon d)^T A_j \leq \hat{p}^T A_j = c_j,\; \forall\, j \in J,\; \forall \epsilon > 0$$

and since $\hat{p}^T A_j < c_j$, $\forall\, j \notin J$,

$$(\hat{p} + \epsilon d)^T A_j \leq c_j,\; \forall\, j \notin J,\; \forall \epsilon \text{ small enough},$$

Thus, d is a direction leading to other feasible points and since $\hat{p}$ is an optimal dual solution, it must be that $d^T b \leq 0$ (i.e., $-d^T b \geq 0$).

So, in conclusion we have that for every $d \in \mathbb{R}^m$ such that $-d^T A_j \geq 0\; \forall j \in J$ it holds that $-d^T b \geq 0$. Thus we are in the $\neg$(ii) case of Farkas lemma restricted to the matrix A_J made up by the columns A_j, $j \in J$.

Hence the corollary of Farkas' Lemma applies, saying that b is a non-negative linear combination of $\{A_j\}_{j \in J}$; i.e., there exists $x_j \geq 0,\ j \in J$ such that $\sum_{j \in J} A_j x_j = b$. Setting $x_j = 0,\ \forall j \notin J$ we obtain $x \geq 0$ and $Ax = b$. Moreover,

$$c^T x = \sum_{j \in J} c_j x_j = \sum_{j \in J} \hat{p}^T A_j x_j = \hat{p}^T b.$$

□

2 Cones and Representation of polyhedra

Definition 2.1 A cone $C \subset \mathbb{R}^n$ is a set with the property $\forall x \in C\ \forall \lambda > 0\ :\ \lambda x \in C$.

A polyhedral cone is generated by a finite set of linear halfspaces

Definition 2.2 A polyhedral cone is a set $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ for some matrix A.

Definition 2.3 The *recession cone* (or also called *characteristic cone*) of a polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$, is the set $C(P) = \{d \in \mathbb{R}^n \mid Ad \geq 0\}$.

It is the set of vectors d with the property that $\forall y \in P\ \forall \lambda > 0$: $y + \lambda d \in P$. For the standard LP-polyhedron $P = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$, the recession cone is easily seen to be $C(P) = \{d \in \mathbb{R}^n \mid Ad = 0, d \geq 0\}$.

Definition 2.4 An extreme ray of an n-dimensional cone is the intersection of $n - 1$ linearly independent active constraints.

We speak about an extreme ray of a polyhedron as an extreme ray of its recession cone.

In an LP $\min\{c^T x \mid Ax \geq b\}$, it is clear that if an extreme ray d of the feasible polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ has negative inner product $c^T d < 0$ then starting in any point $x \in P$ we can make the value of $c^T(x + \lambda d)$ arbitrarily small by choosing λ arbitrarily large. More surprising and indeed somewhat more difficult to prove is that the reverse is also true.

In proving the following theorems we often make use of the following result, stated as a corollary of the weak duality theorem in the book, where I added the last statement, which is in the book verified by example:

Corollary 2.1 (4.1 in [B&T]).
a) If the primal problem is unbounded, i.e., its optimal cost is $-\infty$, then the dual problem must be infeasible.

b) If the dual problem is unbounded, i.e., its optimal cost is $+\infty$, then the primal problem must be infeasible.
c) If the primal problem is infeasible, then its dual is either unbounded or infeasible.

Theorem 2.1 (4.14 in [B&T]) *Assume $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ has at least one extreme point, then the LP $\min\{c^T x \mid x \in P\}$ has optimal unbounded objective value ($-\infty$) if and only if there exists an extreme ray d of P such that $c^T d < 0$.*

Proof. One direction we just proved. For the other direction I just sketch the proof. If the primal LP is unbounded then by Corollary 2.1a, the dual is infeasible. I.e.,

$$\{p \in \mathbb{R}^m \mid p^T A = c^T, p \geq 0\} = \emptyset.$$

This implies also infeasibility of the LP

$$\max\{p^T 0 \mid p^T A = c^T, p \geq 0\}.$$

Hence, by Corollary 2.1c its dual

$$\min\{c^T x \mid Ax \geq 0\}$$

is unbounded or infeasible. Since $x = 0$ is a feasible solution it must be unbounded. This implies that there exists an x in the cone $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ such that $c^T x = -1$. Now use that P has an extreme point, implying that $\text{rank}(A) = n$, and hence that the polyhedron $\{x \in \mathbb{R}^n \mid Ax \geq 0, c^T x = -1\}$ has at least one extreme point. In this point $n-1$ linearly independent inequalities $a_i^T x \geq 0$ must be active, hence the point corresponds to a ray of C, hence to a ray of P. □

Theorem 2.2 (4.18 in [B&T]), Almost Minkowski's resolution theorem. *Let $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$, non-empty, with $x^1, \ldots, x^k$, for some $k \geq 1$, its set of extreme points, and $w^1, \ldots, w^r$ its set of extreme rays. Then*

$$P = Q =: \left\{ \sum_{i=1}^{k} \lambda_i x^i + \sum_{j=1}^{r} \theta_j w^j \mid \lambda_i \geq 0\ \forall i, \theta_j \geq 0\ \forall j, \sum_{i=1}^{k} \lambda_i = 1 \right\} \tag{5}$$

Notice that in this theorem we require P to have at least one extreme point, implying that the recession cone of P is pointed, i.e., having 1 extreme point (the point $\underline{0}$).

Proof. $Q \subset P$ is easily shown and indeed does not require the assumption that $k \geq 1$: $y = \sum_{i=1}^{k} \lambda_i x^i \in P$ and for $z = \sum_{j=1}^{r} \theta_j w^j$, $Az \geq 0$, hence $y + z \in P$.

$P \subset Q$ is proved by contradiction. Let $v \in P \setminus Q$, then we try to construct a hyperplane that separates v from Q and we show that this is impossible. Consider the LP

$$\max \left\{ \underline{0}^T \lambda + \underline{0}^T \theta \mid \sum_{i=1}^{k} \lambda_i x^i + \sum_{j=1}^{r} \theta_j w^j = v,\ \lambda^T \underline{1} = 1,\ \lambda \geq 0,\ \theta \geq 0 \right\} \tag{6}$$

If $v \notin Q$ then this LP is infeasible. By Corollary 2.1c this implies that the dual

$$\min \left\{ p^T v + q \mid p^T x^i + q \geq 0,\ i = 1, \ldots, k,\ p^T w^j \geq 0,\ j = 1, \ldots, r \right\} \tag{7}$$

is either infeasible as well or unbounded. Since, $p = \underline{0}$, $q = 0$ is a solution, the latter must be true. In particular there exists a solution p, q such that $p^T v + q < 0$. Because $p^T x^i + q \geq 0$ we have $p^T v < p^T x^i$, $\forall\ i$. Fix such a p.

Consider the LP

$$\min \left\{ p^T x \mid Ax \geq b \right\} = \min \left\{ p^T x \mid x \in P \right\}$$

If the optimum is finite then some extreme point x^i is optimal and therefore, since $v \in P$, $p^T v \geq p^T x^i$, a contradiction. Otherwise, if the optimum is infinite, then according to Theorem 2.1 there exists an extreme ray w^j of P such that $p^T w^j < 0$, contradicting the feasibility of p in (7). □

As corollaries to both theorems we get

Corollary 2.2 *A non-empty polytope (bounded polyhedron) is the convex hull of its extreme points.*

Corollary 2.3 *A pointed cone is the set of non-negative combinations of its extreme rays.*

The converse to Minkowski's theorem is also true.

Theorem 2.3 (4.16 in [B&T]) *Any set Q defined as*

$$Q = \left\{ \sum_{i=1}^{k} \lambda_i x^i + \sum_{j=1}^{r} \theta_j w^j \mid \lambda_i \geq 0\ \forall i, \theta_j \geq 0\ \forall j, \sum_{i=1}^{k} \lambda_i = 1 \right\}$$

for some vectors $x^1, \ldots, x^k, w^1, \ldots, w^r$ is a polyhedron.

Proof. Take any $v \in Q$. Consider LP (6), which is feasible if and only if the optimum is 0. Hence the dual (7) has a finite optimum. The dual feasible set is a cone. Hence for all (p, q) feasible for (7) we must have $p^T v + q \geq 0$. Clearly, it is sufficient to require this for the finitely many extreme rays of the dual feasible cone. □

This implies the reverse of the previous corollaries.

Corollary 2.4 *The convex hull of a finite set of points is a polytope.*

Corollary 2.5 *The non-negative combination of a finite set of vectors is a pointed cone.*

What follows is Minkowski's theorem, which I have not treated in detail this time. I just gave a picture of what Minkowski's theorem says in case P has no extreme points. I leave the text below for the interested student

By Theorem 2.6 in [B&T], saying a.o. that a polyhedron has an extreme point if and only if it does not contain a line, the condition that P has an extreme point implies that $C(P) \cap -C(P) = \{\underline{0}\}$. $C(P) \cap -C(P)$ is called the *lineality space* of P (lin$P$) and is equal to the space $\text{lin}(P) = \{x \in \mathbb{R}^n \mid Ax = 0\}$, the kernel of the matrix A. Notice that if $y \in P$ and $z \in \text{lin}(P)$ then $y + \mu z \in P$, $\forall \mu \in \mathbb{R}$, *not just* $\mu > 0$. Clearly, lin(P) is a linear subspace of $\mathbb{R}^n$ with dimension t say. Select $z^1, \ldots, z^t \in \text{lin}(P)$ linearly independent. We also notice here that in case P is the standard linear polytope then $\text{lin}(P) = \{x \in \mathbb{R}^n \mid Ax = 0,\ x = \underline{0}\} = \underline{0}$.

This means that any subset $F = \{x \in \mathbb{R}^n \mid A'x = b'\}$, with A' some subset of the rows of A and b' the corresponding subset of the right-hand sides, will have at least dimension t. Such a set is called a face of P. Let $F_1, \ldots, F_k$ be the set of minimal faces (in fact the faces of dimension t). Choose one $x^i \in F_i$ arbitrarily.

Any cone is highly degenerate with only one minimal face, which is its lineality space. We say a cone C is *pointed* if $\text{lin}(C) = \underline{0}$, i.e., if $\text{rank}(A) = n$. The generating faces of a pointed cone are its 1-dimensional extreme rays. As we did already before we call a polyhedron pointed if its recession cone is pointed, and in that case the w-vectors in the last theorem are its 1-dimensional extreme rays. In case lin(P) has dimension t, then the only minimal face of its recession cone $C(P) = \{d \in \mathbb{R}^n \mid Ad \geq 0\}$ has dimension t, and therefore the extreme rays are $t+1$-dimensional objects, and belong to the sets $G_j = \{x \in \mathbb{R}^n \mid a_j^T x \geq 0, A'x = 0\}$, such that $\text{lin}(P) = \{x \in \mathbb{R}^n \mid a_j^T x = 0, A'x = 0\}$ and $\begin{pmatrix} A' \\ a_j \end{pmatrix}$ has rank $n - t$. We call these the *minimal proper faces* of the cone. These sets contain a unique vector up to scalar multiplication. Choose one $w^j \in G_j$.

Then the resolution theorem for general polyhedra is,

Theorem 2.4 Minkowski's resolution theorem. *Let $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$, non-empty, and with $x^1 \in F_1, \ldots, x^k \in F_k$, arbitrarily chosen, $w^1 \in G_1, \ldots, w^r \in G_r$, arbitrarily chosen, and $z^1, \ldots, z^t$, a basis for lin(P).*

Then

$$P = Q = \left\{ \sum_{i=1}^{k} \lambda_i x^i + \sum_{j=1}^{r} \theta_j w^j + \sum_{\ell=1}^{t} \mu_\ell z_\ell \mid \lambda_i \geq 0 \ \forall i, \theta_j \geq 0 \ \forall j, \sum_{i=1}^{k} \lambda_i = 1 \right\}$$

In words: every polyhedron is the (Minkowski- or set-) sum of a polytope, a cone and a lineality space.

The proof is an extension of the previous proof, by adding the μ's to the "infeasible" LP. This would give the extra constraints $p^T z^h = 0$, $h = 1, \ldots, t$. If you like, you could try to prove it by making the exercises 4.46 and 4.47, where the lineality space is not explicitly included in the description, but you will need it in the proof as far as I know.

If you do not like to try yourself, please refer for a complete, and in fact alternative, proof of this theorem and for a clear exposition of faces, minimal faces and facets to Chapters 7 and 8 from the book by Alexander Schrijver, *Theory of Linear and Integer Programming*, Wiley, 1986.

Material of Week 3 from [B& T]

Chapter 4. We skip Sections 4.4, 4.5 and 4.10.

Exercises of Week 3

4.26, 4.31, 4.35, 4.39, 4.40, 4.44b (in view of the general Minkowski Theorem in these notes; you may try to prove Minkowski's Theorem through 4.46 and 4.47).

Next two times

Chapter 7.