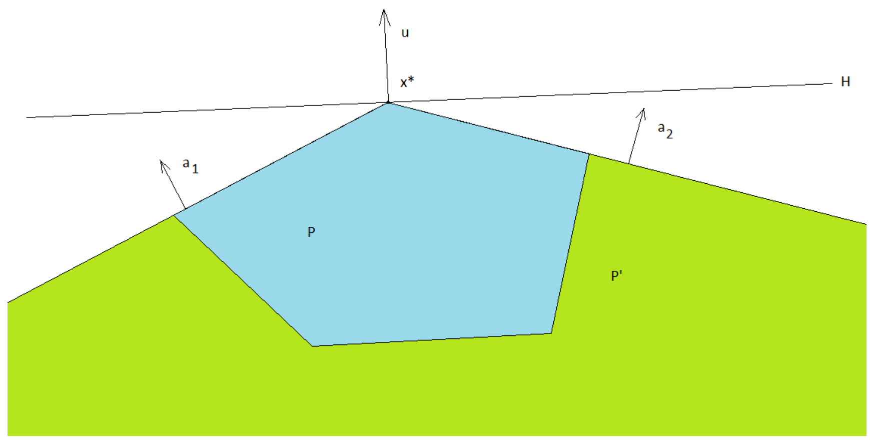

Figure 1: Lemma 7.1: The supporting linear functional u can be represented as a combination of the functionals a_1, a_2 corresponding to the active constraints with nonnegative coefficients.

7 Linear programs

In this lecture we consider solution methods and applications for linear programs (LP). The solution methods can be classified into active set methods and interior-point methods. Active set methods find the optimal solution of the problem in a finite number of steps, but do not have guaranteed polynomial complexity. Interior-point methods (IPM) generate a sequence of iterates which converges to, but never reaches the optimal solution.

Both classes of methods work very well in practice and complement each other, being suited for different kinds of applications. While IPM generally are faster on large problem instances, active set methods are suited for solving multiple linear programs which do not differ much from each other, enabling a warm-start based on the optimal point of the previous LP.

7.1 Theorem on the alternative

Polyhedral sets are particularly simple convex sets. If a hyperplane is supporting to a polyhedral set P, then this relation can be certified in the form of a convex combination of the inequalities defining the polyhedral set.

Lemma 7.1. *Let $P = \{x \,|\, Ax \leq b\}$ be a non-empty polyhedral set, and let $H = \{x \,|\, u^T x = b_0\}$ be a hyperplane containing a point $x^* \in P$ and such that the open half-space $C = \{x \,|\, u^T x > b_0\}$ has an empty intersection with P, i.e., $u^T x^* = b_0$, $u^T x \leq b_0$ for all P. Then there exists a nonnegative vector $\mu \geq 0$ such that $u = A^T \mu$, $b_0 = b^T \mu$.*

Proof. Let I be the set of indices of rows for which the inequality $Ax^* \leq b$ is an equality, i.e., the index set of active constraints at x^*. Then there exists $\epsilon > 0$ such that for all $x \in x^* + B_\epsilon$ we have $(Ax)_j < b_j$ for all $j \notin I$. Define another polyhedral set by $P' = \{x \,|\, (Ax)_i \leq b_i \quad \forall\, i \in I\}$. Then $P \subset P'$. We claim that C has also an empty intersection with P', i.e., $u^T x \leq b_0$ for all $x \in P'$.

Indeed, suppose there exists $z \in P'$ such that $u^T z > u^T x^*$. Then for all $\lambda \in (0, 1]$ we have $z_\lambda = \lambda z + (1 - \lambda)x^* \in P'$ and $u^T z_\lambda > u^T x^*$. But for λ small enough we have $z_\lambda \in x^* + B_\epsilon$, and hence $z_\lambda \in P$, a contradiction.

We now define the polyhedral cone $K = \{A^T \mu \,|\, \mu \geq 0,\ \mu_j = 0 \;\forall\; j \notin I\}$. We claim that $u \in K$.

Indeed, suppose $u \notin K$. Then u can be separated from K, and there exists δ such that $u^T \delta > 0$, $v^T \delta \leq 0$ for all $v \in K$. In particular, $(A\delta)_i \leq 0$ for all $i \in I$. Hence $x^* + \delta \in P'$. But then $u^T(x^* + \delta) \leq u^T x^*$ and $u^T \delta \leq 0$, a contradiction.

Hence there exists $\mu \geq 0$, $\mu_j = 0$ for all $j \notin I$, such that $u = A^T \mu$. It follows that $\mu^T (Ax^* - b) = 0$ and $b_0 = \mu^T A x^* = \mu^T b$, which yields the desired assertion. □

As a consequence, we obtain the following Theorem on the Alternative.

Theorem 7.2. *(Farkas) Let $P = \{x \mid Ax \leq b\}$ be a polyhedral set. Then either $P \neq \emptyset$, or there exists $\mu \geq 0$ such that $\mu^T A = 0$, $\mu^T b = -1$.*

Proof. Clearly if $P \neq \emptyset$, then such a μ cannot exist.

Let $P = \emptyset$. Then the non-empty polyhedral set

$$P' = \{(x,t) \mid Ax - bt = (A, -b)(x^T, t)^T \leq 0\}$$

has an empty intersection with the open half-space $C = \{(x,t) \mid t = (0,1)(x^T,t)^T > 0\}$. By the Lemma 7.1 there exists $\mu \geq 0$ such that $(0,1) = \mu^T(A, -b)$. □

We provide also a version of the theorem with equalities.

Corollary 7.3. *Let $P = \{x \mid Ax \leq b,\ Cx = d\}$ be a polyhedral set. Then either $P \neq \emptyset$, or there exists $\mu \geq 0$, ν such that $\mu^T A + \nu^T C = 0$, $\mu^T b + \nu^T d = -1$.*

Proof. Represent $P = \{x \mid Ax \leq b,\ Cx \leq d,\ -Cx \leq -d\}$. By the previous theorem, either $P \neq \emptyset$, or there exist $\mu, \nu_+, \nu_- \geq 0$ such that $\mu^T A + \nu_+^T C - \nu_-^T C = 0$, $\mu^T b + \nu_+^T d - \nu_-^T d = -1$. In the latter case, set $\nu = \nu_+ - \nu_-$. □

7.2 Standard form

Consider the problem of minimization of a linear objective function under linear equality and inequality constraints,

$$\min_{x \in \mathbb{R}^n} c^T x \ : \qquad Ax = b,\ Cx \leq d.$$

Let m be the number of inequalities. We introduce an additional variable $y \in \mathbb{R}^m$ and reformulate the problem as

$$\min_{x,y} \langle (c, \mathbf{0}), (x,y) \rangle \ : \qquad \begin{pmatrix} A & 0 \\ C & I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}, \quad y \in \mathbb{R}^m_+.$$

The *slack variable* y allowed to turn the inequalities into equalities, at the cost of introducing the *conic constraint* $y \in \mathbb{R}^m_+$.

The variable x can now be eliminated using the linear equality constraints. Indeed, if the kernel L of the matrix $\begin{pmatrix} A \\ C \end{pmatrix}$ is trivial, then there exist n equalities which determine x completely as linear functions of y.

If the kernel of this matrix is non-trivial, then x has degrees of freedom which are not determined by the equality constraints. For every $p \in L$ we have that $(x + p, y)$ is a feasible point whenever (x,y) is feasible. If the linear functional c does not vanish on the kernel identically, then the problem is either infeasible or unbounded. If c vanishes on the kernel, then these degrees of freedom are redundant for the problem.

We can hence assume the linear program (LP) in the following *standard form*:

$$\min_{x \geq 0} c^T x \ : \qquad Ax = b. \tag{1}$$

Without loss of generality A can be assumed of full row rank, otherwise the problem is infeasible or equality constraints are redundant. We shall also assume that $b \geq 0$, by possibly multiplying some rows of A by -1.

7.3 Duality

Consider the LP in standard form

$$\min_{x \geq 0} c^T x \ : \qquad Ax = b \tag{2}$$

with the matrix A being of full row rank of size $m \times n$. Let now $y \in \mathbb{R}^m$ such that $c \geq A^T y$. We claim that the optimal value of (2) is bounded from *below* by the quantity $y^T b$.

Indeed, we have for every feasible x that

$$y^T b = y^T Ax \leq c^T x. \tag{3}$$

In this way y provides a *certificate* that the optimal value of LP (2) is not below a certain value. It is then natural to ask what the *best* such certificate is and which value it provides. We may formulate this question as another optimization problem:

$$\max_y b^T y \ : \qquad A^T y \leq c.$$

Observe that this problem is also a linear program. It is called the *dual* LP, in contrast to (2) which is called the *primal*.

The dual program is hence a *maximization* problem, and every feasible point for the dual problem yields a *lower bound* on the objective value of the primal problem. Vice versa, from (3) it also follows that every feasible point for the primal problem yields an *upper bound* on the objective value of the dual problem.

Introduce the slack variable $s \geq 0$ and reformulate the dual problem as

$$\max_{s\geq 0, y} b^T y \ : \qquad s + A^T y = c. \tag{4}$$

If (x, s) is a primal-dual feasible pair, then the difference of the respective objective values is given by

$$c^T x - b^T y = c^T x - x^T A^T y = c^T x - x^T (c - s) = \langle s, x \rangle. \tag{5}$$

Hence the *complementarity* condition $s^T x = 0$ implies that both x, s are optimal solutions of the respective LP. This condition can equivalently be written as $x_i s_i = 0$ for all $i = 1, \dots, n$ or for every i, $x_i = 0$ or $s_i = 0$.

The dual points s and the objective functional c can be thought of as elements of the dual space to $\mathbb{R}^n$.

We now come to the main result in the duality theory of linear programs.

Theorem 7.4. *(Strong duality for LP) If both problems* (2) *and* (4) *are feasible, then their objective values coincide and are attained at a complementary primal-dual pair of feasible points.*

Proof. If the dual problem is feasible, then the optimal value of the primal problem is lower bounded. Since the primal is also feasible, its optimal value is finite. Likewise, the dual optimal value is finite. Let v^* be the primal optimal value.

Let now $v \in \mathbb{R}$ be arbitrary and suppose that the half-space $\{x \,|\, c^T x \leq v\}$ has an empty intersection with the non-empty polyhedron $C = \{x \,|\, -x \leq 0, \ Ax = b\}$. By the Theorem on the Alternative there exist $\lambda_0, \lambda \geq 0$, μ such that $\lambda_0 c - \lambda^T I - \mu^T A = 0$, $\lambda_0 v - \mu^T b = -1$. If $\lambda_0 = 0$, then $C = \emptyset$, leading to a contradiction. Hence $\lambda_0 > 0$. Set $y = \frac{\mu}{\lambda_0}$. Then the relations can be rewritten as $c - y^T A \geq 0$, $v + \frac{1}{\lambda_0} = y^T b$. But this means that (4) has a feasible solution with value $v + \lambda_0^{-1} > v$.

Therefore for every $v < v^*$ program (4) has a feasible solution with value $> v$. On the other hand, the optimal value of (4) is upper bounded by any number strictly greater than v^*. Hence the optimal value of (4) equals v^*. If LP (2) does not attain its optimal value, then we may set $v = v^*$ and there exists a dual feasible point with value $> v^*$, a contradiction. Hence (2) attains its optimum. In a similar manner, (4) attains its optimum.

Let x^*, y^* be the optimal solutions, and set $s^* = c - A^T y^*$. Then

$$0 = v^* - v^* = c^T x^* - b^T y^* = (x^*)^T s^*,$$

and x^*, s^* are complementary. □

If (4) is unbounded, then (2) must be infeasible, and when (2) is unbounded, then (4) must be infeasible. The converse is not true, however, as the following example shows.

Example: Consider the LP

$$\min_{x=(x_1,x_2)^T\geq 0} -x_2 \ : \qquad x_1 = -1.$$

Clearly this program is infeasible. The dual can be written as

$$\max_y (-y) \ : \qquad (y, 0) \leq (0, -1),$$

which is also infeasible.

Table 1: Optimality conditions maintained by different methods

	Primal simplex	Dual simplex	Interior-point	Infeasible interior-point
$x \geq 0$	Yes	No	Yes	Yes
$s \geq 0$	No	Yes	Yes	Yes
$x \in \mathcal{A}_P$	Yes	Yes	Yes	No
$s \in \mathcal{A}_D$	Yes	Yes	Yes	No
$x_i s_i = 0$	Yes	Yes	No	No

7.4 Optimality conditions

Consider the LP

$$\min_{x\geq 0} \langle c, x\rangle \quad : \qquad Ax = b, \tag{6}$$

where $A \in \mathbb{R}^{m\times n}$ is supposed to be of full row rank, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $x \in \mathbb{R}^n_+$ is the decision variable. We may also assume that $b \geq 0$, by possibly multiplying some rows of A by -1. We shall call problem (6) the *primal* linear program. The feasible set of this program is given by the intersection of the nonnegative orthant $\mathbb{R}^n_+$ with the $(n-m)$-dimensional affine subspace

$$\mathcal{A}_P = \{x \in \mathbb{R}^n \mid Ax = b\}.$$

The dual program can be written in the form

$$\max_{y,s\geq 0} \langle b, y\rangle \quad : \qquad s + A^T y = c \tag{7}$$

with decision variables $s \in \mathbb{R}^n_+$, $y \in \mathbb{R}^m$. Note that since A^T has full column rank, the variables y can be eliminated from the problem by using m of the linear equality relations. Hence we may interpret the linear equality constraint as an inclusion $s \in \mathcal{A}_D$ of s into an m-dimensional affine subspace

$$\mathcal{A}_D = \{s \in \mathbb{R}^n \mid \exists\, y \in \mathbb{R}^m : \quad s + A^T y = c\}$$

of $\mathbb{R}^n$. Note that for every $s \in \mathcal{A}_D$ there exists a unique $y = y(s)$ certifying this inclusion.

As mentioned above, the optimal solutions x^*, s^* of problems (6),(7), respectively, are characterized by the *complementarity condition* $\langle x^*, s^*\rangle = 0$, which can be rewritten as

$$x_i^* s_i^* = 0 \qquad \forall\, i = 1, \dots, n.$$

We obtain the following result.

Lemma 7.5. *The pair $(x, s) \in \mathbb{R}^n \times \mathbb{R}^n$ is the primal-dual pair of optimal solutions if and only if it satisfies the conditions*

$$x \in \mathcal{A}_P, \quad s \in \mathcal{A}_D, \quad x \geq 0, \quad s \geq 0, \quad x_i s_i = 0\ \forall\, i = 1, \dots, n.$$

Finding a pair (x, s) satisfying all five conditions is hence as difficult as solving the LP. However, finding a pair satisfying a subset of the conditions is much simpler. Solution methods start from a pair satisfying a subset and then try to iteratively come closer to satisfaction of the remaining conditions. Table 1 shows which methods maintain which conditions during their iterations, aiming at closing the gaps defined by the remaining conditions.

Thus the simplex method visits points where the some of the inequality constraints are active (hence the name active set method), while interior point methods generate points where the inequalities are strict. In the next sections we consider these methods in more detail.

Now we shall, however, provide a primal-dual symmetric formulation of programs (6),(7). The following fact is easily verified:

The affine subspaces $\mathcal{A}_P, \mathcal{A}_D$ defined by the primal and dual equality constraints are orthogonal to each other.

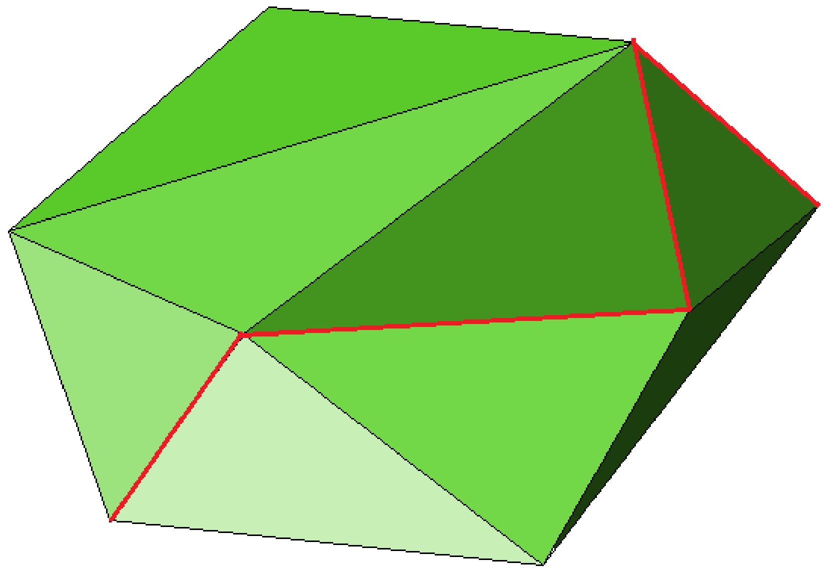

Figure 2: The simplex algorithm jumps from vertex to vertex along the edges of the polyhedron.

Indeed, every vector δ_x which is parallel to $\mathcal{A}_P$ obeys the relation $A\delta_x = 0$, while every vector δ_s parallel to $\mathcal{A}_D$ obeys $\delta_s = A^T\delta_y$ for some δ_y. Thus $\langle \delta_x, \delta_s \rangle = \delta_x^T A^T \delta_y = 0$, which proves our claim.

From (5) it then follows that for every fixed $s \in \mathcal{A}_D$, the linear form $\langle s, x \rangle$ differs by a constant from the objective function of problem (6) on $\mathcal{A}_P$. On the other hand, for every fixed $x \in \mathcal{A}_P$ this form differs by a constant from the objective of problem (7) on $\mathcal{A}_D$.

7.5 Primal simplex method

The feasible set of LP (6) is the intersection of the orthant $\mathbb{R}^n_+$ with an $(n-m)$-dimensional affine subspace $\mathcal{A}_P$. Hence at any extremal point of this polyhedron at least $n-m$ of the inequality constraints $x_i \geq 0$ must be active. Note that each variable x_i is associated to a column of the coefficient matrix A. Specifying $n-m$ indices $i \in \{1, \dots, n\}$ where $x_i \geq 0$, in general allows to recover the values of the remaining entries of x by using the equality constraints. This happens if and only if the remaining m columns of A form a basis of $\mathbb{R}^m$. Such a set of m indices is called *basic*, the set of the other $n-m$ indices *non-basic*. The basic set is called *primal feasible* if the values of the basic primal variables are nonnegative, and hence the corresponding primal vector is feasible.

The optimal value of the objective, if it exists, is attained at a vertex of the polyhedron, which corresponds to one or more basic primal feasible sets. The simplex algorithm goes from such set to the other changing indices one by one, while decreasing monotonically the value of the objective function at the corresponding vertex. Changing one index in the basic set by dropping a basic index and including a non-basic one leads to either staying at the same vertex or moving along an edge of C to a new vertex, under the condition that the basic primal variables stay nonnegative.

Which index will be added to the basic set is decided by a pivoting rule, which also determines the concrete variant of the method. This choice determines the edge along which the iterate moves. The basic index which is dropped is determined by which constraint $x_i \geq 0$, $i \in B$, becomes active at the opposite side of the edge. Equivalently, it is the constraint which first becomes active along the ray in the direction of the edge.

The algorithm terminates in a finite number of steps, by either

- finding that a solution cannot be improved and is optimal, or
- finding an edge with decreasing cost function that recedes to infinity, in which case the LP is unbounded.

The algorithm needs a starting vertex to commence. Such a vertex is found by solving the auxiliary LP

$$\min_{x\geq 0, z\geq 0} \mathbf{1}^T z \ : \qquad Ax + z = b.$$

Let C' be the feasible set of the auxiliary problem. Then the feasible set of the original problem is given by $C = \{x \,|\, (x,0) \in C'\}$. In case $C \neq \emptyset$ the optimal value of the auxiliary problem equals zero, and every optimal vertex of C' corresponds to a vertex of C. In the case $C = \emptyset$ the optimal value of the auxiliary problem is strictly positive. We then launch the simplex algorithm on the auxiliary problem with the set of basic variables z. This set corresponds to the vertex $(0, b)$ of C' (recall that $b \geq 0$).

Although the simplex method works very well in practice, its worst-case performance is exponential in the number of variables in standard form, i.e., in the number of inequality constraints.

We now consider the simplex method in more detail.

We shall denote the basic and the non-basic index sets by B and N, respectively. Define also corresponding sub-vectors $x_B \in \mathbb{R}^m$, $x_N \in \mathbb{R}^{n-m}$, $c_B \in \mathbb{R}^m$, $c_N \in \mathbb{R}^{n-m}$, and sub-matrices $A_B \in \mathbb{R}^{m\times m}$, $A_N \in \mathbb{R}^{m\times(n-m)}$. Then the equality constraints can be written as

$$A_B x_B + A_N x_N = b, \qquad x_B + (A_B^{-1}A_N)x_N = A_B^{-1}b. \tag{8}$$

Thus the basic set B can be associated to the primal vector $x = (x_B, x_N) = (A_B^{-1}b, 0)$. This vector by definition satisfies the equality constraints of problem (6), and it satisfies the inequality constraints if and only if $A_B^{-1}b \geq 0$.

The cost function of problem (6) is given by

$$\langle c_B, x_B\rangle + \langle c_N, x_N\rangle = \langle c_B, A_B^{-1}b - (A_B^{-1}A_N)x_N\rangle + \langle c_N, x_N\rangle = \langle c_N - A_N^T A_B^{-T} c_B, x_N\rangle + \langle c_B, A_B^{-1}b\rangle.$$

Note that the objects

$$M = A_B^{-1}A_N, \quad \mu = A_B^{-1}b, \quad \xi = c_N - A_N^T A_B^{-T} c_B = c_N - M^T c_B, \quad \gamma = \langle c_B, A_B^{-1}b\rangle = \langle c_B, \mu\rangle \tag{9}$$

contain the complete information of the LP. The quantity γ is the value of the objective function at the vertex corresponding to the basic set B.

The basic set B is feasible if $\mu \geq 0$. Moreover, if $\xi \geq 0$, then it is optimal, because for every feasible point $x = (x_b, x_N)$ of problem (6) we have

$$\langle c, x\rangle = \langle \xi, x_N\rangle + \gamma \geq \gamma.$$

The primal simplex method evolves the objects M, μ, ξ, γ such that μ stays nonnegative, and γ monotonely decreases until ξ also becomes nonnegative. The objects are stored in a *tableau*, which is a matrix of size $(m+1)\times(n-m+1)$ given by

$-\gamma$	ξ^T
μ	M

Each row of the tableau except the first one corresponds to a basic index $i \in B$, each column except the first one to a non-basic index $j \in N$.

Exchanging a basic index $i \in B$ with a non-basic index $j \in N$ is equivalent to the transformation

$$\begin{bmatrix} i & \leftarrow & j \\ j & \leftarrow & i \\ \mu_{\tilde B} & \leftarrow & \mu_{\tilde B} - M_{ij}^{-1}M_{\tilde B j}\mu_i \\ \mu_i & \leftarrow & M_{ij}^{-1}\mu_i \\ \xi_{\tilde N} & \leftarrow & \xi_{\tilde N} - M_{ij}^{-1}\xi_j M_{i\tilde N}^T \\ \xi_j & \leftarrow & -M_{ij}^{-1}\xi_j \\ -\gamma & \leftarrow & -\gamma - M_{ij}^{-1}\xi_j\mu_i \\ M_{\tilde B\tilde N} & \leftarrow & M_{\tilde B\tilde N} - M_{ij}^{-1}M_{\tilde B j}M_{i\tilde N} \\ M_{\tilde B j} & \leftarrow & -M_{ij}^{-1}M_{\tilde B j} \\ M_{i\tilde N} & \leftarrow & M_{ij}^{-1}M_{i\tilde N} \\ M_{ij} & \leftarrow & M_{ij}^{-1} \end{bmatrix}, \tag{10}$$

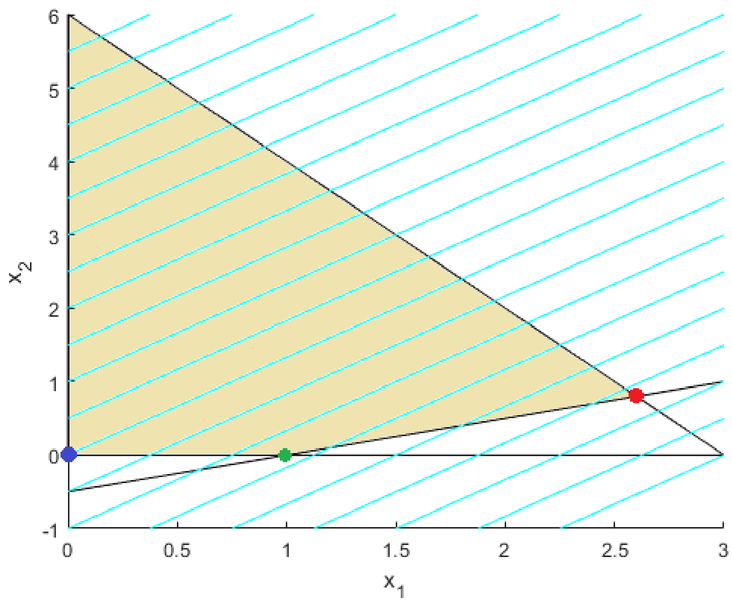

Figure 3: Feasible region, niveau lines of the objective, and visited vertices for LP (11).

where $\tilde{B} = B \setminus \{i\}$, $\tilde{N} = N \setminus \{j\}$.

The method aims at making all elements of ξ nonnegative while decreasing γ and keeping μ nonnegative. From the transformation law of μ_i it follows that the *pivot element* M_{ij} has to be positive. From the transformation law of ξ_j it then follows that the simplex step can transform a negative element ξ_j into a positive one. On the other hand, for all $k \in \tilde{B}$ we have to ensure that $\mu_k \leftarrow \mu_k - M_{ij}^{-1} M_{kj} \mu_i \geq 0$. This is automatically true for those k such that $M_{kj} \leq 0$, but for those k which satisfy $M_{kj} > 0$ we have to ensure $M_{kj}^{-1}\mu_k \geq M_{ij}^{-1}\mu_i$.

Hence the simplex step performs as follows:

- choose $j \in N$ such that $\xi_j < 0$;
- among those $k \in B$ such that $M_{kj} > 0$, let i be the index minimizing the ratio $M_{kj}^{-1}\mu_k$;
- apply rule (10) to update the tableau.

If in first item no index j can be found, then $\xi \geq 0$ and the tableau is already optimal. If in the second item no index i can be found, then $M_{kj} \leq 0$ for all $k \in B$ and the non-basic variable x_j can be increased unbounded while staying in the feasible set of problem (6). This decreases the objective value towards $-\infty$, and the problem is unbounded. Note that the problem is feasible due to the existence of the tableau.

Example: Consider the LP

$$\min_{x \in \mathbb{R}^2_+} (3x_2 - 4x_1) : \qquad x_1 - 2x_2 \leq 1,\ 2x_1 + x_2 \leq 6. \tag{11}$$

The feasible region and the niveau lines of the objective are depicted in Fig. 3.

In standard form this LP can be written as

$$\min_{x \in \mathbb{R}^4_+} (3x_2 - 4x_1) : \qquad \begin{pmatrix} 1 & -2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 1 \\ 6 \end{pmatrix}.$$

Obviously the point $x = (0,0)^T$ is feasible for the original LP (11), which corresponds to the point $x = (0,0,1,6)^T$ for the standard form. This vertex is represented by the basic index set $B = (3,4)$ and the non-basic index set $(1,2)$. The elements of the corresponding tableau are by virtue of (9) given by

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}, \quad \mu = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \end{pmatrix},$$

$$\xi = \begin{pmatrix} -4 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^T \begin{pmatrix} 1 \\ 6 \end{pmatrix} = 0.$$

This yields the tableau

0	−4	3
1	1	−2
6	2	1

It corresponds to the blue vertex in Fig. 3. The non-basic index to be moved to the basic set is $j = 1$, because $\xi_1 = -4$ is the only negative element of ξ. Both coefficients in the corresponding column are positive, but the ratio $\frac{\mu_3}{M_{31}} = 1$ is smaller than $\frac{\mu_4}{M_{41}} = 3$, and the basic index $i = 3$ has to be moved to the non-basic set. Exchanging the two indices by virtue of (10) yields the tableau

4	4	−5
1	1	−2
4	−2	5

We are now located in the point $x = (1, 0, 0, 4)$, corresponding to the green vertex on Fig. 3. Note that the objective value has decreased to -4. For the next iteration we have to choose the non-basic index $j = 2$. In the corresponding column only the element $M_{42} = 5$ is positive, and we move $i = 4$ to the non-basic set. This yields the tableau

8	2	1
$\frac{13}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
$\frac{4}{5}$	$-\frac{2}{5}$	$\frac{1}{5}$

This tableau corresponds to the point $x = (\frac{13}{5}, \frac{4}{5}, 0, 0)$ and the blue vertex on Fig. 3. Now we have $\xi \geq 0$, and hence we have reached the optimal solution, with value -8.

7.6 Dual simplex method

The simplex tableau can also be interpreted in terms of the dual variables s. In order to ensure the complementarity condition $x_i s_i = 0$ for all i, we declare the subvector s_B of the dual variables to be non-basic, and the subvector s_N to be basic. Then for every $i \in \{1, \dots, n\}$, exactly one variable in the pair (x_i, s_i) is non-basic, and the product $x_i s_i$ is always zero at the primal and dual vertices represented by the tableau.

The primal variables obey the relation $x_B + Mx_N = (I\ M)(x_B; x_N) = \mu$, while the cost is given by $\langle \xi, x_N \rangle + \gamma = \langle (0; \xi), (x_B; x_N) \rangle + \gamma$. Hence the equality constraint in (7) can be written as

$$\begin{pmatrix} s_B \\ s_N \end{pmatrix} + \begin{pmatrix} I \\ M^T \end{pmatrix} y = \begin{pmatrix} 0 \\ \xi \end{pmatrix}.$$

Eliminating y, we obtain the relation

$$s_N - M^T s_B = \xi$$

on the dual variables s, while the dual cost function is given by

$$\langle \mu, y \rangle + \gamma = -\langle \mu, s_B \rangle + \gamma.$$

Taking into account, that the dual problem maximizes the cost and we hence have to multiply it by -1 to obtain a minimization problem, the simplex tableau for the dual problem is given by

γ	μ^T
ξ	$-M^T$

Here the rows of the tableau are indexed by the dual basic set N, while the columns are indexed by the dual non-basic set B. The dual simplex tableau can hence be obtained from the primal one by transposition and multiplying the diagonal blocks by -1. The tableau is feasible if $\xi \geq 0$ and optimal if in addition $\mu \geq 0$. The value of γ increases monotonely.

Instead of applying primal simplex operations to the dual simplex tableau, we may apply the equivalent operations to the primal tableau. This is how the *dual simplex algorithm* operates. Each of its steps consists of the following stages:

- choose $i \in B$ such that $\mu_i < 0$;
- among those $k \in N$ such that $M_{ik} < 0$, let j be the index minimizing the ratio $-M_{ik}^{-1}\xi_k$;
- apply rule (10) to update the tableau.

The algorithm stops if

- all μ_i are nonnegative (optimality);
- all M_{ik} are nonnegative (unbounded-ness of the dual, or infeasibility of the primal).

Dual feasible simplex tableaux can be created from an optimal simplex tableau for a given LP if new constraints are added to the LP or existing constraints are modified. In this case new rows are added to the tableau, possibly creating negative elements in the vector μ, or existing elements of μ are modified, while the vector ξ remains unchanged and hence nonnegative.

The dual simplex method is hence suitable for the warm-start of the solution process of a slightly modified LP if an optimal solution of the original LP is available. A prime example of the application of the dual simplex algorithm will be considered in the next section.

Consider again LP (11) with optimal solution $(x_1, x_2) = (\frac{13}{5}, \frac{4}{5})$. Suppose we add a new constraint $x_1 \leq 2$. This constraint diminishes the feasible set of the LP, and the formerly optimal point becomes infeasible. However, we may use this point to construct a dual feasible simplex tableau, which we can then optimize by the dual simplex algorithm.

Introduce a new slack variable

$$x_5 = 2 - x_1$$

This variable is basic, and we have to add a new row corresponding to it to the tableau. Hence the basic index set becomes $B = (1, 2, 5)$, while the non-basic set stays the same, $N = (3, 4)$. Let us compute the new row. We have to determine the slack x_5 as a function of the non-basic variables x_3, x_4. Since it is defined by means of the basic variable x_1, we have to use the expression of x_1 as a function of x_3, x_4, which is encoded in the corresponding row 1 of the tableau. We have

$$x_5 = 2 - (\frac{13}{5} - \frac{1}{5}x_3 - \frac{2}{5}x_4) = -\frac{3}{5} - (-\frac{1}{5}x_3 - \frac{2}{5}x_4),$$

which yields the new tableau

8	2	1
$\frac{13}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
$\frac{4}{5}$	$-\frac{2}{5}$	$\frac{1}{5}$
$-\frac{3}{5}$	$-\frac{1}{5}$	$-\frac{2}{5}$

We now apply a dual simplex step to it. The only negative element of μ is $\mu_5 = -\frac{3}{5}$, hence we choose the corresponding row as a pivot. In this row, both coefficients are negative, and we have to look which column k yields the minimum ratio $-\frac{\xi_k}{M_{5k}}$. Since $-\frac{\xi_4}{M_{54}} = \frac{5}{2} < 10 = -\frac{\xi_3}{M_{53}}$, we choose the pivot column 4. Exchanging the indices 4 and 5 leads to the basic set $B = (1, 2, 4)$, the non-basic set $N = (3, 5)$, and the tableau

$\frac{13}{2}$	$\frac{3}{2}$	$\frac{5}{2}$
2	0	1
$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{5}{2}$

This tableau is now optimal, it corresponds to the point $(x_1, x_2) = (2, \frac{1}{2})$ with value $-\frac{13}{2}$. The evolution of the vertex during the dual simplex algorithm is depicted in Fig. 4.

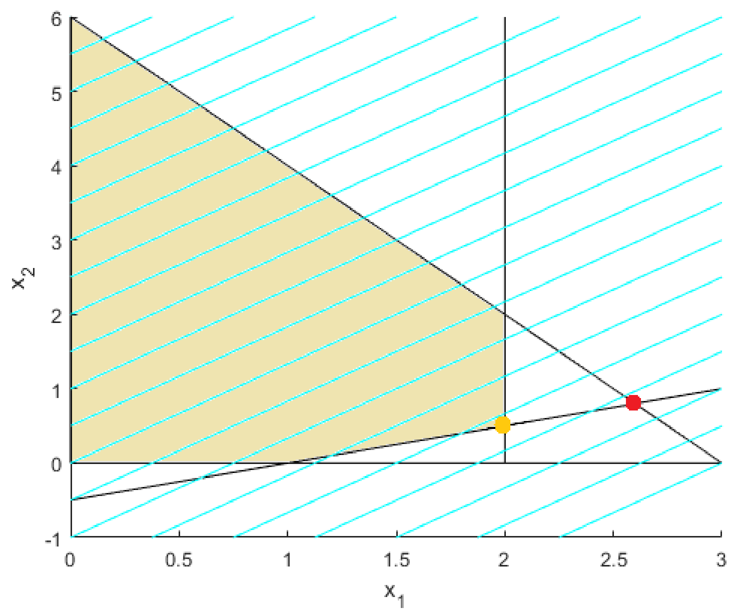

Figure 4: Feasible region, niveau lines of the objective, and visited vertices: originally optimal vertex (red), newly optimal vertex (orange).

7.7 Application: mixed integer linear programs

A mixed integer linear program (MILP) is an LP with additional integrality constraints on a part of the decision variables:

$$\min_{x\geq 0} \langle c, x\rangle : \qquad Ax = b, \quad x_i \in \mathbb{Z} \quad \forall\, i \in I,$$

where I is a subset of indices. Such a problem is non-convex and actually in general NP-hard.

By removing the integrality constraints we obtain the *linear relaxation* of the program, namely the LP

$$\min_{x\geq 0} \langle c, x\rangle : \qquad Ax = b. \tag{12}$$

Since the feasible set of the LP is larger than that of the original MILP, its optimal value is a lower bound on the value of the MILP. Let x^* be the solution of this LP.

If the subvector x_I^* of the solution happens to be integral, then x^* is feasible for the MILP and hence yields its optimal solution. However, in general there exists an index $i \in I$ such that x_i^* is fractional. Consider the two linear programs

$$\min_{x\geq 0} \langle c, x\rangle : \qquad Ax = b, \quad x_i \leq \lfloor x_i^* \rfloor, \tag{13}$$

$$\min_{x\geq 0} \langle c, x\rangle : \qquad Ax = b, \quad x_i \geq \lceil x_i^* \rceil. \tag{14}$$

The feasible sets of LPs (13),(14) are disjoint, but their union contains the feasible set of the original MILP. On the other hand, neither of the feasible sets of LPs (13),(14) contains the solution x^* of the original linear relaxation (12). Hence the minimum of the two values of LPs (13),(14) is a better lower bound on the optimal value of the MILP than the optimal value of LP (12).

The process of splitting LP (12) into two stronger LPs (13),(14) by constraining one of the integer variables is called *branching*. MILP solvers proceed by recursively splitting the feasible set of the MILP into smaller parts by branching on integer variables whose value happened to be fractional in the solutions of the relaxations. Since the LP relaxations yield bounds on the value of the original MILP, the whole algorithm is called *branch-and-bound*.

Modern MILP solvers usually bring forward additional features strengthening the LP relaxations, such as presolve algorithms tightening the bounds on the integer variables or cuts separating fractional solutions from the feasible set of the MILP.

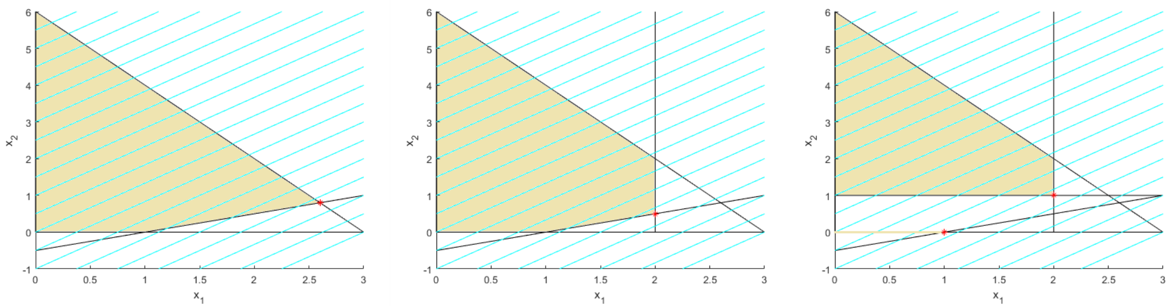

Figure 5: Feasible region and optimal solution for LP (11) (left), LP (15) (center), and LPs (16),(17) (right).

We shall stress one property which makes the dual simplex method a particularly useful method for solving the LP relaxations appearing in the course of the branch-and-bound algorithm.

Suppose we use the simplex method to solve LP (12) and obtained an optimal simplex table. LPs (13),(14) differ from (12) by the addition of one constraint. Introducing corresponding slack variables, we see that at the optimal point x^* of LP (12) these variables are negative, because x^* is not feasible for LPs (13),(14). Hence these slacks enter the basic set of variables. By adding a new row to the table corresponding to the slack, the table remains dual feasible, but loses primal feasibility at just one entry, which will be contained in the interval $(-1, 0)$. It will therefore in general take only a few dual simplex iterations to return the table to optimality and thus to solve LPs (13),(14).

During a general branching step, the constraint on the integer variable defining the branching will not be added, but merely tightened. This corresponds to changing the value of one entry in the vector b of the table, after which again the dual simplex method can be started to return the table to optimality.

Example: Consider the MILP

$$\min_{x \in \mathbb{R}^2_+} (3x_2 - 4x_1): \qquad x_1 - 2x_2 \leq 1,\ 2x_1 + x_2 \leq 6,\ x \in \mathbb{Z}^2.$$

Its first LP relaxation is given by (11) with optimal solution $x^* = (\frac{13}{5}, \frac{4}{5})$ and value -8.

This solution is not integer, and both integer variables have fractional values. We may thus branch on either variable. Choosing the variable x_1, we obtain the infeasible LP

$$\min_{x \in \mathbb{R}^2_+} (3x_2 - 4x_1): \qquad x_1 - 2x_2 \leq 1,\ 2x_1 + x_2 \leq 6,\ x_1 \geq 3$$

and the LP

$$\min_{x \in \mathbb{R}^2_+} (3x_2 - 4x_1): \qquad x_1 - 2x_2 \leq 1,\ 2x_1 + x_2 \leq 6,\ x_1 \leq 2, \tag{15}$$

whose solution is given by $x^* = (2, \frac{1}{2})$ with value $-\frac{13}{2}$. Thus the lower bound on the optimal value of the MILP has improved from -8 to $\min(+\infty, -\frac{13}{2}) = -\frac{13}{2}$.

We need to pursue only the branch defined by the second LP (15). Its solution has only one fractional variable x_2, with value $\frac{1}{2}$, branching on which yields the two LPs

$$\min_{x \in \mathbb{R}^2_+} (3x_2 - 4x_1): \qquad x_1 - 2x_2 \leq 1,\ 2x_1 + x_2 \leq 6,\ x_1 \leq 2,\ x_2 \geq 1, \tag{16}$$

$$\min_{x \in \mathbb{R}^2_+} (3x_2 - 4x_1): \qquad x_1 - 2x_2 \leq 1,\ 2x_1 + x_2 \leq 6,\ x_1 \leq 2,\ x_2 \leq 0. \tag{17}$$

Both LPs yields integer solutions, namely $(2, 1)$ and $(1, 0)$, with values -5 and -4, respectively.

Thus the optimal value of the MILP is the lower of these values, namely -5, and the corresponding solution is $(2, 1)$.

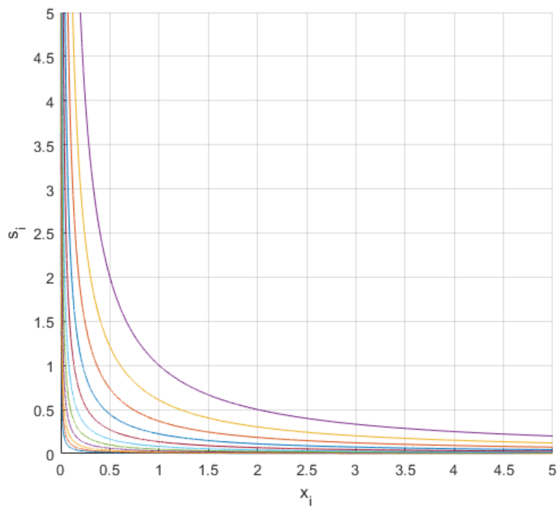

Figure 6: Approximations of the complementarity condition $x_i s_i = 0$.

7.8 Interior-point methods

As mentioned above, interior-point methods for LPs maintain the inequality constraints $x \geq 0$, $s \geq 0$, and possibly the equality constraints $x \in \mathcal{A}_P$, $s \in \mathcal{A}_D$, while trying to get increasingly closer to satisfaction of the complementarity condition $x_i s_i = 0$, $i = 1, \dots, n$.

In the two-dimensional space of the variables x_i, s_i, the complementarity condition defines a non-convex, non-smooth set, namely the union of the positive axes. We approximate this set by a family of hyperbolas $x_i s_i = \mu$. As $\mu \to 0$, these hyperbolas tend to the union of the axes (see Fig. 6).

The method aims at decreasing $\mu = \frac{\langle x,s \rangle}{n}$ in a way such that the individual products $x_i s_i$ tend to zero uniformly. More precisely, the iterates should not leave the set

$$N_\gamma = \{(x,s) \in \mathbb{R}_+^n \times \mathbb{R}_+^n \mid x_i s_i \geq \gamma\mu \ \forall\, i = 1, \dots, n\},$$

where $\gamma < 1$ is a positive constant, typically of the order 10^3. An interior-point iteration consists of a Newton step towards the solution of the non-linear system of equations

$$x \in \mathcal{A}_P, \quad s \in \mathcal{A}_D, \quad x_i s_i = \mu_k \ \forall\, i = 1, \dots, n,$$

where μ_k measures the accuracy achieved at the current iteration and $\mu_{k+1} = \sigma\mu_k$ for some suitable $\sigma \in (0,1)$.

Feasible methods generate iterates in the affine subspace $(x,s) \in \mathcal{A}_P \times \mathcal{A}_D$. The innovations δ_x, δ_s are hence located in the linear subspaces underlying $\mathcal{A}_P, \mathcal{A}_D$. For convenience, we consider also the innovation δ_y of the auxiliary variable $y \in \mathbb{R}^m$ which parameterizes the affine subspace $\mathcal{A}_D$. For given values (x,s,y) at the current iterate, satisfying the relations $Ax = b$, $s + A^T y = c$, and given $\mu > 0$ we wish to solve the system of equations

$$A(x + \delta_x) = b, \quad s + \delta_s + A^T(y + \delta_y) = c, \quad (x_i + \delta_{x,i})(s_i + \delta_{s,i}) = \mu. \tag{18}$$

Linearizing the system around $(\delta_x, \delta_s, \delta_y) = 0$ and solving for the innovations, we obtain the linear system

$$\begin{pmatrix} A & 0 & 0 \\ 0 & I & A^T \\ \operatorname{diag}(s) & \operatorname{diag}(x) & 0 \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_s \\ \delta_y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mu \cdot \mathbf{1} - x \bullet s \end{pmatrix},$$

where $\mathbf{1}$ is the all-ones vector and $\bullet$ stands for element-wise multiplication.

The solution of this system cannot always be used to determine the innovations. We have also to ensure that the next iterate stays in the set N_γ. To this end, we introduce a damping coefficient $\alpha \in (0,1]$ and make a step

$$(x,s,y) \leftarrow (x,s,y) + \alpha(\delta_x, \delta_s, \delta_y). \tag{19}$$

Inclusion into N_γ then can be written as

$$(x_i + \alpha\delta_{x,i})(s_i + \alpha\delta_{s,i}) \geq \gamma\mu \quad \forall\, i = 1, \dots, n.$$

The actual step-size α is then chosen as the minimum of the value 1 and the largest positive value α which satisfies these n scalar quadratic inequalities.

The algorithm can be designed with an adaptive step-size: if the values of α become smaller, the quantity μ can be updated less aggressively, and if the algorithm makes full steps, then we may decrease the factor σ which is used to drive μ towards zero.

Since the scalar product $\langle x, s\rangle$ equals the duality gap between the primal and dual objective values at the current iterates x, s, the progress of the algorithm can be monitored by tracking the evolution of μ.

Clearly for a feasible method to work, we have to require existence of a strictly feasible primal-dual pair (x, s). In this case we can parameterize the set of all such pairs by the positive orthant $\mathbb{R}^n_{++}$. Namely, we have the following result.

Lemma 7.6. *Suppose problems* (6),(7) *are both strictly feasible. Then for every $z \in \mathbb{R}^n_{++}$, there exists a unique strictly primal-dual feasible pair (x, s) such that $z = x \bullet s$.*

Proof. Let $z > 0$ be arbitrary. Consider the optimization problem

$$\min_{x \geq 0} \left(\langle c, x\rangle - \langle z, \log x\rangle\right) : \qquad Ax = b. \tag{20}$$

The feasible set of this problem is the same as in the primal LP, while the objective function is strictly convex. Since (6) is strictly feasible, problem (20) is strictly feasible too.

Note that on any compact subset of the feasible set the objective in problem (20) is bounded from below, and the objective tends to $+\infty$ as x tends to the boundary of the feasible set. We now show that there exists a solution to this problem. Let ϕ be its optimal value, including $-\infty$.

Suppose for the sake of contradiction that no solution exists. Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence such that $\langle c, x_k\rangle - \langle z, \log x_k\rangle \to \phi$ as $k \to \infty$. Then we must have $x_k \to \infty$ too. Choose a subsequence such that $\frac{x_k}{\|x_k\|}$ has a limit x^* on the unit sphere as $k \to \infty$. Since the objective values approach the limit ϕ from above, we must have $\langle c, x^*\rangle \leq 0$, otherwise the linear term dominates the logarithmic one and the values tend to $+\infty$. On the other hand, x^* is a recessive direction of the feasible set, i.e., $Ax^* = 0$. This yields for every dual feasible vector $s = c - A^T y$ that

$$0 \leq \langle s, x^*\rangle = \langle c, x^*\rangle - \langle y, Ax^*\rangle = \langle c, x^*\rangle \leq 0.$$

Thus s lies on the boundary of the nonnegative orthant, and the dual problem (7) is not strictly feasible, a contradiction.

Let x^* be the unique solution of problem (20). The constraints $x \geq 0$ are not active at this solution. Hence the optimality conditions at x^* amount to the existence of a vector $y \in \mathbb{R}^m$ such that the Lagrangian

$$\langle c, x\rangle - \langle z, \log x\rangle - \langle y, Ax - b\rangle$$

has a vanishing derivative with respect to x at $x = x^*$. This yields

$$c - z \bullet (x^*)^{-1} - A^T y = 0,$$

and $s^* = z \bullet (x^*)^{-1}$ is strictly dual feasible. Then (x^*, s^*) provide the sought primal-dual strictly feasible pair.

On the other hand, every such pair (x, s) satisfies the optimality condition of problem (20) and must hence coincide with (x^*, s^*). □

Clearly for every strictly primal-dual feasible pair (x, s) we have $x \bullet s > 0$, and hence the set of such pairs is in bijection with the open orthant $\mathbb{R}^n_{++}$ under the quadratic map $(x, s) \mapsto x \bullet s$.

If the point $z \in \mathbb{R}^n_{++}$ tends to the origin, then its pre-image in $\mathcal{A}_P \times \mathcal{A}_D$ tends to the solutions of problems (6),(7). The interior-point method tries to trace the primal-dual strictly feasible curve defined by the pre-image of the ray generated by the all-ones vector $\mathbf{1}$ in $\mathbb{R}^n_{++}$.

Feasible methods have the inconvenience of necessitating a primal-dual feasible point in the interior of the orthants. This requires a preliminary phase to find such a point. Infeasible methods, on the contrary, try to decrease the difference $Ax - b$ and $s + A^T y - c$ along with μ. Linearizing equations (18) we obtain the linear system

$$\begin{pmatrix} A & 0 & 0 \\ 0 & I & A^T \\ \mathrm{diag}(s) & \mathrm{diag}(x) & 0 \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_s \\ \delta_y \end{pmatrix} = \begin{pmatrix} b - Ax \\ c - s - A^T y \\ \mu \cdot \mathbf{1} - x \bullet s \end{pmatrix}$$

with a modified right-hand side. When determining the step-size α in the iterate (19), we ensure not only inclusion of the next point in the set N_γ, but also the condition

$$||Ax - b|| + ||s + A^T y - c|| \leq \beta \cdot \mu$$

with β a suitable constant. This ensures that the feasibility gap tends to zero along with the duality gap. Checking this condition also amounts to a solution of a scalar quadratic equation with respect to the step size α.

Note that the mismatch $Ax - b, s + A^T y - c$ in the linear equality conditions decreases at each step by the factor $1 - \alpha$. Thus as soon as we make a full step with $\alpha = 1$, the primal-dual pair (x, s) becomes feasible and stays feasible during all subsequent iterations.

More details on interior-point methods for linear programming can be found, e.g., in [1].

7.9 Liftings

The complexity of an LP depends primarily on the number of linear inequalities. In some situations it may be beneficial to add additional design variables to *decrease* the number of inequalities.

Example: Consider the unit ball of the 1-norm in $\mathbb{R}^n$, $B_1 = \{x \,|\, ||x||_1 = \sum_i |x_i| \leq 1\}$. This polytope has 2^n facets, and hence a direct description will need 2^n linear inequalities. Let us, however, add n additional variables $s_1, \dots, s_n$, and consider the set

$$C = \left\{ (x, s) \,|\, -s_i \leq x_i \leq s_i,\ \sum_i s_i \leq 1 \right\}.$$

Then $x \in B_1$ if and only if there exists s such that $(x, s) \in C$. In other words, B_1 can be represented as a *projection* of C. In any LP involving B_1 this polytope can be replaced by C. The latter is described by only $2n + 1$ inequalities at the cost of n additional variables.

A description of a convex set by a projection of another (usually simpler) convex set is called a *lifting*. For a given polytope

$$X = \{x \,|\, A_{eq} x = b_{eq},\ A_{ineq} x \leq b_{ineq}\},$$

one can then ask what the minimal number of inequalities needed to describe X by a lifting is. The answer can be obtained by considering the *slack matrix* of X.

Let $v_1, \dots, v_N$ be the vertices of X, and let m be the number of inequalities in the description of X. The slack matrix of X is the nonnegative $m \times N$ matrix S with elements

$$S_{ij} = -(A_{ineq} v_j - b_{ineq})_i.$$

We have the following result [2].

Theorem 7.7. *Let X be a polytope and let S be its slack matrix. Suppose that every inequality is attained with equality somewhere on X, i.e., the inequalities are tight. Then the minimal number of inequalities required for a lifted representation of X is equal to the nonnegative rank of S.*

Here the nonnegative rank of S is the minimal number k such that there exists a *nonnegative* factorization $S = S_1 S_2$ with $S_1 \in \mathbb{R}^{m \times k}$, $S_2 \in \mathbb{R}^{k \times N}$.

Proof. Let $X = \{x \,|\, Cx = d,\ Ax \leq b\}$, and let $S = S_1 S_2$ be a nonnegative factorization of the slack matrix. Define the polytope

$$X' = \{(x, y) \,|\, Cx = d,\ Ax + S_1 y = b,\ y \geq 0\}.$$

The polytope X' is described by k inequalities, where k is the number of columns of S_1. We claim that X' is a lifting of X.

Indeed, every point $(x,y) \in X'$ projects to a point of X since $S_1 y \geq 0$. On the other hand, let v_i be the i-th vertex of X and s_i the i-th column of S_2. Then $(v_i, s_i) \in X'$, because $S_1 s_i$ is the i-th column of S and equals $-(Av_i - b)$. Hence the projection of X' is contained in X and contains all extreme points of X. Thus this projection equals X.

Let now $X = \{x \,|\, Ax \leq b\}$ be a polytope described by m inequalities and $X' = \{(x,y) \,|\, Cx + Dy = c,\ y \geq 0\}$ be an arbitrary lifting of X with k inequalities. Then for every vertex v_i, $i = 1,\dots,N$ of X there exists $(v_i, s_i) \in X'$ with $s_i \geq 0$. The polytope X' has an empty intersection with the half-space $\{(x,y) \,|\, a_j^T x > b_j\}$, where a_j is the j-th row of A^T. Since the hyperplane $\{(x,y) \mid a_j^T x = b_j\}$ is supporting to X', it follows that the inequality $a_j^T x \leq b_j$ is a linear combination of the equalities $Cx + Dy = c$ and inequalities $y \geq 0$ with arbitrary and nonnegative coefficients, respectively. Therefore there exist vectors $\mu_j \geq 0$, λ_j such that $\lambda_j^T C = a_j^T$, $\lambda_j^T D - \mu_j^T = 0$, $\lambda_j^T c = b_j$. The slack matrix of X is then given by the elements

$$S_{ji} = -a_j^T v_i + b_j = -\lambda_j^T C v_i + \lambda_j^T c = \lambda_j^T D s_i = \mu_j^T s_i.$$

Hence $S = S_1 S_2$ with $S_2 = (s_1, \dots, s_N)$ and $S_1^T = (\mu_1, \dots, \mu_m)$ is a nonnegative factorization of S with factors of sizes $m \times k$ and $k \times N$. □

However, finding the nonnegative rank of a given nonnegative matrix is a hard problem.

Example: Lifting of a regular 2^{n+2}-gon.

Let us consider the planar polyhedral set P given by

$$\Bigg\{ x \mid \exists\, u_0,\dots,u_n,v_1,\dots,v_n :\ v_i = \begin{pmatrix} \cos\frac{\pi}{2^{i+1}} & \sin\frac{\pi}{2^{i+1}} \\ -\sin\frac{\pi}{2^{i+1}} & \cos\frac{\pi}{2^{i+1}} \end{pmatrix} u_{i-1},\ u_{i,1} = v_{i,1},\ u_{i,2} \geq |v_{i,2}|,\ i = 1,\dots,n;$$
$$|x| \leq u_0,\ \begin{pmatrix} 1 & 0 \\ -\tan\frac{\pi}{2^{n+2}} & 1 \end{pmatrix} u_n \leq (1,0)^T \Bigg\},$$

where $n \geq 0$ is an integer and the inequalities are meant element-wise.

For every point $x \in P$ and feasible vectors $u_0,\dots,u_n,v_1,\dots,v_n$ we have that

$$\|u_0\| \geq \|x\|; \quad \|v_i\| = \|u_{i-1}\|,\ \|u_i\| \geq \|v_i\|,\ i = 1,\dots,n; \quad \|u_n\| \leq \sqrt{1 + \tan^2\frac{\pi}{2^{n+2}}}.$$

Hence P is contained in the disc with radius $\sqrt{1 + \tan^2\frac{\pi}{2^{n+2}}} = \frac{1}{\cos\frac{\pi}{2^{n+2}}}$. Let φ_i, ξ_i be the arguments of u_i, v_i, respectively.

If a point $x \in P$ has the maximal norm $\frac{1}{\cos\frac{\pi}{2^{n+2}}}$, then all vectors u_i, v_i must have this length. It follows that $u_n = (1, \tan\frac{\pi}{2^{n+2}})$ and the argument of u_n is $\varphi_n = \frac{\pi}{2^{n+2}}$. Depending on whether $\varphi_i = \xi_i$ or $\varphi_i = -\xi_i$ for $i = 1,\dots,n$, the angle φ_0 of u_0 takes values $\frac{\pi}{2^{n+2}},\dots,\frac{\pi}{2} - \frac{\pi}{2^{n+2}}$ on a regular grid with step length $\frac{\pi}{2^{n+1}}$. But then the argument of x can take values $\frac{\pi}{2^{n+2}},\dots,2\pi - \frac{\pi}{2^{n+2}}$ on a regular grid with step length $\frac{\pi}{2^{n+1}}$.

There are hence vertices of the polyhedron P given by

$$w_k = \frac{1}{\cos\frac{\pi}{2^{n+2}}} \begin{pmatrix} \cos\left(\frac{k\pi}{2^{n+1}} - \frac{\pi}{2^{n+2}}\right) \\ \sin\left(\frac{k\pi}{2^{n+1}} - \frac{\pi}{2^{n+2}}\right) \end{pmatrix}, \quad k = 1,\dots,2^{n+2}.$$

It is directly checked that there are no vertices with strictly lower norm.

The supporting linear functionals are given by

$$L_l(x) = 1 - \cos\frac{l\pi}{2^{n+1}} x_1 - \sin\frac{l\pi}{2^{n+1}} x_2, \quad l = 1,\dots,2^{n+2}.$$

Hence the slack matrix is of size $2^{n+2} \times 2^{n+2}$ and has elements

$$S_{kl} = 1 - \frac{1}{\cos\frac{\pi}{2^{n+2}}} \left(\cos\frac{l\pi}{2^{n+1}} \cos\left(\frac{k\pi}{2^{n+1}} - \frac{\pi}{2^{n+2}}\right) + \sin\frac{l\pi}{2^{n+1}} \sin\left(\frac{k\pi}{2^{n+1}} - \frac{\pi}{2^{n+2}}\right) \right) = 1 - \frac{\cos\left(\frac{(l-k)\pi}{2^{n+1}} + \frac{\pi}{2^{n+2}}\right)}{\cos\frac{\pi}{2^{n+2}}}.$$

On the other hand, the polyhedron P is given by a lifting with $2n+6$ inequalities. This corresponds to a nonnegative factorization of $S = FG^T$ into factors F, G of size $2^{n+2} \times (2n+6)$.

7.10 Application: resource allocation

Suppose we may fabricate a number of products which we can sell at prices $p_1, \dots, p_n$, respectively. The production of a unit of product l consumes a_{kl} units of raw material k, $k = 1, \dots, K$, of which a total quantity of r_k units is available. We wish to choose the quantities $x_1, \dots, x_n$ of each product to be produced in order to maximize the revenue.

The problem can be formalized as the LP

$$\min_x -\langle p, x\rangle : \qquad Ax \le r,\ x \ge 0,$$

where A is the $K \times n$ matrix made up of the coefficients a_{kl}, and x, r, p are the vectors made up of the corresponding elements. The constraint $x \ge 0$ is necessary to prevent the conversion of products back to raw materials, which would correspond to a negative quantity x_l.

If all or some of the products can only be produced in integer numbers, the problem becomes a MILP.

7.11 Application: sparse recovery

Suppose we observe a noisy linear image of a *sparse* vector x,

$$y = Ax + \xi,$$

where $A \in \mathbb{R}^{m\times n}$ encodes the linear map, and ξ is a noise term bounded by a constant δ by absolute value. Here we suppose that $m \ll n$, i.e., the number of observations is smaller than the dimension of the vector x.

Our goal is to recover the vector x. Obviously, even in the absence of noise, the linear system $Ax = y$ is underdetermined. We hence cannot just employ linear regression, but we have to use somehow the information that the vector x is sparse. Ideally, we should hence solve the problem

$$\min_x \|x\|_0 : \qquad \|Ax - y\|_\infty \le \delta,$$

where $\|x\|_0$ denotes the number of non-zero components of x.

This is a highly non-convex difficult problem. However, we may relax this problem by replacing the 0-"norm" by the 1-norm, obtaining the convex problem

$$\min_x \|x\|_1 : \qquad \|Ax - y\|_\infty \le \delta.$$

This problem in turn can be rewritten as

$$\min_{x,t} \langle \mathbf{1}, t\rangle : \qquad -t \le x \le t, \quad -\delta \le Ax - y \le \delta.$$

Obviously, this is a linear program.

References

[1] Stephen J. Wright. *Primal-dual interior-point methods.* Other Titles in Applied Mathematics. SIAM, 1997.

[2] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. Syst. Sci.*, 43:441–466, 1991.