Chapter 2
Geometry of the Feasible Region

Feasible region is of a basic significance to optimization. Due do the linear structure of the standard LP problem (1.8), the feasible region P, defined by (1.9), has special features. In this chapter, we explore this theme from a geometric point of view.

Denote the n-dimensional Euclidean space by $\mathcal{R}^n$, as usual. Points or column vectors are the basic geometrical elements. Denote a point, or the vector from the origin to this point by

$$x = (x_1, \ldots, x_n)^T \in \mathcal{R}^n,$$

whose components (coordinates) are $x_1, \ldots, x_n$. Thereafter, points and vectors will not be distinguished. Denote by $e_j, \ j = 1, \ldots, n$ the jth coordinate vector, i.e., the unit vector with its jth component 1, and denote by e the vector of all ones. The reader is referred to related literature for basic concepts and operations in Euclidean space, such as linear dependency and independency, set of points and its boundedness and unboundedness, the inner product $x^T y$ of vectors $x, y \in \mathcal{R}^n$, $\cos < x, y >= x^T y / (\|x\| \|y\|)$ of the angle between them, their orthogonality $x \perp y$ or $x^T y = 0$, i.e., $< x, y >= \pi/2$, the Euclidean module or norm $\|x\| = \sqrt{x^T x}$ of vector x, and so on.

Involved in the standard LP problem, both c and x may be viewed as vectors in $\mathcal{R}^n$, columns $a_j, \ j = 1, \ldots, n$ of A as vectors in $\mathcal{R}^m$, b as a vector in $\mathcal{R}^m$, and the feasible region P as a closed polyhedral in $\mathcal{R}^n$ in general (as will be clear a little later), though P could be degenerate, or even empty. The following lemma, proposed by Farkas (1902), renders a sufficient and necessary condition for nonempty P (the proof is delayed to at the end of Sect. 4.2).

Lemma 2.1 (Farkas). *Assume $A \in \mathcal{R}^{m \times n}$ and $b \in \mathcal{R}^m$. The feasible region P is nonempty if and only if*

$$b^T y \geq 0, \qquad \forall \ y \in \{y \in \mathcal{R}^m \mid A^T y \geq 0\}.$$

Fig. 2.1 A geometrical
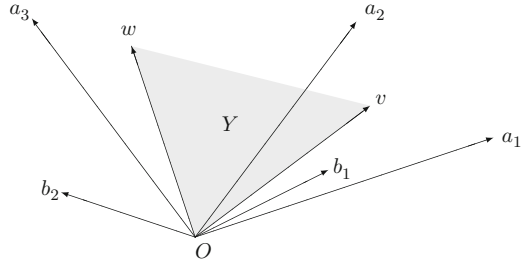explanation for Farkas lemma

Fig. 2.1 serves as a geometric explanation for the preceding Lemma, where
a_1, a_2, a_3 are columns of $A \in \mathcal{R}^{2 \times 3}$. Y is the set of all vectors which forms with
every column of A an angle no more than $\pi/2$ (corresponding to the shaded area
between vectors v and w in the figure). b_1 forms with each $y \in Y$ an angle no more
than $\pi/2$ but b_2 does not. Thereby, P is nonempty when $b = b_1$, whereas empty
when $b = b_2$.

In addition to rank $A = m$, discussions in this chapter will be based on the
following.

Assumption. The feasible region P is nonempty and infinite.

2.1   Polyhedral Convex Set and the Feasible Region

For any given two points $x, y \in \mathcal{R}^n$, set

$$S = \{\alpha x + (1 - \alpha)y \mid \alpha \in \mathcal{R}\}$$

is a *straight line*; if $0 < \alpha < 1$, it is an *open segment* with end points x and y,
denoted by (x, y); if $0 \le \alpha \le 1$, it is a *closed segment*, denoted by $[x, y]$. Hereafter
so-called "segment" will be all closed.

Definition 2.1.1. Π is an affine set if, whenever it includes any two points, it
includes the whole straight line passing through them. The smallest affine set
including a set is the affine hull of the latter.

Straight lines in $\mathcal{R}^2$ and planes in $\mathcal{R}^3$ are instances of affine sets. The whole space
$\mathcal{R}^n$ is an affine set. An empty set and a single point set are viewed as affine sets. It
is clear that the intersection of affine sets is an affine set.

For any given α_i, $i = 1, \ldots, k$ satisfying $\sum_{i=1}^{k} \alpha_i = 1$, point

$$x = \sum_{i=1}^{k} \alpha_i x^i$$

is called an *affine combination* of $x^1, \ldots, x^k$. It is easy to show that the set of all such affine combinations, i.e.,

$$\left\{ \sum_{i=1}^{k} \alpha_k x^i \mid \sum_{i=1}^{k} \alpha_i = 1, \ \alpha_i \in \mathcal{R}, \ i = 1, \ldots, k \right\}$$

is an affine set, called *affine hull* of these points. The two points in Definition 2.1.1 can be generalized to multiple points: it is easy to show that Π is an affine set if and only if the affine hull of any finitely many points within Π belongs to Π.

Set L is a *subspace* of $\mathcal{R}^n$ if it is closed for all linear operations, that is, for any $x, y \in L$ and $\alpha, \beta \in \mathcal{R}$ it holds that $\alpha x + \beta y \in L$. An affine set is an affine subspace if it is a subspace.

Theorem 2.1.1. *An affine set is an affine subspace if and only if it includes the origin.*

Proof. The necessity is clear. Sufficiency. Let Π be an affine set including the origin. Then for any $x \in \Pi$ and $\alpha \in \mathcal{R}$, it holds that

$$\alpha x = \alpha x + (1 - \alpha)0 \in \Pi.$$

On the other hand, it holds for any $x, y \in \Pi$ that

$$\frac{x + y}{2} = \frac{1}{2}x + (1 - \frac{1}{2})y \in \Pi.$$

Therefore, Π is closed for linear operations, and is thus an affine subspace.   □

Theorem 2.1.2. *For any nonempty affine set Π, there exists vector p so that*

$$L = \{x + p \mid x \in \Pi\}$$

is an affine subspace, and such subspace is unique.

Proof. According to Theorem 2.1.1, Π is an affine subspace if $0 \in \Pi$. Note that $L = \Pi$ corresponds to $p = 0$. Assume that $0 \notin \Pi$. Since $\Pi \neq \emptyset$, there exists $0 \neq y \in \Pi$. Letting $p = -y$, it is clear that

$$L = \{x + p \mid x \in \Pi\}$$

is an affine set including the origin, and is hence an affine subspace.

Now let us show the uniqueness. Assume that L_1, L_2 are affine subspaces such that

$$L_1 = \{y + p_1 \mid y \in \Pi\}, \qquad L_2 = \{y + p_2 \mid y \in \Pi\}.$$

It is clear that

$$\Pi = \{x - p_1 \mid x \in L_1\}.$$

If $p = -p_1 + p_2$, therefore, it holds that

$$L_2 = \{x + p \mid x \in L_1\},$$

from which it follows that $x + p \in L_2$ for any $x \in L_1$. Since $0 \in L_1$, $p \in L_2$ holds. Further, since L_2 is a subspace, $x = (x + p) - p \in L_2$ holds. Therefore, $L_1 \subset L_2$. $L_2 \subset L_1$ can be similarly derived. So it can be asserted that $L_1 = L_2$. $\square$

Geometrically, the affine subspace L may be viewed as a parallelism of affine set Π along vector p. It is therefore called *parallel subspace* of Π. The dimension of L is said to be that of Π, which is equal to the number of independent components (coordinates) of elements in Π. It is clear that an affine set with one or more than one dimension is unbounded.

Let a be a nonzero vector and let η be a real number. Set

$$H = \{x \in \mathcal{R}^n \mid a^T x = \eta\}$$

is called *superplane*, whose normal vector is a $(a \perp H)$; in fact, for any two points $x, y \in H$, it holds that

$$a^T (x - y) = a^T x - a^T y = \eta - \eta = 0.$$

It is easy to show that any superplane is an affine set.

Any straight line in $\mathcal{R}^2$ and any plane in $\mathcal{R}^3$ are instances of superplane.

The "signed" distance from any point $\bar{x}$ to superplane H is defined by $r/\|a\|$, where r is the residual

$$r = a^T \bar{x} - \eta.$$

If $r = 0$, point $\bar{x}$ is within H. It might be well to assume $\eta > 0$. Then, if $r < 0$, the origin and $\bar{x}$ are in the same side of H; if $r > 0$, the two points are in different sides of H.

Superplanes associated with the objective function are of significance to LP. Regarding objective value f as a parameter, the sets

$$H(f) = \{x \in \mathcal{R}^n \mid c^T x = f\}$$

are a family of contour surfaces of the objective function. The gradient $\nabla f = c$ of the objective function is the common normal vector of all the contour surfaces, pointing to the increasing side of objective value f.

The following gives a mathematical expression of an affine set.

Theorem 2.1.3. *Set Π is an affine set if and only if there exists $W \in \mathcal{R}^{k \times n}$ and $h \in \mathcal{R}^k$ such that*

$$\Pi = \{x \mid Wx = h\}. \tag{2.1}$$

Proof. It might be well to rule out the trivial case when Π is empty or the whole space.

Sufficiency. Let Π is defined by (2.1). For any $x, y \in \Pi$ and $\alpha \in R^1$, it holds that

$$W(\alpha x + (1 - \alpha)y) = \alpha Wx + (1 - \alpha)Wy = \alpha h + (1 - \alpha)h = h,$$

leading to $\alpha x + (1 - \alpha)y \in \Pi$. Therefore, Π is an affine set.

Necessity. Let Π be an affine set. Assume that L is a parallel affine subspace, and $w^1, \ldots, w^k$ are basis of the orthogonal complementary space of it. Then, it follows that

$$L = \{y \mid (w^i)^T y = 0, \ i = 1, \ldots, k\} \triangleq \{y \mid Wy = 0\},$$

where rows of $W \in \mathcal{R}^{k \times n}$ are $(w^1)^T, \ldots, (w^k)^T$. Introduce notation $h = Wp$. Since L is a parallel subspace of Π, there exists a vector p such that

$$L = \{x - p \mid x \in \Pi\}.$$

Thus, for any $x \in \Pi$, it holds that $x - p \in L$; and $W(x - p) = 0$ means that $x \in \{x \mid Wx = h\}$. If $x \in \{x \mid Wx = h\}$, conversely, then $Wx - Wp = W(x - p) = 0$, hence from $x - p \in L$ it follows that $x \in \Pi$. So, Π has expression (2.1). $\square$

The preceding Theorem says that a set is affine set if and only if it is the intersection of finitely many superplanes. It is easy to show that the dimension of the affine set is equal to $n - \text{rank}(W)$. In particular, the solution set

$$\Delta = \{x \mid Ax = b\}$$

of the constraint system of the standard LP problem is an affine set. Since rank $A = m$, $\Delta \neq \emptyset$ and dim $\Delta = n - m$.

Definition 2.1.2. C is a convex set if it includes the whole segment whenever it includes its two end points. The smallest convex set including a set is the convex hull of the latter.

Any disks and the first quadrant in $\mathcal{R}^2$ and spheres in $\mathcal{R}^3$ are instances of convex sets. Clearly, segments and the whole space are convex sets too. Empty sets and single point sets are regarded as convex sets. Intersections of convex sets are convex. Any affine set is convex, but a convex set is not an affine set in general; e.g., disks

and spheres are not affine sets. It is clear that any convex set has an affine hull. The dimension of the latter is said to be the dimension of the former. Hereafter, so-called "convex set" is a closed convex set.

For any $\alpha_i \geq 0$, $i = 1, \ldots, k$ satisfying $\sum_{i=1} \alpha_i = 1$, the point

$$x = \sum_{i=1}^{k} \alpha_k x^i$$

is called *convex combination* of points $x^1, \ldots, x^k$. It is easy to show that the set of all such convex combinations, i.e.,

$$\left\{ \sum_{i=1}^{k} \alpha_k x^i \;\middle|\; \sum_{i=1}^{k} \alpha_i = 1; \alpha_i \geq 0, \; i = 1, \ldots, k \right\}$$

is the convex hull of these points. It is easy to show that C is a convex set if and only if the convex hull of any finitely many points within C belongs C.

Any superplane H divides the whole space to two *closed half spaces*, i.e.,

$$H_L = \{x \mid a^T x \leq \eta\}, \quad \text{and} \quad H_R = \{x \mid a^T x \geq \eta\}. \tag{2.2}$$

The intersection of infinitely many closed half spaces is called *polyhedral*, and bounded polyhedral called *polyhedron*. It could degenerate to a segment or point, or even empty set.

It is clear that a half space is convex. Therefore, a polyhedral or polyhedron is convex, as termed *polyhedral convex set*.

A convex set C is called *polyhedral convex cone* if $\alpha x \in C$ holds for any $x \in C$ and $\alpha \geq 0$. It is easy to show that a set is a polyhedral convex cone if and if it is the intersection of finitely many closed half spaces passing through the origin, as expressed $\{x \in \mathcal{R}^n \mid Ax \geq 0\}$.

The nonnegative constraints $x \geq 0$ in the standard problem correspond to the positive octant, which is the intersection of the n closed half spaces with the coordinate planes as its boundary. Therefore, the feasible region P is the intersection of the affine set Δ and the positive octant. As any superplane $a^T x = \eta$ may be viewed as the intersection of two closed half spaces (2.2), P may be also viewed as the intersection of finitely many closed half spaces. Therefore, we make the following statement.

Proposition 2.1.1. *The feasible region P is a polyhedral convex set.*

Such a set could be degenerate, however. The following result concerns the dimension of P.

Proposition 2.1.2. *Define index set*

$$J' = \{j \in A \mid x_j = \mu_j, \; x \in P\}, \tag{2.3}$$

where μ_j, $j \in A$ are nonnegative constants and denote by $I_{J'}$ the coefficient matrix of system $x_j = \mu_j$, $j \in J'$. Then it holds that

$$n - \min\{m + |J'|, n\} \le \dim P = n - r \le n - \max\{m, |J'|\}, \qquad (2.4)$$

where

$$r = \text{rank} \begin{pmatrix} A \\ I_{J'} \end{pmatrix}.$$

Proof. Note that rank $A = m$ and P is a nonempty infinite set, according to the basic assumption.

It is clear that $|J'| \le r$ and

$$P = \{x \in \mathcal{R}^n \mid Ax = b, x \ge 0; x_j = \mu_j, j \in J'\}.$$

IF $|J'| = r$, then $x_j = \mu_j$, $j \in J'$ is a canonical form of system

$$Ax = b; \qquad x_j = \mu_j, \quad j \in J'.$$

If $|J'| < r$, besides x_j $j \in J'$, there are additional $r - |J'|$ basic variables, and hence a canonical form of the preceding system. It is to say that there exists a canonical form, whose $n - r$ nonbasic variables do not belong to J'. Therefore, it holds that $\dim P = n - r$. Hence (2.4) follows from $\min\{m + |J'|, n\} \ge r \ge \max\{m, |J'|\}$. $\square$

In particular, $\dim P = n - m$ if $J' = \emptyset$.

The special case of $\mu_j = 0$ is of significance to the standard LP problem. Introduce sets

$$J = \{j \in A \mid x_j = 0, x \in P\}, \qquad \bar{J} = A \backslash J, \qquad (2.5)$$

where J is said to be *index set of zero components*.

According to the preceding definition, it is clear that $P \subset P'$, and in addition

$$x_j = 0, \qquad \forall j \in J, x \in P. \qquad (2.6)$$

If $x \in P$ and $x_j > 0$, then $j \in \bar{J}$.

If $J \ne \emptyset$, the feasible region P has no interior point in the normal sense, as is often the case for real problems. For convenience of applications, the following concept is introduced instead.

Definition 2.1.3. Assume $\bar{x} \in P$. If there exists $\delta > 0$ such that a neighborhood of $\bar{x}$ is included in P, i.e.,

$$\Omega(\delta) = \{x \in \Delta \mid x_j = 0, j \in J; \|x - \bar{x}\| < \delta\} \subset P, \qquad (2.7)$$

then $\bar{x}$ is an *interior point* of P; otherwise, it is a *boundary point*.

The set of all interior points of P is called its *interior*, denoted by int P. P and its interior have the same dimension.

For sake of distinguishing, the point is said to be *relative interior point* when $J \neq \emptyset$. The set of relative interior points is *relative interior*, while *(strict) interior point* or *interior* stands for the case of $J = \emptyset$.

Theorem 2.1.4. *Assume* $\bar{x} \in P$. *Then* $\bar{x} \in$ int P *if and only if*

$$\bar{x}_j > 0, \qquad j \in \bar{J}. \tag{2.8}$$

Proof. Sufficiency. Let point $\bar{x} \in P$ satisfy (2.8). Using

$$\delta = \min_{j \in \bar{J}} \bar{x}_j > 0, \tag{2.9}$$

and

$$\Omega(\delta) = \{x \in \Delta \mid x_j = 0, \ j \in J; \ \|x - \bar{x}\| < \delta\},$$

then for any $x \in \Omega(\delta)$, it holds that

$$x \in \Delta \qquad \text{and} \qquad x_j = 0, \ j \in J.$$

Moreover, sine

$$\|x - \bar{x}\| = \sqrt{\sum_{j=1}^{n} (x_j - \bar{x}_j)^2} < \delta,$$

it holds that

$$|\bar{x}_j - x_j| \leq \|x - \bar{x}\| < \delta, \qquad j \in \bar{J},$$

which and (2.9) together give

$$x_j > \bar{x}_j - \delta \geq 0, \qquad j \in \bar{J}.$$

Thus $x \in P$. Therefore $\Omega(\delta) \subset P$, and $\bar{x}$ is an interior point of P.

Necessity. Assuming $\bar{x} \in$ int P, then there is $\delta > 0$ such that $\Omega(\delta) \subset P$; also there is $p \in \bar{J}$ such that $\bar{x}_p = 0$. According to the definition of $\bar{J}$, there is $x' \in P$ such that $x'_p > 0$. It is clear that for any $\alpha > 0$ there exists

$$x = -\alpha x' + (1 + \alpha)\bar{x} = \bar{x} + \alpha(\bar{x} - x') \in \Delta. \tag{2.10}$$

Hence, when α is sufficiently small, it holds that

$$\|x - \bar{x}\| = \alpha\|\bar{x} - x'\| < \delta.$$

In addition, it is clear that x'_j, $\bar{x}_j = 0$, $j \in J$, and hence $x_j = 0$, $j \in J$. Therefore $x \in \Omega(\delta)$. On the other hand, from (2.10), $\bar{x}_p = 0$ and $\alpha > 0$, $x'_p > 0$ it follows that

$$x_p = -\alpha x'_p + (1 + \alpha)\bar{x}_p = -\alpha x'_p < 0,$$

which contradicts $\Omega(\delta) \subset P$. Therefore, (2.8) holds if $\bar{x} \in \text{int } P$. $\square$

Proposition 2.1.3. *If* $\dim P \geq 1$, *then* P *has a relative interior point.*

Proof. The assumption of the proposition implies $\bar{J} \neq \emptyset$, because otherwise it holds that $J = A$ and hence $\dim P = 0$, leading to contradiction. On the other hand, for any $j \in \bar{J}$ there exists $x \in P$ such that $x_j > 0$; hence from the convexity of P, it follows that there is $x \in P$ satisfying $x_j > 0$, $j \in \bar{J}$. According to Theorem 2.1.4, it is known that $x \in \text{int } P \neq \emptyset$. $\square$

Example 2.1.1. Investigate the interior of the feasible region of the following problem:

$$\begin{aligned}
\min \ & x_1 + x_2 + x_3 + x_4 + x_5, \\
\text{s.t.} \ \ & x_1 \quad\quad - \ x_3 + x_4 + x_5 = 6, \\
& x_2 - \ x_3 - x_4 - x_5 = 0, \\
& -x_2 + 3x_3 + x_4 + x_5 = 0, \\
& x_j \geq 0, \ j = 1, \ldots, 5.
\end{aligned}$$

Answer   Adding the second constraint equality to the third gives

$$2x_3 = 0.$$

It is clear that the feasible region is nonempty, and the x_3 component of all feasible points equals 0. Eliminating x_3 from the problem leads to

$$\begin{aligned}
\min \ & x_1 + x_2 + x_4 + x_5, \\
\text{s.t.} \ \ & x_1 \quad + \ x_4 + x_5 = 6, \\
& x_2 - x_4 - x_5 = 0, \\
& x_j \geq 0, \ j = 1, 2, 4, 5,
\end{aligned}$$

the interior of the feasible region of which is clearly nonempty, corresponding to the relative interior of the feasible region of the original problem.

2.2   Geometric Structure of the Feasible Region

This section will discuss the geometric structure of the feasible region P. Actually, most results in this section are also valid for general polyhedral convex sets.

Definition 2.2.1. Let P' be a nonempty convex subset of P. It is a face of P if for any $x \in P'$, satisfying $x \in (y, z) \subset P$, it holds that $y, z \subset P'$.

The preceding means that a face includes the whole segment if it includes a interior point of any segment of P.

The following renders a mathematical expression of P.

Theorem 2.2.1. *A nonempty convex subset $P(Q)$ of P is its face if and only if there exists index set $Q \subset A$ such that*

$$P(Q) = \{x \in \mathcal{R}^n \mid Ax = b, \ x \geq 0; \ x_j = 0, \ j \in Q\}. \tag{2.11}$$

Proof. Sufficiency. Let $\emptyset \neq P(Q) \subset P$ be defined by (2.11). If $v \in P(Q)$ is an interior point of segment (y, z) and $y, z \in P$, then there exists $0 < \alpha < 1$ such that

$$v = \alpha y + (1 - \alpha)z.$$

Hence from $\alpha > 0, \ 1 - \alpha > 0, \ y, z \geq 0$ and

$$v_j = \alpha y_j + (1 - \alpha)z_j = 0, \qquad j \in Q,$$

it follows that

$$y_j, z_j = 0, \qquad j \in Q,$$

Therefore, $y, z \in P(Q)$, and hence $P(Q)$ is a face of P.

Necessity. Let $P(Q) \neq \emptyset$ be a face of P. Introduce notation

$$Q = \{j \in A \mid x_j = 0, \ x \in P(Q)\}. \tag{2.12}$$

Now we show that $P(Q)$ is equivalent to

$$P(Q)' = \{x \in \mathcal{R}^n \mid Ax = b, \ x \geq 0; \ x_j = 0, \ j \in Q\}. \tag{2.13}$$

It is clear that $P(Q) \subset P(Q)' \subset P$.

If $P(Q)$ includes the origin 0 only, it follows that $b = 0$ and $Q = A$. And $P(Q)'$ clearly includes 0, hence $P(Q) = P(Q)'$. Now assuming that $P(Q)$ does not include 0, we will show $P(Q)' \subset P(Q)$.

Assume $x \in P(Q)'$. If $x = 0$ ($b = 0$), then for any $0 \neq v \in P(Q)$ and

$$y = 2v$$

it holds that

$$v = \frac{y}{2} + \frac{0}{2},$$

which means that $v \in (y, 0)$. Since $P(Q)$ is a face and $y, 0 \in P$, we have $x = 0 \in P(Q)$. On the other hand, if $x \neq 0$, i.e.,

$$S = \{j \in A \mid x_j > 0\}$$

is nonempty, then from $x \in P(Q)'$ and (2.13), it is known that $j \notin Q$ for any $j \in S$. Therefore there exists $u \in P(Q)$ such that $u_j > 0$. If S includes two or more indices, then $u, w \in P(Q)$ satisfy $u_i, w_j > 0$ for any $i, j \in S$, hence $z = u/2 + w/2 \in P(Q)$ satisfies $z_i, z_j > 0$. This means that there exists $v \in P(Q)$ such that $v_j > 0$ for all $j \in S$. As for the relation between x and v, there are the following two cases only:

(i) $\{j \in S \mid x_j > v_j\} = \emptyset$. It is clear that

$$z = 2v - x = v + (v - x) \in P.$$

Since $P(Q)$ is a face of P, and $v = z/2 + x/2 \in (x, z)$, it holds that $x \in P(Q)$.

(ii) $\{j \in S \mid x_j > z_j\} \neq \emptyset$. Define

$$z = x + \beta(v - x), \qquad \beta = \min\{x_j/(x_j - v_j) \mid x_j - v_j > 0, j \in S\} > 1.$$

It is easy to verify that $z \in P$, and

$$v = \alpha z + (1 - \alpha)x \in (x, z), \qquad 0 < \alpha = 1/\beta < 1.$$

In addition, $P(Q)$ is a face of P, hence $x \in P(Q)$. Therefore $P(Q)' \subset P(Q)$.
□

Clearly, P itself is a face ($Q = \emptyset$). If a face $P(Q) \neq P$, it is called a *proper face*. It is easy to show that face $P(Q)$ is a proper face if and only if $\dim P(Q) < \dim P$. From the proof of Proposition 2.1.2, it is know that $\dim P(Q) \leq n - \max\{m, |Q|\}$. Face of face is a face.

If $\dim P(Q) = \dim P - 1$, face $P(Q)$ is called *facet* of P. An 1-dimensional face is also called *edge*; an 0-dimensional face is called *vertex* or *extreme point*.

It is clear that the feasible region P has infinitely many faces. In fact, it is known that the number of $(n - m - k)$-dimensional faces of P is no more than C_n^k ($k = 1, \ldots, n - m$); in particular, there exist, at most, an $(n - m)$-dimensional face (that is P itself), C_n^{m+1} edges and $C_n^{n-m} = C_n^m$ vertices.

"Vertex" can also be defined by the following alternatively.

Definition 2.2.2. x is a vertex of P if $x \in [y, z]$ leads to $x = y$ or $x = z$ for any $y, z \in P$.

The preceding implies that a vertex is not an interior point of any segment of P. It is clear that a vertex of face is a vertex of P, and that the origin is a vertex if it belongs to P.

Vertex has its distinctive algebraical attribute.

Lemma 2.2.1. *A point $x \in P$ is a vertex if and only if columns of A corresponding to its positive components are linearly independent.*

Proof. It might be well to let the first s components of x be great than 0 and let the rest be 0. Assume that $\bar{x}$ is the subvector consisting of the first s components of x and $\bar{A}$ is the submatrix consisting of the first s columns of A. Then $\bar{A}\bar{x} = b$.

Necessity. Let x be a vertex of P. If columns of $\bar{A}$ are linearly dependent, there is a nonzero vector $\bar{v}$ such that $\bar{A}\bar{v} = 0$. Introduce notation

$$\bar{y} = \bar{x} + \alpha\bar{v}, \qquad \bar{z} = \bar{x} - \alpha\bar{v}.$$

It is clear that for any real α it holds that

$$\bar{A}\bar{y} = \bar{A}\bar{z} = b.$$

Take sufficiently small $\alpha > 0$, such that $\bar{y}, \bar{z} \geq 0$. Construct vectors y and z, so that the first s components of them respectively constitute $\bar{y}$ and $\bar{z}$, and the others are 0. Then it is clear that $y, z \in P$ and $x = y/2 + z/2$. Thus x is not a vertex of P, as leads to a contradiction. Therefore, columns corresponding to all positive components of x are linear independent if x is a vertex of P.

Sufficiency. Assume that columns of $\bar{A}$ are linearly independent. If $x \in P$ is not a vertex, there are two points $y, z \in P$ and a real $\alpha \in (0, 1)$ such that

$$x = \alpha y + (1 - \alpha)z,$$

from which it is known that the last $n - s$ components of y and z are both 0. Therefore, $v = x - y \neq 0$ and

$$\bar{A}\bar{v} = Av = Ax - Ay = b - b = 0,$$

which means that columns of $\bar{A}$ are linearly dependent, as a contradiction. Therefore x is a vertex if columns of $\bar{A}$ are linearly independent. $\square$

In view of rank$(A) = m$, the preceding theorem implies that the number of positive components of a vertex of P is no more than m.

Lemma 2.2.2. *$\bar{x}$ is a vertex of the feasible region if and only if it is a basic feasible solution.*

Proof. It is from Theorem 1.6.2 and Lemma 2.2.1. $\square$

The preceding Lemma says that vertex and basic feasible solution of P are the same; thus, the two may be regarded as geometrical and algebraical names of the same mathematical item. It is usually difficult to determine if a point is a vertex based on the definition itself, while a basic feasible solution can be conveniently determined algebraically. Recall that every canonical form of the constraint system $Ax = b$ corresponds to a basic solution, which is a basic feasible solution if it is nonnegative (Sect. 1.6).

Lemma 2.2.3. *Nonempty feasible region has a vertex.*

Proof. It is clearly the case when P is a single point set. Let P be infinite set, and $\bar{x} \in P$. If $\bar{x}$ is not a vertex, there are two distinct points $y, z \in P$ and a real $\bar{\alpha} \in (0, 1)$ such that

$$\bar{x} = \bar{\alpha}y + (1 - \bar{\alpha})z = z + \bar{\alpha}(y - z).$$

Thus, a component of $\bar{x}$ is 0 if and only if the corresponding components of both y, z are 0. Introduce

$$T = \{j \in A \mid \bar{x}_j > 0\}.$$

Then $T \neq \emptyset$, since $T = \emptyset$ implies that $\bar{x} = 0$ is a vertex. It might be well to assume that for some $i \in \{1, \dots, n\}$ it holds that $z_i > y_i$, and hence $i \in T$. This means that

$$\{j \in T \mid z_j - y_j > 0\} \neq \emptyset.$$

It is easy to show that redefined

$$\bar{x} = \alpha_1 y + (1 - \alpha_1)z, \quad \alpha_1 = z_q/(z_q - y_q) = \min\{z_j/(z_j - y_j) \mid z_j - y_j > 0, \ j \in T\}$$

satisfies $\bar{x} \in P$ and $\bar{x}_q = 0$. Thus $|T|$ for the new $\bar{x}$ is less than that for the old by 1, at least. Repeating no more than n times, therefore, the preceding process terminates at a vertex.   □

The above proof produces a series of feasible points, corresponding to faces, each of which is a proper face of its predecessor, until reaching a 0-dimensional face (vertex). Such a technique for shifting to faces of lower dimensions will often be used.

For any given point x and vector $d \neq 0$, set $\{x + \alpha d \mid \alpha \geq 0\}$ is said to be *ray* (*or half-line*), emanating from x along the direction of d. It is clear that a ray is an infinite set.

Definition 2.2.3. A nonzero vector d is an unbounded direction if P includes rays, emanating from all $x \in P$ along d.

Two unbounded directions having the same direction are regarded as the same.

Theorem 2.2.2. *Vector d is an unbounded direction of the nonempty feasible region if and only if*

$$Ad = 0, \qquad d \neq 0, \qquad d \geq 0. \tag{2.14}$$

Proof. Sufficiency. With the assumptions, it is easy to verify that for any given $x \in P$ and $\alpha \geq 0$ it holds that $x + \alpha d \in P$, therefore $d \neq 0$ is an unbounded direction.

Necessity. Let $d \neq 0$ be an unbounded direction. Thus, there is $x \in P$, satisfying $x + \alpha d \in P$ for any $\alpha \geq 0$. Hence, from $Ax = b$ and $A(x + \alpha d) = b$, it follows that $Ad = 0$. In addition, $d \geq 0$ holds, because, otherwise, d has a negative component, and hence the corresponding component of $x + \alpha d$ is negative whenever $\alpha > 0$ becomes sufficiently large, as contradicts $x + \alpha d \in P$. $\square$

Corollary 2.2.1. *A nonzero vector d is a unbounded direction if P includes the ray, emanating from some $x \in P$ along d.*

It is clear that any nonnegative linear combination of finitely many unbounded directions is an unbounded direction if the combination coefficients are not all zero. Note that "unbounded direction" is meaningless to an empty P.

Theorem 2.2.3. *The feasible region is unbounded if and only if it has an unbounded direction.*

Proof. Sufficiency is clear, it is only needed to show necessity.

If $v \in P$, then the translation of P, i.e.,

$$C = \{x - v \mid x \in P\}$$

clearly includes the origin, and P is unbounded if and only if C is unbounded. It might be well to assume that $0 \in P$.

Let $S' = \{x^k\} \in P$ be an unbounded sequence of points. Without loss of generality, assume that

$$\|x^k\| \rightarrow \infty \quad \text{as} \quad k \rightarrow \infty.$$

Then the sequence

$$S'' = \{x^k / \|x^k\|\}$$

on the unit sphere is bounded, hence has a cluster point. Letting x be its cluster point, then S'' includes a subsequence converging to x. It might be well to assume

$$x^k / \|x^k\| \rightarrow x.$$

Now it should be shown that x is an unbounded direction of P. Let M be a given positive number. Since $\|x^k\| \rightarrow \infty$, there is a positive integer K such that $\|x^k\| > M$ or $M/\|x^k\| < 1$ when $k \geq K$. Introduce

Fig. 2.2 d^1, d^2, d^3 are
unbounded directions. d^1, d^2
are extreme directions, but d^3
not
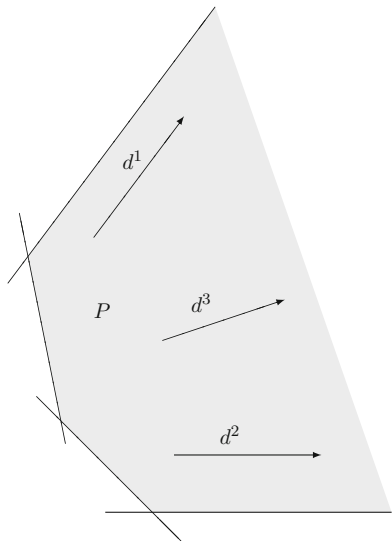
$$y^k = (M/\|x^k\|)x^k, \qquad k = K, K + 1, \ldots.$$

Since P is convex, we have $0, x^k \in P$ and $y^k \in (0, x^k)$, and hence $y^k \in P$ and $y^k \to Mx$ when $k \geq K$. As P is closed, it can be asserted that $Mx \in P$.   □

Definition 2.2.4. An unbounded direction of P is extreme direction if it can not be expressed by a positive linear combination of two distinct unbounded directions.

According to the preceding, that d is an extreme direction means that if there are unbounded directions d', d'' and positive number $\sigma_1, \sigma_2 > 0$, satisfying $d = \sigma_1 d' + \sigma_2 d''$, then there must be $d' = \sigma d'$, where $\sigma > 0$. Two extreme directions having the same direction are regarded as the same.

In Fig. 2.2, d^1, d^2, d^3 are unbounded directions of P. d^1 and d^2 are extreme directions, but d^3 is not.

Theorem 2.2.4. *An unbounded direction is extreme direction if and only if the rank of columns, corresponding to its positive components, is less than the number of columns by* 1.

Proof. It might be well to assume that k positive components of unbounded direction d correspond to the set of columns $a_1, \ldots, a_k$. The satisfaction of $Ad = 0$ implies that the columns are linearly dependent. Denote the rank of the set by is r, then it is clear that $r < k$. Without loss of generality, assume that the first r columns are linear independent. Introduce $B = (a_1, \ldots, a_r)$. It is clear that

$$\text{rank } B = r \leq \text{rank } A = m.$$

Note that $k \geq 2$, because if $k = 1$, otherwise, then from $Ad = 0$ it follows that $a_1 = 0$, as leads to a contradiction.

Necessity. Assume that d is an extreme direction, but $r \neq k - 1$, that is, $a_1, \ldots, a_{k-1}$ are linearly dependent. Thus, there exists point

$$y = (y_1, \ldots, y_{k-1}, 0, \ldots, 0)^T \neq 0$$

such that

$$Ay = \sum_{j=1}^{k-1} y_j a_j = 0.$$

Clearly, for sufficiently small $\delta > 0$, it holds that

$$0 \neq d' = d + \delta y \geq 0, \qquad 0 \neq d'' = d - \delta y \geq 0,$$

and $Ad' = Ad'' = 0$. Therefore, d', d'' are unbounded directions of P, and hence not of the same direction. But $d = (d' + d'')/2$, as contradicts that d is an extreme direction. It therefore holds that $r = k - 1$.

Sufficiency. Assume $r = k - 1$. If there exist unbounded directions d', d'' and $\sigma_1, \sigma_2 > 0$ such that

$$d = \sigma_1 d' + \sigma_2 d'',$$

then zero components of d clearly correspond to zero components of d' and d''. So, the last $n - k$ components of d' and d'' are all zero. In addition, since d', d'' are unbounded directions, it holds that $Ad' = Ad'' = 0$ (Theorem 2.2.2), and hence that

$$Bd'_B + d'_k a_k = 0, \qquad Bd''_B + d''_k a_k = 0.$$

Note that $d'_k, d''_k > 0$; because if $d'_k = 0$, otherwise, then $d'_B = 0$, and hence $d' = 0$, as is a contradiction. Premultiplying the two sides of the preceding two equalities by B^T gives

$$B^T Bd'_B + d'_k B^T a_k = 0, \qquad B^T Bd''_B + d''_k B^T a_k = 0,$$

from which it follows that

$$d'_B = -d'_k (B^T B)^{-1} B^T a_k, \qquad d''_B = -d''_k (B^T B)^{-1} B^T a_k.$$

Therefore, it holds that

$$d'' = (d''_k / d'_k) d',$$

which means that unbounded directions d' and d'' have the same direction, and hence d is an extreme direction. $\qquad\square$

The extreme direction and 1-dimensional face (edge) have close relationship.

Theorem 2.2.5. *A vector is an extreme direction if and only if it is the unbounded direction of a edge.*

Proof. Necessity. Assume that d is an extreme direction. Based on Theorem 2.2.4, assume that columns corresponding to its positive components are

$$a_1, \ldots, a_r, a_{m+1},$$

where the first $r \le m$ columns are linearly independent. Thus

$$d_1, \ldots, d_r, d_{m+1} > 0, \qquad d_{r+1}, \ldots, d_m, d_{m+2}, \ldots, d_n = 0. \qquad (2.15)$$

Since the rank of A is m, there are $m - r$ columns which together with $a_1, \ldots, a_r$ form a basis when $r < m$. Without loss of generality, assume that the first m columns of A constitute a basis, i.e.,

$$B = (a_1, \ldots, a_r, a_{r+1}, \ldots, a_m), \qquad N = \{a_{m+1}, \ldots, a_n\}. \qquad (2.16)$$

From (2.15) and $Ad = 0$, it is follows that

$$\sum_{i=1}^{r} d_i a_i + d_{m+1} a_{m+1} = 0,$$

hence

$$a_{m+1} = -\sum_{i=1}^{r} (d_i / d_{m+1}) a_i.$$

Assume that basis B corresponds to the canonical form below:

$$x_B = \bar{b} - \bar{N} x_N. \qquad (2.17)$$

Since its augmented matrix $(I \ \bar{N} \mid \bar{b})$ comes from $(A \mid b) = (B \ N \mid b)$ by elementary transformations, and a_{m+1} is a linear combination of $a_1, \ldots, a_r$, the last $m - r$ components of column $\bar{a}_{m+1}$ in the canonical form are all zero.

Let $\bar{x}$ belong to P. If $\bar{x}_N = 0$, then $\bar{x}$ is the basic feasible solution corresponding to basis B. Now assume, otherwise, that $\bar{x}_N \ne 0$. We will create a new basis B associated with a basic feasible solution by a series of elementary transformations and some solution updating.

Assume that $\bar{x}_j > 0$ holds for some $j \in N$, $j \neq m + 1$. Reducing $\bar{x}_j$ and keep the other nonbasic components unchanged, we determine the corresponding value of $\bar{x}_B$ such that $\bar{x}$ satisfies (2.17), resulting in a new solution $\bar{x}$. For the column corresponding to $\bar{x}_j$, of the canonical form (2.17), there are only two cases arising:

(i) $\bar{a}_j \geq 0$.

It is clear that $\bar{x}_j$ may decrease to 0 and associated $\bar{x}_B$ remains nonnegative. Thus, setting $\bar{x}_j = 0$ gives a new feasible solution $\bar{x}$.

(ii) $\bar{a}_j \ngeq 0$.

If the first r components of $\bar{a}_j$ are nonnegative, one of the last $m-r$ components of $\bar{x}_B$ decreases to 0 first (so-called "blocking") as $\bar{x}_j$ decreases. Assume that the blocking is component i ($r + 1 \leq i \leq m$). Set $\bar{x}_j$ to the according value, and interchange a_j and a_i to update B and N. By exchanging their indices at the same time, the nonbasic component $\bar{x}_j$ of the new solution $\bar{x}$ becomes 0.

If some of the first r components of $\bar{a}_j$ are negative, then we can determine a $\sigma > 0$ such that the first r basic components of

$$\bar{x} := \bar{x} + \sigma d$$

are sufficiently large and all the nonbasic components remain unchanged, except for $\bar{x}_{m+1}$, such that no broking happens to the first r components of $\bar{x}_B$ as $\bar{x}_j$ decreases. If no broking happens to the last $m - r$ components either, we set $\bar{x}_j = 0$. If broking happens to component $r + 1 \leq i \leq m$, we set $\bar{x}_j$ to the according value, and interchange a_j and a_i to updated B and N. Consequently, by exchanging their indices, the nonbasic component $\bar{x}_j$ of the new solution $\bar{x}$ is now equal to 0.

As such, we can transform all $\bar{x}_j$, $j \in N$, $j \neq m + 1$ to 0, without affecting the first r indices of B. Then, if $\bar{x}_{m+1} = 0$, we are done.

If $\bar{x}_{m+1} > 0$, we reduce it and keep the other nonbasic components unchanged. Since the last $m - r$ components of $\bar{a}_{m+1}$ are zero, the corresponding components of $\bar{x}_B$ remain unchanged, and there will be only two cases arising:

(i) The first r components of $\bar{a}_{m+1}$ are all nonnegative. Then it is clear that $\bar{x}_{m+1}$ can decrease to 0 and according $\bar{x}_B$ remain nonnegative, thus we set $\bar{x}_{m+1} = 0$.

(ii) Some of the first r components of $\bar{a}_{m+1}$ are negative. If the according $\bar{x}_B$ remains nonnegative as $\bar{x}_{m+1}$ decreases to 0, we set $\bar{x}_{m+1} = 0$; otherwise, if broking happens for component $1 \leq i \leq r$ of $\bar{x}_B$, we set $\bar{x}_{m+1}$ to the associated value, and exchange a_{m+1} and a_i to update B and N. Then, by exchanging their indices, the nonbasic component $\bar{x}_{m+1}$ of the new $\bar{x}$ is equal to zero.

Therefore, it might be well to assert that the basis B, defined by (2.16), corresponds to basic feasible solution $\bar{x}$.

Consider the following face

$$P' = \{x \in \mathcal{R}^n \mid Ax = b, \ x \geq 0, \ x_j = 0, \ j = m+2,\ldots,n\}$$
$$= \{x \in \mathcal{R}^n \mid Bx_B + x_{m+1}a_{m+1} = b, \ x_B, x_{m+1} \geq 0;$$
$$x_j = 0, \ j = m+2,\ldots,n\}$$
$$= \{x \in \mathcal{R}^n \mid x_B + x_{m+1}\bar{a}_{m+1} = \bar{b}, \ x_B, x_{m+1} \geq 0; \quad (2.18)$$
$$x_j = 0, \ j = m+2,\ldots,n\}.$$

It is clear that $\bar{x} \in P'$. Hence, from (2.15) and

$$Bd_B + d_{m+1}a_{m+1} = 0, \quad (2.19)$$

it is known that

$$\bar{x} + \alpha d \in P', \qquad \forall \, \alpha \geq 0.$$

Therefore, d is a unbounded direction of P'. It is now only needed to show $\dim P' = 1$.

For any $x' \in P'$ and $x' \neq \bar{x}$, introduce $d' = x' - \bar{x}$. It is clear that

$$d'_1,\ldots,d'_r, d'_{m+1} > 0, \quad d'_{r+1},\ldots,d'_m, d'_{m+2},\ldots,d'_n = 0 \quad (2.20)$$

and

$$Bd'_B + d'_{m+1}a_{m+1} = 0. \quad (2.21)$$

From (2.19) and (2.21) it follows respectively that

$$d_B = -d_{m+1}B^{-1}a_{m+1}, \qquad d'_B = -d'_{m+1}B^{-1}a_{m+1},$$

Therefore, $d' = (d'_{m+1}/d_{m+1})d$, where $d'_{m+1}/d_{m+1} > 0$. This implies that $\dim P' = 1$. Therefore, P' is an 1-dimensional face or edge, and d is an unbounded direction of it.

Sufficiency. Assume that d is an unbounded direction of edge P' (2.19), and hence satisfies (2.19). If d is a positive linear combination of unbounded directions d', d'' of P, then there exists a correspondence between zero components of d and of d', d'', and hence d', d'' are also unbounded directions of P'. Since $\dim P' = 1$, in addition, d' and d'' have the same direction. Therefore, d is an extreme direction. $\qquad\square$

Lemma 2.2.4. *An unbounded direction that is not an extreme one is a positive linear combination of two unparallel unbounded directions.*

Proof. Without loss of generality, assume that columns $a_1, \ldots, a_k$ correspond to positive components of an unbounded direction d, and the first r columns are linearly independent. As d is not an extreme direction, it holds by Theorem 2.2.4 that $r < k - 1$, or

$$k - r \geq 2. \tag{2.22}$$

Introduce matrix $B_1 = (a_1, \ldots, a_r)$. Within the set of $a_{k+1}, \ldots, a_n$, determine $m - r$ columns, which might be assumed to correspond to $B_1 = (a_{k+1}, \ldots, a_{k+m-r})$, to construct basis $B = (B_1, B_1)$ (it is clear that $m - r \leq n - k$). Then the nonbasic matrix is $N = (N_1, N_2)$, where $N_1 = (a_{r+1}, \ldots, a_k)$, $N_2 = (a_{k+m-r+1}, \ldots, a_n)$. Assume that B corresponds to the canonical form $x_B = \bar{b} - \bar{N} x_N$. Then d satisfies

$$B_1 d_{B_1} = -\bar{N}_1 d_{N_1}, \qquad d_{N_2} = 0.$$

or equivalently,

$$d_{B_1} = -(B_1^T B_1)^{-1} B_1^T \bar{N}_1 d_{N_1}, \qquad d_{N_2} = 0.$$

Introduce $e \in \mathcal{R}^{k-r}$ and

$$d'_{B_1} = -(B_1^T B_1)^{-1} B_1^T \bar{N}_1 (d_{N_1} + \epsilon e), \qquad d''_{B_1} = -(B_1^T B_1)^{-1} B_1^T \bar{N}_1 (d_{N_1} - \epsilon e).$$

Letting $\epsilon = (\epsilon_1, \ldots, \epsilon_{k-r})^T > 0$, then vectors

$$d' = \begin{pmatrix} d'_{B_1} \\ 0 \\ d_{N_1} + \epsilon \\ 0 \end{pmatrix}, \qquad d'' = \begin{pmatrix} d''_{B_1} \\ 0 \\ d_{N_1} - \epsilon \\ 0 \end{pmatrix}$$

satisfy $d = d'/2 + d''/2$. It is clear that $d', d'' \neq 0$ and $Ad' = 0, Ad'' = 0$. As (2.22) holds, it is known that there exists a sufficiently small ϵ such that $d', d'' \geq 0$ are unparallel unbounded directions. $\square$

Theorem 2.2.6. *If P has an unbounded direction, it has a extreme direction.*

Proof. Let d be an unbounded direction. Assume that it has k positive components and the rank of the corresponding columns is $r \leq k - 1$. If $r = k - 1$, then d is an extreme direction (Theorem 2.2.4). Otherwise, by Lemma 2.2.4, d can be expressed as

$$d = \sigma_1 d' + \sigma_2 d'',$$

where d', d'' are unparallel unbounded directions, and $\sigma_1, \sigma_2 > 0$. Without loss of generality, assume that d'' has at least one component greater than the corresponding component of d'. Thus, the following vector

$$\hat{d} = d'' - \alpha(d'' - d'), \quad \alpha = \min\{d_j''/(d_j'' - d_j') \mid d_j'' - d_j' > 0, \ j = 1, \ldots, n\} > 0.$$

is well-defined. Consider

$$A\tilde{d} = 0, \qquad \tilde{d} \geq 0, \qquad \tilde{d} \neq 0,$$

where the first two equalities clearly hold. If the third does not hold, then $d'' - \alpha(d'' - d') = 0$, as leads to

$$d' = \frac{\alpha - 1}{\alpha} d'',$$

implies that d', d'' are parallel, as a contradiction. Therefore, the third equality also holds, and hence $\tilde{d}$ is an unbounded direction. In addition, it is known that the number of zero components of $\tilde{d}$ is less than that of d by 1, at least. Then set $d = \tilde{d}$ and repeat the preceding steps. It is clear that such a process can only repeat finitely many times, and terminates at some extreme direction.   □

According to Theorem 2.2.5, extreme directions and unbounded edges of P are 1-to-1 correspondent (extreme directions having the same direction are viewed as the same). As there are finitely many edges, the number of extreme directions are finite.

It is now time to lay a theoretical basis to Dantzig-Wolfe decomposition method for solving large-scale LP problems (Chap. 8).

Theorem 2.2.7 (Representation Theorem of the Feasible Region). *Let P be nonempty. Assume that $\{u^1, \ldots, u^s\}$ is the vertex set and $\{v^1, \ldots, v^t\}$ the extreme direction set. Then $x \in P$ if and only if*

$$x = \sum_{i=1}^{s} \alpha_i u^i + \sum_{j=1}^{t} \beta_j v^j,$$

$$\sum_{i=1}^{s} \alpha_i = 1, \qquad \alpha_i \geq 0, \quad i = 1, \ldots, s, \tag{2.23}$$

$$\beta_j \geq 0, \qquad j = 1, \ldots, t.$$

Proof. When (2.23) holds, it is clear that $x \in P$. So it is only needed to show necessity. We use inductive method to dimensions.

If dim $P = 0$, P is a single point set, including a vertex. The conclusion holds clearly. Assume that it holds for dim $P < k$. We will show that it holds for dim $P = k \geq 1$.

From Proposition 2.1.3, it follows that int $P \neq \emptyset$. Assume $x \in$ int P, and consider

$$x' = x - \lambda(u^1 - x). \tag{2.24}$$

Note that $u^1 \neq x$. There are the following two cases arising:

(i) $u^1 - x \not\leq 0$.

Determine λ such that

$$\lambda = x_q/(u_q^1 - x_q) = \min\{x_j/(u_j^1 - x_j) \mid u_j^1 - x_j > 0, \ j = 1, \ldots, n\} > 0.$$

Then, it is easy to verify that x', defined by (2.24), satisfies $x' \in P$ and $x'_q = 0$. Therefore, x' belongs to some proper face with its dimension less than k. According to the assumption of induction, x' can be expressed as the sum of a convex combination of vertices and a nonnegative combination of extreme directions of the proper face. Since vertices and extreme directions of a face are also that of P, therefore, x' can be expressed as the sum of a convex combination of vertices and a nonnegative combination of extreme directions of P, i.e.,

$$x' = \sum_{i=1}^{s_1} \alpha'_i u^i + \sum_{j=1}^{t_1} \beta'_j v^j,$$

$$\sum_{i=1}^{s_1} \alpha'_i = 1, \qquad \alpha'_i \geq 0, \quad i = 1, \ldots, s_1,$$

$$\beta'_j \geq 0, \qquad j = 1, \ldots, t_1,$$

where u^i and v^j are vertices and extreme directions of P, respectively. Substituting the preceding into

$$x = \frac{1}{1+\lambda} x' + (1 - \frac{1}{1+\lambda}) u^1,$$

which is equivalent to (2.24), leads to the expression of the form of (2.23), i.e.,

$$x = \sum_{i=1}^{s_1} \frac{1}{1+\lambda} \alpha'_i u^i + (1 - \frac{1}{1+\lambda}) u^1 + \sum_{j=1}^{t_1} \frac{1}{1+\lambda} \beta'_j v^j.$$

(ii) $u^1 - x \leq 0$.

Then x', defined by (2.24), are all feasible points for any $\lambda \geq 0$, hence $-(u^1 - x)$ is an unbounded direction. According to Lemma 2.2.6, there is an extreme direction, say v^1. Now take a sufficiently large μ such that

$$\tilde{x} = u^1 + \mu v^1$$

has at least one component greater than the corresponding component of x. Consequently, the point defined by

$$x' = x - \lambda(\tilde{x} - x),$$

$$\lambda = x_q/(\tilde{x}_q - x_q) = \min\{x_j/(\tilde{x}_j - x_j) \mid \tilde{x}_j - x_j > 0, \ j = 1, \ldots, n\} > 0.$$

is a feasible point, satisfying $x'_q = 0$. Therefore, this point belongs to some proper face with dimension less than k, and hence can be expressed as the sum of a convex combination of vertices and a nonnegative combination of extreme directions of P. As a result, an expression of the form (2.23) can be obtained in an analogous manner as case (i). $\square$

The preceding Theorem indicates that a feasible point is the sum of a convex combination of vertices and a nonnegative combination of extreme directions, and vice versa. In particular, the following is a direct corollary.

Corollary 2.2.2. *Let the feasible region be bounded. A point is feasible if and only if it is a convex combination of vertices.*

2.3   Optimal Face and Vertex

We describe a basic result without proof (see, e.g., Rockafellar 1997).

Theorem 2.3.1 (Partition Theorem). *Let $\bar{x}$ be a boundary point of convex set S. Then there exists a superplane including $\bar{x}$ and partitioning the total space to two half spaces, one of which includes S.*

The superplane involved in the preceding Theorem is said to be *supporting superplane* of S. That is to say, there is a supporting superplane through every boundary point of a convex set. Although the result is applicable to any convex set, we are only concerned with the feasible region P, in particular.

Supporting superplane of P is closely related to its face, as the following reveals.

Lemma 2.3.1. *The intersection of P and a supporting superplane is a face.*

Proof. Assume that $H = \{x \in \mathcal{R}^n \mid a^T x = \eta\}$ is the superplane of P, and $P' = P \cap H$. Without loss of generality, let $a^T x \leq \eta$ hold for all $x \in P$.

Assume that $v \in P'$ is an interior point of segment (y, z), and $y, z \in P$, i.e.,

$$v = \alpha y + (1 - \alpha)z, \tag{2.25}$$

where $0 < \alpha < 1$. Note that, P' is a nonempty convex set.

It is only needed to show $y, z \in H$, as leads to $y, z \in P'$, hence P' is a face of P. Assume $y, z \notin H$. Then it holds that

$$a^T y < \eta, \qquad a^T z < \eta.$$

Multiplying the preceding two formulas respectively by $\alpha > 0$ and $1 - \alpha > 0$, and then adding the results gives

$$a^T (\alpha y + (1 - \alpha)z) < \eta \alpha + \eta(1 - \alpha) = \eta,$$

combining which and (2.25) leads to $v \notin H$, and hence $v \notin P'$. This contradicts the assumption $v \in P'$. Therefore, at least one of y and z belongs to H.

Without loss of generality, assume $z \in H$. Then, it follows from (2.25) that

$$y = (1/\alpha)v + (1 - 1/\alpha)z,$$

implying that y is in the straight line through v and z. Moreover, $z, v \in H$ and H is a superplane, therefore $y \in H$. □

Lemma 2.3.2. *Let $\bar{f}$ be the optimal value of the standard LP problem. Set F is the set of optimal solutions if and only if it is the intersection of P and the objective contour plane*

$$\bar{H} = \{x \in \mathcal{R}^n \mid c^T x = \bar{f}\}. \tag{2.26}$$

Proof. Assume $F = P \cap \bar{H}$. It is clear that any optimal solution $\bar{x} \in P$ satisfies $c^T \bar{x} = \bar{f}$, implying $\bar{x} \in \bar{H}$, hence $\bar{x} \in F$. Therefore, F is the set of optimal solutions. If F is the set of optimal solutions, conversely, then $c^T \bar{x} = \bar{f}$ holds for any $\bar{x} \in F \subset P$. Therefore $\bar{x} \in \bar{H}$, and hence $\bar{x} \subset P \cap \bar{H}$. □

It is clear that $\bar{H}$ is a supporting superplane of P, as is referred to as *objective supporting superplane*.

A face is optimal if its elements are all optimal solutions. A vertex is optimal if it is an optimal solution.

Lemma 2.3.3. *If there exists an optimal solution, then there exists an optimal face.*

Proof. According to Lemma 2.3.2, a nonempty set of optimal solutions is the intersection of feasible region P and objective contour plane $\bar{H}$. Therefore it is an optimal face, according to Lemma 2.3.1 and the definition of an optimal face. □
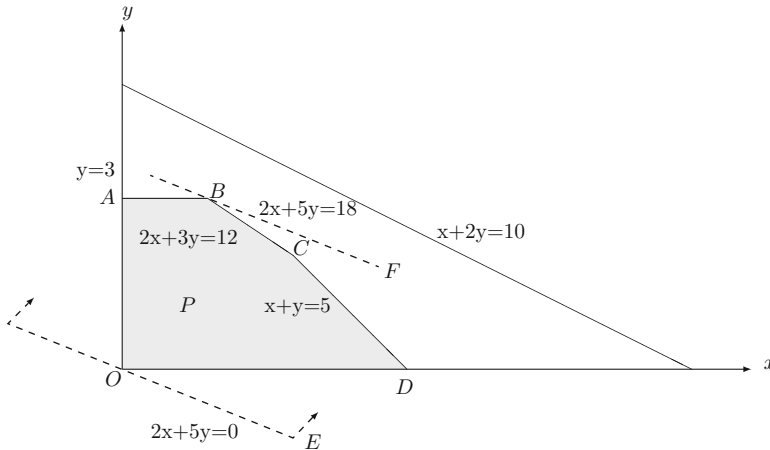
Fig. 2.3 Graphic solution to Example 1.2.2

Theorem 2.3.2. *If there exists a feasible solution, there exists a basic feasible solution. If there is an optimal solution, there is a basic optimal solution.*

Proof. By Lemmas 2.2.2 and 2.2.3, it is known that nonempty feasible region has a basic feasible solution. By Lemma 2.3.3, if there is an optimal solution, there is an optimal face, which is a nonempty polyhedral convex set, and hence having an optimal convex or basic optimal solution. □

In presence of optimal solution, there exists an optimal 0-dimensional face (or vertex). In general, there could exist optimal faces of higher dimensions. It is clear that the optimal face of the highest dimension is the set of all optimal solutions, as is referred to as *optimal set*. After a LP problem is solved by the simplex method, the optimal set can be obtained easily (Sect. 25.2).

2.3.1 Graphic Approach

A LP problem of 2-dimension can be solved via a graphic approach. To do so, let us return to Example 1.2.2. The shaded area enclosed by polygon OABCD in Fig. 1.1 is the feasible region (ignore the straight line $x + 2y = 10$, at the moment). It is required to determine a point over the area such that the objective function reaches the highest value at the point (Fig. 2.3).

In the figure, the equation $2x + 5y = 0$ of the contour line of the objection function corresponds to the dashed line OE, going through the origin, all points on which correspond to the same objective value 0. The line's slope, i.e., the tangent of the angle between it and x axis, is $-2/5 = -0.4$. Therefore, the corresponding contour line shifts parallel to the upper-right side as the objective value increases

from 0. Points in the intersection of the line and the area of OABCD are all feasible points. The parallel shifting should be carried out as far as the intersection remains nonempty to attain the biggest possible objective value. It is seen from the figure that the contour line shifting the farthest is the dashed line BF though vertex B, that is, the "objective supporting plane". The optimal set, i.e., the intersection of the line and the feasible region includes a single vertex B, corresponding to the basic optimal solution. Consequently, problem (1.2) is solved after measuring coordinates of point B, and calculating the associated objective value.

If the figure or measurement is not relatively accurate, however, the end solution would involve unacceptable errors. For a graphic approach, therefore, it would be better to use coordinate paper, with the help of algebraic calculating. Once B is known to be the optimal vertex, for instance, its coordinates can be obtained by solving the following system of equations:

$$
\begin{cases}
2x + 3y = 12 \\
\quad\quad y = \;\; 3
\end{cases}
\tag{2.27}
$$

from which the optimal basic solution $\bar{x} = 1.5$, $\bar{y} = 3$ to problem (1.2) follows, with the optimal value $f = 18$. That is to say, the manufacturer should arrange production of 1,500 laths and 3,000 sheep piles daily, gaining 18,000 dollars profit.

The graphic approach is not suitable for cases of $n \geq 3$, though it is simple. Even for case of $n = 2$, in fact, its application is rare seen. However, it still offers some inspiration, as is the topic of Sect. 2.5.

2.4 Feasible Direction and Active Constraint

Methods for solving mathematical problems fall into two categories: direct and iterative methods. The latter produces a sequence of points by iterations, offering an exact or approximate solution. Methods presented in this book belong to the iterative category.

Line search is the mostly used iterative approach in optimization. At a current point $\bar{x}$, in each iteration, a new point $\hat{x}$ is determined along a ray starting from $\bar{x}$ along a nonzero vector d, i.e.,

$$
\hat{x} = \bar{x} + \alpha d,
\tag{2.28}
$$

where d is referred to as *search direction*, $\alpha > 0$ as *stepsize*. Once the two are available, $\hat{x}$ can be calculated and then one iteration is complete. Repeating this process yields a sequence of points, until a solution is reached. Formula (2.28) is referred to as *line search* or *iterative scheme*.

The determination of a search direction is crucial. In presence of constraints, d should be such that the intersection of the ray (2.28) and the feasible region is nonempty. More precisely, we introduce the following concept:

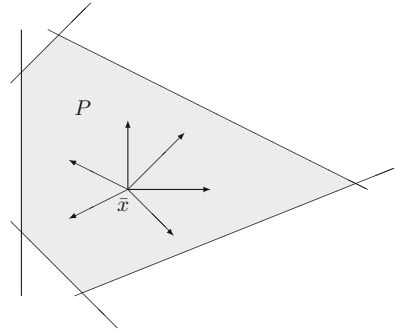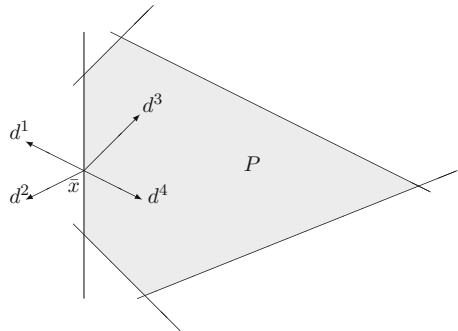**Fig. 2.4** Any direction is feasible at $\bar{x}$

Fig. 2.5 d^3, d^4 are feasible at $\bar{x}$, but d^1, d^2 not

Definition 2.4.1. Let P be the feasible region. Assume that $\bar{x} \in P, \ d \neq 0$. If there exists $\bar{\alpha} > 0$ such that

$$\bar{x} + \alpha d \in P, \qquad \forall \, \alpha \in [0, \bar{\alpha}],$$

d is a feasible direction at $\bar{x}$.

The preceding is relevant to general constrained optimization problems, including the LP problem. The following are some instances, in conjunction with feasible region P.

Example 2.4.1. In Fig. 2.4, $\bar{x}$ is an interior point of P, and hence any direction is feasible at it.

Example 2.4.2. In Fig. 2.5, $\bar{x}$ is a boundary point of P. It is seen that d^3, d^4 are feasible directions at $\bar{x}$, but d^1, d^2 are not.

Example 2.4.3. In Fig. 2.6, $\bar{x}$ is a vertex of P. Any direction, e.g., d^4 or d^5, within the angle area between d^6 and d^7 (which are respectively along two sides of P) is feasible at $\bar{x}$. Vectors d^1, d^2, d^3 are not feasible.

Let d be a feasible search direction. It is possible to maintain feasibility of some new iterate $\hat{x}$. In order for $\hat{x}$ to be close to an optimal solution, it is needed to take

Fig. 2.6 d^4, d^5, d^6, d^7 are feasible at $\bar{x}$, but d^1, d^2, d^3 not
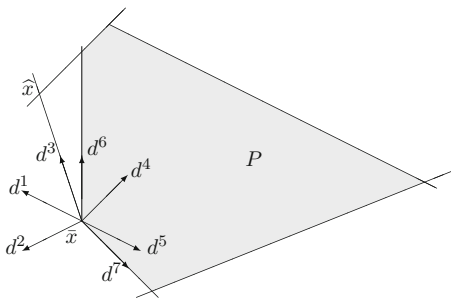
into account the objective function. To this end, it might be well to consider the following problem:

$$
\begin{aligned}
\min \quad & f = c^T x, \\
\text{s.t.} \quad & a_i^T x \geq b_i, \qquad i = 1, \ldots, m,
\end{aligned}
\tag{2.29}
$$

where $m > n$, and whose feasible region is

$$
P = \{x \in \mathcal{R}^n \mid a_i^T x \geq b_i, \ i = 1, \ldots, m\}.
$$

Definition 2.4.2. Vector d satisfying condition $c^T d < 0$ is a descent direction. If d is also a feasible direction at $\bar{x} \in P$, it is a feasible descent direction at $\bar{x}$.

It is clear that d is a feasible descent direction at $\bar{x}$ if and only if it is a feasible direction and forms an obtuse angle with the objective gradient c. Once such a direction d is available, a stepsize $\alpha > 0$ can be determined, and hence a new iterate $\hat{x} \in P$, obtained by (2.28), corresponding to a smaller objective value. Then, one iteration is complete.

It is noticeable that not all constraints affect the determination of a feasible descent direction at a current point.

Definition 2.4.3. A constraint which is violated, or satisfied as an equality, by the current point is an active constraint.

Aimed at a *feasible* point, the "active" constraint is usually defined as one satisfied as equality, or binding at the point. But it seems to be useful to include infeasible point by regarding a constraint violated as active.

Let us bring up (2.29) as an example. If $a_i^T \bar{x} = b$, then $a_i^T x \geq b$ is active at $\bar{x}$; if, otherwise, $a_i^T \bar{x} > b$, then $a_i^T x \geq b$ is not active at that point. So, the current point is on a boundary of an active constraint. From the simple instances given in the preceding figures, it is seen that a feasible descent direction can be determined by taking into account active constraints only.

In practice, the preceding definition for active constraint does not come up to expectations, as a current point close to boundary could lead to too small stepsize,

and hence insignificant progress. In view of this, Powell (1989) proposes the so-called "ϵ-active" constraint, where ϵ is a small positive number. For example, if $a_i^T \bar{x} - b \le \epsilon$, then $a_i^T x \ge b$ is an ϵ-active constraint at $\bar{x}$; whereas it is not if $a_i^T \bar{x} - b > \epsilon$.

The LP problem can be solved by the so-called "active set method", which is usually used for solving nonlinear programming problems though some scholars prefer it to be a LP problem solver (e.g., Fletcher 1981; Hager 2002). We outline the method in conjunction with problem (2.29) in the remainder of this section.

Assume that the current vertex $\bar{x}$ is an unique solution to the linear system below:

$$a_i^T x = b_i, \qquad i \in \mathcal{A},$$

where $\mathcal{A}$ is called *the active set (of constraints)*, consisting of n indices of (total or part) active constraints, with linearly independent gradients a_i. If $\bar{x}$ is judged to be optimal under some criterion, we are done; otherwise, some index $p \in \mathcal{A}$ is selected such that the $n - 1$ equalities

$$a_i^T x = b_i, \qquad i \in \mathcal{A} \backslash \{p\}, \tag{2.30}$$

determines a descent edge (1-dimensional face).

In fact, since the rank of the coefficient matrix of (2.30) is $n - 1$, the associated homogeneous system

$$a_i^T d = 0, \qquad i \in \mathcal{A} \backslash \{p\},$$

has a solution d such that

$$d \ne 0, \qquad c^T d < 0.$$

It is easily verified that for $\alpha \ge 0$, all points on the ray

$$\hat{x} = \bar{x} + \alpha d,$$

satisfy (2.30). Under the condition that the objective value is lower bounded over the feasible region, the following stepsize is well defined:

$$\alpha = (b_q - a_q^T \bar{x})/a_q^T d = \min\{(b_i - a_i^T \bar{x})/a_i^T d \mid a_i^T d < 0, \ i \notin \mathcal{A}\} \ge 0.$$

In fact, such an α is the largest stepsize possible to maintain feasibility of $\hat{x}$. Since $a_q^T \hat{x} = b_q$, the $a_q^T x \ge b_q$ is an active constraint at $\hat{x}$.

Consequently, setting $\bar{x} = \hat{x}$ and redefining $\mathcal{A} = \mathcal{A} \backslash \{p\} \cup \{q\}$ completes an iteration of the active set method. If $\alpha > 0$, the new vertex corresponds to a lower objective value. Note however that if there are multiple active constraints at the current vertex $\bar{x}$, then α defined by (2.30) would vanish, so that the descent edge

degenerates to a vertex. That is to say, the "new vertex" actually coincides with the old, though set $\mathcal{A}$ changes. Such a vertex is called *degenerate*. In Fig. 2.6, e.g., $\bar{x}$ is a degenerate vertex, at which the three edges along respective d^3, d^6, d^7 intersect.

The simplex method can be viewed as a special scheme of the active set method, as will be discussed in Sect. 3.9. In history, however, the former emerged before the latter from another path by fully taking advantage of the linear structure.

2.5   Heuristic Characteristic of Optimal Solution

From the graphic approach demonstrated in Fig. 2.3, it is seen that the optimal solution to the LP problem is attained at a vertex of the feasible region. It is imaginable that if the feasible region has a side going through the vertex, which is parallel to objective contour lines, then the whole side corresponds the optimal set, associated with the same optimal value.

So, the solution key lies on how to determine lines intersecting at an optimal vertex. In other words, it is only needed to know which inequalities are active at an optimal solution. Once these active inequalities are known, what left to do is just to solve an linear system; for the instance in Fig. 2.3, the optimal solution was quickly calculated through solving the system (2.27).

Thereby, we make the observation that normal directions (pointing to the interior of the feasible region) of lines AB and BC, intersecting at the optimal vertex B, form *the largest angles* with the parallel shifting direction of the contour line BF, among all the lines.

Now turn to the more general minimization problem

$$\begin{aligned} \min \quad & f = c^T x, \\ \text{s.t.} \quad & Ax \geq b, \end{aligned} \qquad (2.31)$$

where $A \in \mathcal{R}^{m \times n}$, $c \in \mathcal{R}^n$, $b \in \mathcal{R}^m$, $m > 1$. Note that constraints here are all inequalities of "$\geq$" type.

We could imagine analogously in the space of multiple dimensions. Now the constraint inequalities correspond to half spaces, and vertices correspond to intersection points formed by half spaces. What we should do is to examine angles between normal directions and parallel shifting direction of the contour plane. For the minimization problem, the shifting direction is the negative gradient direction of the objective function. This leads to the following plausible statement (Pan 1990).

Proposition 2.5.1 (Heuristic characteristic of optimal solution). *Gradients of active constraints at an optimal solution of a minimization problem tend to form the largest angles with the negative objective gradient.*

It is now needed to quantize magnitude of the angles. If we denote the ith row vector of A by $\bar{a}_i^T$, the cosine of the angle between the ith constraint gradient and negative objective gradient is then

$$\cos < \bar{a}_i, c >= -\bar{a}_i^T c/(\|\bar{a}_i\|\|c\|).$$

For simplicity, we ignore the constant factor $1/\|c\|$ and introduce the following

Definition 2.5.1. The pivoting-index of the ith constraint is defined as

$$\alpha_i = -\bar{a}_i^T c/\|\bar{a}_i\|. \tag{2.32}$$

Then, we are able to compare the angles by pivoting-indices, and hence Proposition 2.5.1 may be reformulated as

Gradients of active constraints at an optimal solution tend to have the smallest pivoting-indices.

Example 2.5.1. Investigate problem (1.2) via pivoting-indices:

$$
\begin{aligned}
\min \ & f = -2x - 5y, \\
\text{s.t.} \quad -\,& 2x - 3y \geq -12, \\
-\,& x - \ y \geq - \ 5, \\
-\,& \ y \geq - \ 3, \\
& x, y \geq 0.
\end{aligned}
\tag{2.33}
$$

Answer   Calculate indices of the constraints, and put them in the following table in the order of increasing pivoting-indices:

Constraints	α_i
$-2x - 3y \geq -12$	-5.26
$-y \geq -3$	-5.00
$-x - y \geq -5$	-4.95
$x \geq 0$	2.00
$y \geq 0$	5.00

From the preceding table, it is seen that $-2x - 3y \geq -12$ and $-y \geq -3$ are the two constraints with the smallest pivoting-indices. Thus, the two are active constraints at an optimal solution. This immediately leads to solving system (2.27), as coincides with the outcome from the graphic approach.

If Proposition 2.5.1 were true in general, solving the LP problem amounts to solving a system of linear equations by $O(m^3)$ basic arithmetics, in contrast to existing iterative methods (see Sect. 3.8)! Unfortunately, the characteristic of

Fig. 2.7 The normal vectors (pointing to the interior of the polyhedron) of planes *ABC*, *ABD* and *CBD* form the most obtuse three angles with the negative objective gradient $-c$. But the optimal vertex is not their intersection *B*, but *A*, the intersection of planes *ABC*, *ABD* and (*vertical*) *EAF*
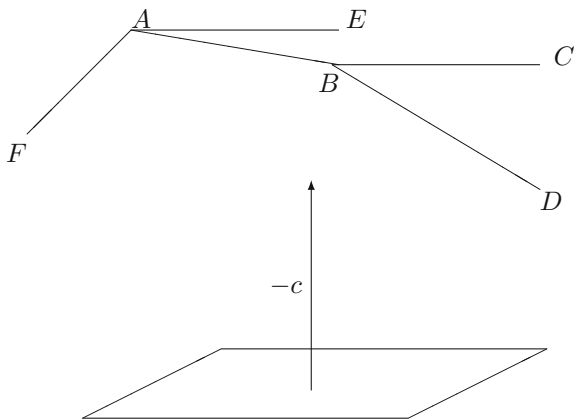
optimal solution is only heuristic. Counterexamples are easily constructed. If adding a constraint $x + 2y \leq 10$ to (1.2) (see Fig. 2.3), e.g., it is then clear that the added constraint is not active at the optimal solution, though its gradient forms the largest angle with the objective gradient (with pivoting-index -5.37); in fact, added is a redundant constraint, not affecting the feasible region at all. It is also not difficult to construct counterexamples without redundant constraint. Someday in the Fall of 1986 when the author was visiting the Mathematics Department of University of Washington, after he talked with Professor Rockafellar about his idea on the characteristic of an optimal solution, the latter quickly he showed a counterexample by sketching on a piece of paper, as is seen in Fig. 2.7.

For all that, Proposition 2.5.1 might still offer some clue toward optimal solution, shedding a light on LP research. Such a trick well be referred to as "the most-obtuse-angle heuristics" in this book.

In some cases, in fact, it is possible to detect unboundedness of a problem simply from signs of pivoting-indices.

Theorem 2.5.1. *Assume that the feasible region is nonempty. If pivoting-indices of constraints are all nonnegative, then problem (2.31) is unbounded.*

Proof. By contradiction. Assume that the problem is bounded under the assumptions. It follows that

$$c^T v \leq 0, \qquad \forall \, v \in \{v \in \mathcal{R}^n \mid Av \geq 0\}. \tag{2.34}$$

In fact, there is a vector v such that

$$Av \geq 0, \qquad c^T v > 0. \tag{2.35}$$

Thus, for any $\alpha \geq 0$ and feasible solution $\bar{x}$, vector $x = \bar{x} + \alpha v$ satisfies $Ax \geq b$, that is, v is an unbounded direction. Further, it is known from (2.35) that
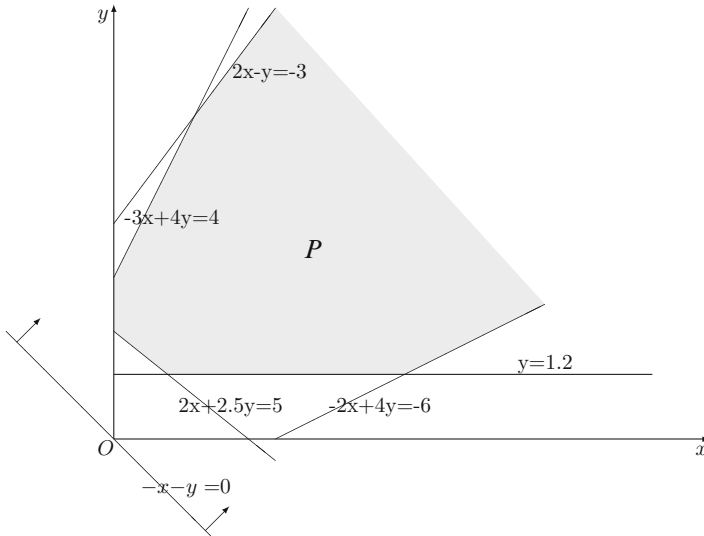
Fig. 2.8 Unbounded problem

$$c^T x = c^T \bar{x} + \alpha c^T v \to \infty, \quad (\text{as } \alpha \to \infty),$$

which contradicts that (2.31) is upper bounded. Therefore, (2.34) holds.

Thus, according to Farkas' Lemma 2.1, there is $y \geq 0$ such that

$$-c = A^T y,$$

premultiplying which by $-c^T$ gives

$$0 < c^T c = -y^T A c.$$

Hence $Ac \geq 0$ follows from nonnegativeness of pivoting-indices. This implies that the right-hand side of the preceding is less than or equal to 0, as is a contradiction. Therefore, the problem is unbounded. □

An alternative proof of the preceding Theorem is via showing $-c$ to be an unbounded direction of the feasible region. The following Corollary gives a necessary condition for the existence of an optimal solution.

Corollary 2.5.1. *If there is an optimal solution, then there is at least one constraint bearing negative pivoting-index.*

Example 2.5.2. Investigate the following LP problem by pivoting-indices:

$$
\begin{array}{lrcl|c}
\min & f = -x - y & & & \alpha_i \\
\hline
\text{s.t.} & 2x - & y \geq & -3 & 0.71 \\
& -3x + 4y \geq & & 4 & 0.20 \\
& 2x + 2.5y \geq & & 5 & 1.41 \\
& -2x + 4y \geq & & -8 & 0.45 \\
& y \geq & & 1.2 & 1.00 \\
& x \geq & & 0 & 1.00 \\
& y \geq & & 0 & 1.00 \\
\hline
\end{array}
\tag{2.36}
$$

Answer   Calculate indices of the constraints, and fill in the right-hand side of the preceding table. According to Theorem 2.5.1, the problem is unbounded since all indices are nonnegative (see Fig. 2.8).