

## The Geometry of Linear Optimization

The most fundamental geometric entity occurring in linear optimization is the hyperplane

$$H_a^b \triangleq \{x \in R^n \mid ax = b\},$$

whose description involves some nonzero vector  $a \in R^n$  and some scalar  $b \in R$ . Such a description is not unique, because  $H_{sa}^{sb}$  is clearly identical to  $H_a^b$  for each nonzero scalar  $s$ .

From elementary analytic geometry, we know that the hyperplanes in  $R^1$  are the points  $b/a$  in  $R^1$ , while the hyperplanes in  $R^2$  are the (straight) lines in  $R^2$ . Moreover, the hyperplanes in  $R^3$  are the planes in  $R^3$ , while the hyperplanes in  $R^n$  for  $n > 3$  have no other name (and hence are sometimes the only entities termed hyperplanes).

Each hyperplane  $H_a^b$  in  $R^n$  separates  $R^n$  into two closed halfspaces

$$P_a^b \triangleq \{x \in R^n \mid ax \leq b\}$$

and

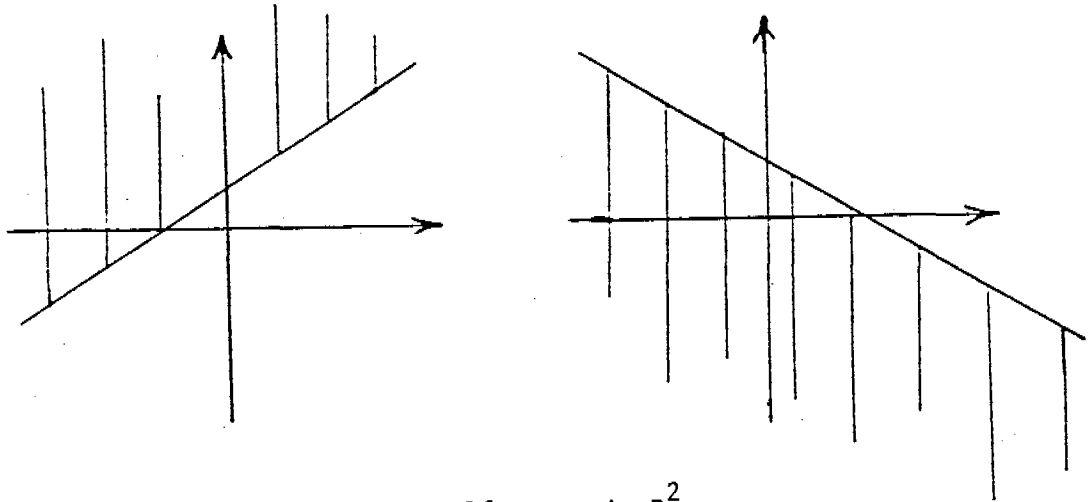
$$P_{-a}^{-b} = \{x \in R^n \mid -ax \leq -b\} = \{x \in R^n \mid ax \geq b\},$$

whose intersect is obviously just  $H_a^b$  and whose union is clearly  $R^n$ .

Needless to say,  $P_{sa}^{sb}$  is identical to  $P_a^b$  for each positive scalar  $s$  while  $P_{sa}^{sb}$  is identical to  $P_{-a}^{-b}$  for each negative scalar  $s$ .

The closed halfspaces in  $R^1$  are, of course, the "closed semi-infinite intervals"  $(-\infty, b/a]$  and  $[b/a, +\infty)$ , sometimes called "closed halflines". They always occur in pairs, with the two in a given pair intersecting at, and separated by, a single point  $b/a$  -- the corresponding hyperplane.

Similarly, the closed halfspaces in  $\mathbb{R}^2$  are called "closed halfplanes". They always occur in pairs, with the two in a given pair intersecting at, and separated by, a line -- the corresponding hyperplane. Two examples of pairs of closed halfspaces in  $\mathbb{R}^2$  are illustrated below, while the visualization



Closed Halfspaces in  $\mathbb{R}^2$   
(Closed Halfplanes)

of closed halfspaces in  $\mathbb{R}^3$  and higher dimensional spaces  $\mathbb{R}^n$  is left to the imagination of the reader.

Intimately related to the pair of closed halfspaces  $P_a^b$  and  $P_{-a}^{-b}$  are the pair of open halfspaces

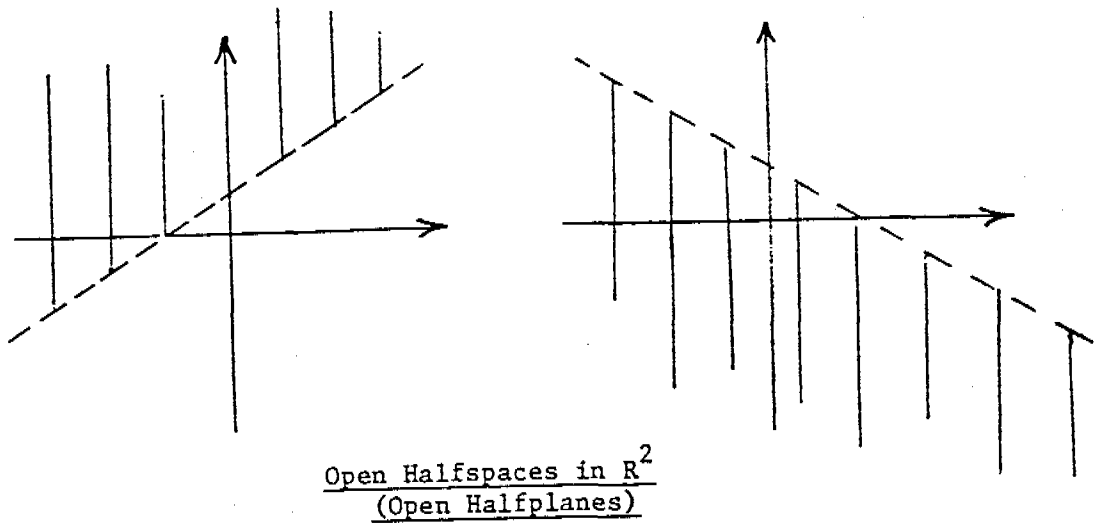
$$(\text{int } P_a^b) \triangleq \{x \in \mathbb{R}^n \mid ax < b\}$$

and

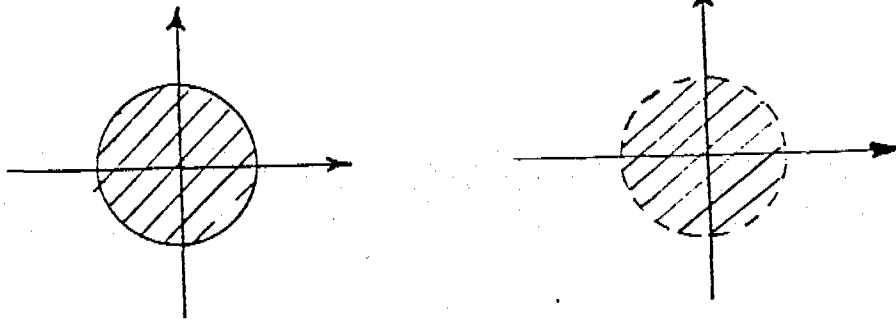
$$(\text{int } P_{-a}^{-b}) = \{x \in \mathbb{R}^n \mid -ax < -b\} = \{x \in \mathbb{R}^n \mid ax > b\},$$

whose intersect is obviously empty and whose union is clearly  $\mathbb{R}^n$  with  $H_a^b$  excluded. The notation expresses the fact that  $(\text{int } P_a^b)$  is just the "interior" of  $P_a^b$ , namely, the points of  $P_a^b$  not on the "boundary"  $H_a^b$  of  $P_a^b$ .

The open halfspaces in  $R^1$  are, of course, the "open semi-infinite intervals"  $(-\infty, b/a)$  and  $(b/a, +\infty)$ , sometimes called "open halflines". They always occur in pairs, with the two in a given pair separated by a single point  $b/a$  — the boundary of each. Similarly, the open halfspaces in  $R^2$  are called "open halfplanes". They always occur in pairs, with the two in a given pair separated by a single line — the boundary of each. Two examples of pairs of open halfspaces in  $R^2$  are illustrated below, while the visualization of open halfspaces in  $R^3$  and higher dimensional spaces  $R^n$  is left to the imagination of the reader.



The boundary lines in the preceding two dimensional examples are illustrated with broken line segments to indicate they are not part of the sets being considered. We will follow the convention of using solid line segments and, more generally, solid curves to illustrate those set boundaries that are to be part of the sets in  $R^2$  being considered. For example, the closed disk  $\{(x_1, x_2) | x_1^2 + x_2^2 \leq 1\}$  and its interior, the open disk  $\{(x_1, x_2) | x_1^2 + x_2^2 < 1\}$ , are respectively illustrated as follows.



### Closed and Open Disks in $R^2$

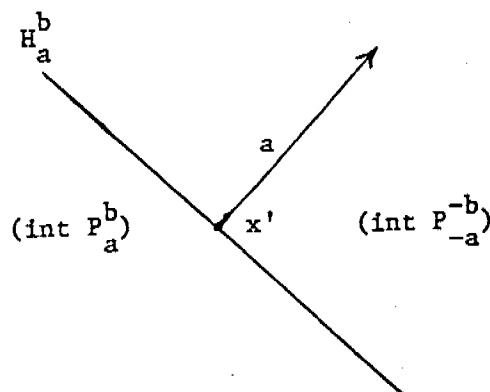
The boundary of each is the circle  $\{(x_1, x_2) \mid x_1^2 + x_2^2 = 1\}$ , which is contained in the closed disk but is not contained in the open disk.

Boundaries of sets in  $R^3$  are normally surfaces and may, or may not, be contained in their corresponding sets. In any event, we will not attempt to illustrate sets in  $R^3$  and hence will not need a graphical device to indicate whether or not given boundary points are contained in their corresponding sets. As might be expected, boundaries of sets in  $R^n$  for  $n > 3$  are normally "hypersurfaces" and may, or may not, be contained in their corresponding sets. Obviously, sets in  $R^n$  for  $n > 3$  can not easily be illustrated within the confines of our physical space  $R^3$ ; so we will not need a graphical device to indicate whether or not given boundary points are contained in their corresponding sets.

By definition, closed sets are those sets that contain all of their boundary points, while open sets are those sets that contain none of their boundary points. Needless to say, some sets are neither closed nor open (such as the half closed, half open interval  $[c,d)$ , which contains the boundary point  $c$  but excludes the other boundary point  $d$ ). The interior,  $(\text{int } S)$ , of a given set  $S \subseteq R^n$  is, by definition, all points in  $S$  that are not on the boundary of  $S$ . The  $(\text{int } S)$  is either empty (e.g.,  $(\text{int } \{c\}) = \emptyset$ ) or open (e.g.,  $\text{int } [c,d) = (c,d)$ ). The sets to be considered here can usually be described by (nonstrict) inequalities, with

the boundary points being those set points that make at least one of the inequalities an equality. In particular, the sets to be considered are usually closed, but it is sometimes helpful to think in terms of their interiors, which consist of those points (if any) that make each of the defining inequalities strict.

To further elaborate on the geometric relations between a hyperplane  $H_a^b$  and its corresponding open halfspaces  $(\text{int } P_a^b)$  and  $(\text{int } P_{-a}^{-b})$ , first note that a given point  $x'$  on  $H_a^b$  must satisfy the equation  $ax' = b$ . Then, note that for any other point  $x$  in  $R^n$ , either  $x$  is in the  $(\text{int } P_a^b)$  because  $ax < b$ , or  $x$  is on  $H_a^b$  because  $ax = b$ , or  $x$  is in the  $(\text{int } P_{-a}^{-b})$  because  $ax > b$  (one, and only one). Consequently, subtracting the equation  $ax'=b$  from each of the preceding relations shows that either  $a(x - x') < 0$ , or  $a(x - x') = 0$ , or  $a(x-x') > 0$ , respectively. This means, of course, that the "angle" between  $a$  and  $x-x'$  is either "obtuse", or "right", or "acute", respectively. In particular, if the vector  $a$  is represented by an "arrow" whose "tail" is at  $x'$ , the arrow is "orthogonal to"  $H_a^b$  and "points away from" the open half space  $(\text{int } P_a^b)$  while "pointing into" the open half space  $(\text{int } P_{-a}^{-b})$  -- a geometric relation that is graphically illustrated by the following diagram in  $R^2$ .



Hyperplanes and Their Open Halfspaces in  $R^2$

In summary, the vector  $a$  determines the "angular orientation" of  $H_a^b$  and its corresponding halfspaces (both open and closed). In fact, the vector  $a$  itself determines only the infinite family of all hyperplanes orthogonal to  $a$ . It is the scalar  $b$  that selects one such hyperplane  $H_a^b$  -- in such a way that  $H_a^b$  is "translated" (without "rotation") in the direction of the vector  $a$  as  $b$  increases (because the halfspace  $P_a^b$  clearly expands while the halfspace  $P_{-a}^{-b}$  obviously contracts as  $b$  increases).

A polyhedral set (or "polyhedron") is the intersect of a finite, nonempty collection of closed halfspaces. However, since each hyperplane is the intersect of its corresponding closed halfspaces, a polyhedral set may also (implicitly) involve the intersect of a finite collection of hyperplanes (in which case its interior is empty). In any event, each polyhedral set  $P$  in  $R^n$  can be described via some  $m \times n$  matrix  $A$  with nonzero rows  $a_i \in R^n$ , and some vector  $b \in R^m$ , as

$$P_A^b \triangleq \{x \in R^n \mid Ax \leq b\}.$$

In particular, the  $i$ 'th row  $a_i$  of  $A$  along with the  $i$ 'th component  $b_i$  of  $b$  describe a hyperplane

$$H_i \triangleq H_{a_i}^{b_i} = \{x \in R^n \mid a_i x = b_i\}$$

whose "normal vector"  $a_i$  points away from the halfspace

$$P_i \triangleq P_{a_i}^{b_i} = \{x \in R^n \mid a_i x \leq b_i\}$$

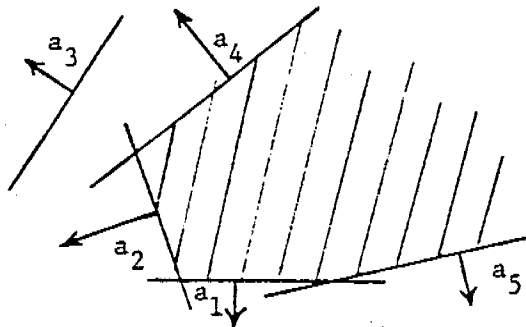
in which  $P_A^b$  lies. Needless to say, if the corresponding constraint

$a_i x \leq b_i$  is not "redundant" (in that it is not automatically satisfied when the other constraints are satisfied), there are points from  $P_A^b$  on  $H_i$ , with each such point being a boundary point of  $P_A^b$ . Of course, the

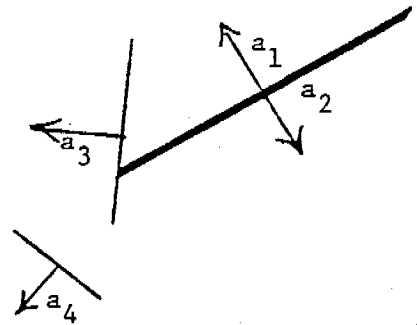
$$(\text{int } P_A^b) = \{x \in \mathbb{R}^n \mid Ax < b\},$$

a set that might be empty even when  $P_A^b$  is not empty. Finally, a polytope is just a polyhedral set that is "bounded"—in the sense that there is a finite upper bound on the "distance" of any of its points from the origin.

Two examples of a polyhedral set in  $\mathbb{R}^2$  are illustrated below.

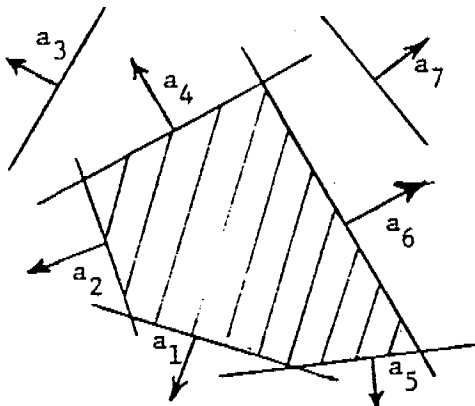


Polyhedral Set with Nonempty Interior

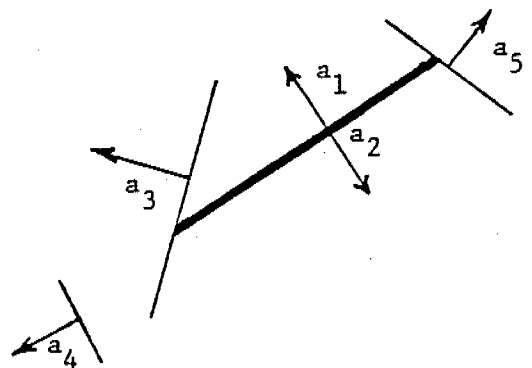


Polyhedral Set with Empty Interior

Two examples of a polytope in  $\mathbb{R}^2$  are illustrated below.



Polytope with Nonempty Interior



Polytope with Empty Interior

Needless to say, the feasible solution set for a consistent linear optimization problem is always polyhedral, and is sometimes polytopal. Moreover, the optimal solution set for a bounded problem is always polyhedral, and is usually polytopal (in fact, is almost always a singleton).

Exercise: 1. Prove the following facts.

(a) The intersect of a finite collection of polyhedral sets is either empty or polyhedral,

(b) The intersect of a finite collection of polytopes is either empty or polytopal.

2. Show that Statements (a) and (b) of Exercise 1 become false when the word "intersect" is replaced by the work "union."

3. Show that the complement of a polyhedral set is never polyhedral; and show that the complement of a polytope is never polytopal. [Hint: Use the fact that the complement of a closed set is open, as well as the fact that the complement of a bounded set is unbounded.]

A more detailed study of polyhedral sets and polytopes requires the following fundamental definition.

Definition. A linear combination  $\delta_1 x^1 + \delta_2 x^2$  of points  $x^1, x^2 \in \mathbb{R}^n$  is:

(i) an affine combination when

$$\delta_1 + \delta_2 = 1,$$

(ii) a convex combination when

$$\delta_1 + \delta_2 = 1 \text{ with } 0 \leq \delta_1, \delta_2 \leq 1$$

Since the equation  $\delta_1 + \delta_2 = 1$  has the general parametric solution  $\delta_1 = 1 - s$  and  $\delta_2 = s$  for  $s \in \mathbb{R}$ , we see that

$$\delta_1 x^1 + \delta_2 x^2 = (1-s)x^1 + sx^2 = x^1 + s(x^2 - x^1).$$

Consequently, the set of all affine combinations of distinct points  $x^1, x^2 \in \mathbb{R}^n$  is simply the whole "line" determined by  $x^1$  and  $x^2$ , while the set of all convex combinations of  $x^1$  and  $x^2$  is just the "line segment" joining  $x^1$  and  $x^2$ . Obviously, each convex combination of  $x^1$  and  $x^2$  is an affine combination of  $x^1$  and  $x^2$ , but each affine combination of  $x^1$  and  $x^2$  is a convex combination of  $x^1$  and  $x^2$  only when  $x^1 = x^2$  (in which case all such combinations  $\delta_1 x^1 + \delta_2 x^2 = \delta_1 x^1 + \delta_2 x^1 = (\delta_1 + \delta_2) x^1 = x^1 = x^2$ ).

The preceding definition leads in a natural way to the following definition.

Definition. A nonempty set  $S \subseteq \mathbb{R}^n$  is:

- (i) affine (or "flat" or an "affine variety" or a "linear variety") if it is closed under affine combinations (i.e., if  $\delta_1 x^1 + \delta_2 x^2 \in S$  when  $x^1, x^2 \in S$  and  $\delta_1 + \delta_2 = 1$ ),
- (ii) convex if it is closed under convex combinations (i.e., if  $\delta_1 x^1 + \delta_2 x^2 \in S$  when  $x^1, x^2 \in S$  and  $\delta_1 + \delta_2 = 1$  with  $0 \leq \delta_1, \delta_2 \leq 1$ ).

It is obvious that affine sets are convex, but convex sets need not be affine (with the latter fact being easy to establish by examples, as well as by the following exercise).

- Exercise:
1. Prove that hyperplanes are affine (and hence convex).
  2. Prove that closed halfspaces are convex, but not affine.
  3. Prove that open halfspaces are convex, but not affine.
  4. Give an example of a halfplane that is not convex  
(by an appropriate diagram, if desired). [HINT: Consider those that are neither closed nor open.] Can you give an example of a halfline that is not convex? If not, why not?

Affine combinations and convex combinations of more than two points occur in a natural way in the study of affine sets and convex sets. The definition of such combinations is an obvious extension of the previously given definition for two points. Moreover, the following theorem shows how such combinations can be interpreted in terms of combinations of just two points. The additional flexibility provided by this theorem is needed to eventually clarify the nature of affine sets and convex sets.

Theorem. Given an integer  $p > 2$ , each affine combination  $\sum_{i=1}^p \delta_i x^i$  of  $p$  points  $x^1, x^2, \dots, x^p \in \mathbb{R}^n$  (for which  $\sum_{i=1}^p \delta_i = 1$ ) is an affine combination of one of these points with an affine combination of the remaining  $p-1$  points; and each convex combination  $\sum_{i=1}^p \delta_i x^i$  of these  $p$  points (for which  $\sum_{i=1}^p \delta_i = 1$  with  $0 \leq \delta_1, \delta_2, \dots, \delta_p \leq 1$ ) is a convex combination of one of these points with a convex combination of the remaining  $p-1$  points. Conversely, each affine

combination of one of these points with an affine combination of the remaining  $p - 1$  points is an affine combination of all  $p$  points; and each convex combination of one of these points with a convex combination of the remaining  $p - 1$  points is a convex combination of all  $p$  points.

Proof: An affine combination  $\sum_{i=1}^p \delta_i x^i$  involves at least one  $\delta_i \neq 1$ ,

because  $\sum_{i=1}^p \delta_i = 1$  and  $p > 1$ . In particular, the option of renumbering

the points  $x^1, x^2, \dots, x^p$  lets us assume, without loss of generality,

that  $\delta_p \neq 1$ . This assumption in turn implies that  $\sum_{j=1}^{p-1} \delta_j \neq 0$ , because

$\sum_{j=1}^{p-1} \delta_j + \delta_p = 1$ . Consequently,

$$\sum_{i=1}^p \delta_i x^i = \left[ \sum_{j=1}^{p-1} \delta_j \right] \left[ \sum_{i=1}^{p-1} (\delta_i / \sum_{j=1}^{p-1} \delta_j) x^i \right] + \delta_p x^p.$$

This equation shows that  $\sum_{i=1}^p \delta_i x^i$  is an affine combination of  $x^p$

with an affine combination  $\sum_{i=1}^{p-1} (\delta_i / \sum_{j=1}^{p-1} \delta_j) x^i$ , because  $\left[ \sum_{j=1}^{p-1} \delta_j \right] + \delta_p = 1$

and  $\sum_{i=1}^{p-1} (\delta_i / \sum_{j=1}^{p-1} \delta_j) = 1$ .

Now, a convex combination  $\sum_{i=1}^p \delta_i x^i$  is a fortiori an affine

combination and hence can be represented, without loss of generality, by the preceding displayed equation. That equation shows that  $\sum_{i=1}^p \delta_i x^i$  is

also a convex combination of  $x^p$  with a convex combination

$\sum_{i=1}^{p-1} (\delta_i / \sum_{j=1}^{p-1} \delta_j) x^i$ , because the condition  $0 \leq \delta_1, \delta_2, \dots, \delta_p \leq 1$

clearly implies that  $0 \leq \left[ \sum_{j=1}^{p-1} \delta_j \right], \delta_p \leq 1$  and  $0 \leq (\delta_1 / \sum_{j=1}^{p-1} \delta_j), \dots,$

$$(\delta_{p-1} / \sum_{j=1}^{p-1} \delta_j) \leq 1.$$

Conversely, an affine combination of one of these points, say  $x_1$ , with an affine combination of the remaining  $p - 1$  points  $x^2, x^3, \dots, x^p$  must have the form  $\delta_1 x^1 + (1 - \delta_1) \left( \sum_{i=2}^p \gamma_i x^i \right)$  where  $\sum_{i=2}^p \gamma_i = 1$ . Letting  $\delta_i \triangleq (1 - \delta_1) \gamma_i$  for  $i = 2, 3, \dots, p$ , we see that

$$\delta_1 x^1 + (1 - \delta_1) \left( \sum_{i=2}^p \gamma_i x^i \right) = \sum_{i=1}^p \delta_i x^i,$$

which is an affine combination of all  $p$  points because  $\sum_{i=1}^p \delta_i = \delta_1 +$

$$(1 - \delta_1) \sum_{i=2}^p \gamma_i = \delta_1 + (1 - \delta_1) = 1.$$

Now, a convex combination of  $x^1$  with a convex combination of  $x^2, x^3, \dots, x^p$  is a fortiori an affine combination of  $x^1$  with an affine combination of  $x^2, x^3, \dots, x^p$  and hence can be represented by the preceding displayed equation. That equation shows that  $\delta_1 x^1 + (1 - \delta_1) \left( \sum_{i=2}^p \gamma_i x^i \right)$  is also a convex combination of all  $p$  points, because the condition  $0 \leq \delta_1 \leq 1$  and the condition  $0 \leq \gamma_2, \gamma_3, \dots, \gamma_p \leq 1$  clearly imply that  $0 \leq \delta_1, \delta_2, \dots, \delta_p \leq 1$ . q.e.d.

We already know that the set of all affine combinations of distinct points  $x^1, x^2 \in \mathbb{R}^3$  is the line determined by  $x^1$  and  $x^2$ . Consequently, the preceding theorem with  $p = 3$  implies that the set of all affine combinations of noncollinear points  $x^1, x^2, x^3 \in \mathbb{R}^3$  is the plane determined by  $x^1, x^2$  and  $x^3$ . Moreover, the preceding theorem with  $p = 4$  then implies that the set of all affine combinations of nonplanar points  $x^1, x^2, x^3, x^4 \in \mathbb{R}^3$  is all of  $\mathbb{R}^3$ .

Similarly, we already know that the set of all convex combinations of distinct points  $x^1, x^2 \in \mathbb{R}^3$  is the line segment joining  $x^1$  and  $x^2$ . Consequently, the preceding theorem with  $p = 3$  implies that the set of all convex combinations of noncollinear points  $x^1, x^2, x^3 \in \mathbb{R}^3$  is the triangular disk determined by  $x^1, x^2$  and  $x^3$ . Moreover, the preceding theorem with  $p = 4$  then implies that the set of all convex combinations of nonplanar points  $x^1, x^2, x^3, x^4 \in \mathbb{R}^3$  is a tetrahedron. Needless to say, more complicated polytopes arise from this construction as  $p$  increases further.

Exercise: 1. How should the statements in each of the preceding two paragraphs be altered if the terms "distinct", "noncollinear" and "nonplanar" are replaced respectively by the terms "identical", "collinear but distinct" and "planar but noncollinear"?

The following corollary to the last theorem is important.

Corollary. Each nonempty affine set  $S \subseteq \mathbb{R}^n$  is closed under all finite affine combinations (i.e.,  $\sum_{i=1}^p \delta_i x^i \in S$  when  $x^1, x^2, \dots, x^p \in S$  and

$\sum_{i=1}^p \delta_i = 1$ ). Moreover, each nonempty convex set  $S \subseteq \mathbb{R}^n$  is closed

under all finite convex combinations (i.e.,  $\sum_{i=1}^p \delta_i x^i \in S$  when  $x^1, x^2, \dots, x^p \in S$  and  $\sum_{i=1}^p \delta_i = 1$  with  $0 \leq \delta_1, \delta_2, \dots, \delta_p \leq 1$ ).

Proof: By definition, both assertions are true for  $p = 2$ . Consequently, the last theorem implies that both assertions are true for  $p > 2$

(essentially by mathematical induction). Finally, both assertions are obviously true for  $p = 1$ . q.e.d.

The following theorem reduces the study of affine sets to the study of linear equations (and hence to the study of linear algebra).

Theorem. If a nonempty set  $S \subseteq \mathbb{R}^n$  is a "linear manifold" (i.e., if  $S$  is the solution set for a finite system  $Ax = b$  of linear equations), then  $S$  is affine (and hence convex); and conversely, if  $S$  is affine, then  $S$  is a linear manifold.

Proof. If  $Ax^i = b$  for  $i = 1, 2$ , then  $A(\delta_1 x^1 + \delta_2 x^2) = \delta_1 Ax^1 + \delta_2 Ax^2 = \delta_1 b + \delta_2 b = (\delta_1 + \delta_2)b = b$ ; so linear manifolds are affine.

Conversely, if  $S$  is affine (and nonempty), either  $S$  is just a singleton  $\{x^0\}$  or  $S$  has at least one additional point  $x^1$ . In the first case,  $S$  is the solution set for the finite system  $Ix = x^0$ , and hence is a linear manifold. In the second case, the theory of linear algebra asserts that at least one, and at most  $n$ , vectors  $x^1, \dots, x^p$  in  $S$  can be chosen so that the vectors  $x^1 - x^0, \dots, x^p - x^0$  are "maximally linearly independent". Then,  $x^0 + s_j(x^j - x^0) = (1 - s_j)x^0 + s_j x^j \in S$  for each  $s_j \in \mathbb{R}$  and for  $j = 1, \dots, p$ . Consequently,  $\sum_{j=1}^p \delta_j [x^0 + s_j(x^j - x^0)] \in S$ ; that is,  $x^0 + \sum_{j=1}^p \delta_j s_j (x^j - x^0) \in S$ . In particular, the vector  $x^0$  plus each vector in the vector space with basis vectors  $x^1 - x^0, \dots, x^p - x^0$  is in  $S$ . Moreover, if there were a vector  $x$  in  $S$  that is not in this

linear manifold, our assumption about the maximality of  $x^1 - x^0, \dots, x^p - x^0$  would be violated. q.e.d.

We already know that all linear manifolds are polyhedral, but many polyhedral sets are not linear (manifolds). Similarly, we know that all affine sets are convex, but many convex sets are not affine. With this information, one might hope that the preceding theorem would remain valid when the terms "linear manifold" and "affine" are replaced, respectively, by the less restrictive terms "polyhedral set" and "convex". The next theorem shows that this is indeed the case for the first assertion; and although it is not the case for the converse assertion, it would be the case if the term "polyhedral set" were broadened to include solution sets for infinite systems of linear inequalities (a broadening that would take us beyond the scope of finite linear systems, and hence this book).

Theorem. If a nonempty set  $S \subseteq \mathbb{R}^n$  is a polyhedral set (i.e., if  $S$  is the solution set for a finite system  $Ax \leq b$  of linear inequalities), then  $S$  is convex; but not conversely.

Proof. If  $Ax^i \leq b$  for  $i = 1, 2$ , then  $A(\delta_1 x^1 + \delta_2 x^2) = \delta_1 Ax^1 + \delta_2 Ax^2 \leq \delta_1 b + \delta_2 b = (\delta_1 + \delta_2) b = b$ ; so polyhedral sets are convex.

It is not hard to show that a circular disk in  $\mathbb{R}^2$  is convex, but not polyhedral; so the converse is not valid. q.e.d.

From the preceding theorem, we infer that polytopes are convex, as are the feasible solution sets and optimal solution sets for linear optimization problems. In particular then, each convex combination of

given optimal solutions (such as the basic optimal solutions produced by the simplex method) is optimal.

Exercise: 1. Prove the following facts:

(a) The intersect of a collection (either finite or infinite) of convex sets is either empty or convex,

(b) The intersect of a collection (either finite or infinite) of affine sets is either empty or affine.

2. Show that Statements (a) and (b) of Exercise 1 become false when the word "intersect" is replaced by the word "union."

3. Show that the complement of a convex set need not be convex; and show that the complement of an affine set is never affine.

We already know from the Farkas lemma and duality theory that cones play a significant role in the study of linear optimization. With a view toward using them in a more thorough study of linear optimization, we now develop their properties in a more thorough way.

Definition. A nonempty set  $K \subseteq \mathbb{R}^n$  is conical if it is closed under nonnegative scalar multiplication (i.e., if  $sx \in K$  when  $x \in K$  and  $s \geq 0$ ).

Since  $sx = 0$  when  $s = 0$ , we see that each cone contains the zero vector (sometimes termed its "vertex"). Moreover, a cone that contains

at least one nonzero vector  $x$  must contain at least one "ray", namely  $\{sx \mid s \geq 0\}$ . Such cones can clearly be viewed as the union of rays.

The origin  $\{0\}$  and the whole space  $\mathbb{R}^n$  are obviously cones -- sometimes termed the "trivial cones". Being nonempty and closed under linear combinations, each subspace of  $\mathbb{R}^n$  is, a fortiori, a cone, but not conversely (because a single ray is clearly a cone but not a subspace). Each cone contains the subspace  $\{0\}$ , and is said to be "pointed" if it contains no other subspace (i.e., if it does not contain  $-x$  when it contains a nonzero  $x$ ).

Exercise: 1. There are only a finite number of cones in  $\mathbb{R}^1$ .

- (a) Describe each one.
- (b) Which ones are subspaces?
- (c) Which ones are pointed?
- (d) Which ones are polyhedral?
- (e) Which ones are polytopal?
- (f) Which ones are affine?
- (g) Which ones are convex?

2. Prove the following facts:

(a) The intersect of a nonempty collection (either finite or infinite) of cones is a cone (and hence is nonempty).

(b) The union of a nonempty collection (either finite or infinite) of cones is a cone.

(c) The complement of a cone is never a cone.

3. Prove the following facts:

(a) The intersect of a finite nonempty collection of polyhedral cones is a polyhedral cone (and hence is nonempty).

(b) The intersect of a nonempty collection (either finite or infinite) of convex cones is a convex cone (and hence is nonempty).

4. Show that statements (a) and (b) of Exercise 3 become false when the word "intersect" is replaced by the word "union".

5. Describe the cones that are polytopal.

6. Prove that a set  $K \subseteq \mathbb{R}^n$  is an affine cone if, and only if,

$$K = \{x \mid Ax = 0\} \text{ for some matrix } A.$$

7. Prove that a set  $K \subseteq \mathbb{R}^n$  is a polyhedral cone if, and only if,

$$K = \{x \mid Ax \leq 0\} \text{ for some matrix } A \text{ with non-zero rows. [Hint: You}$$

should need no help in proving that a set  $K$  defined in terms of such

a matrix  $A$  is a polyhedral cone. To establish the converse, first

recall the definition of a polyhedral set  $P_A^b$ , and then use your knowledge

of cones to prove that  $0 \leq b$ . Finally, if a component of  $b$  is actually

positive, prove that the set is either not conical or the positive

component can be replaced by zero without altering the set.]

8. Prove that a polyhedral cone  $\{x \mid Ax \leq 0\}$  is pointed if, and only if, the null space of  $A$  is  $\{0\}$  (i.e.,  $\{x \mid Ax = 0\} = \{0\}$ ).

Conicality and convexity together are most conveniently characterized by the following theorem. In fact, this characterization is frequently used as a definition for convex cones (with the term convex frequently deleted when there is no need or desire to mention the nonconvex variety). Furthermore, this characterization is a natural extension of an analogous characterization for vector spaces (which we have already identified as a special, but very important, variety of convex cones).

Theorem. A nonempty set  $K \subseteq \mathbb{R}^n$  is a convex cone if, and only if, it is closed under non-negative linear combinations (i.e., if, and only if,  $s_1 x^1 + s_2 x^2 \in K$  when  $x^1, x^2 \in K$  and  $0 \leq s_1, s_2$ ).

Proof. A nonempty set  $K$  that is closed under nonnegative linear combinations is, a fortiori, closed under convex combinations (with the additional restriction  $s_1 + s_2 = 1$ ). It is also, a fortiori, closed under nonnegative scalar multiplication (with the additional restriction  $s_1 = 0$  or  $s_2 = 0$ ).

Conversely, we already know that a convex cone  $K$  must contain the zero vector, and hence it must contain  $s_1 x^1 + s_2 x^2$  when  $s_1$  and  $s_2$  are both zero. Moreover, when  $0 \leq s_1, s_2$  with  $s_1$  and  $s_2$  not both zero,  $s_1 x^1 + s_2 x^2 = (s_1 + s_2) \left[ \left\{ \frac{s_1}{s_1 + s_2} \right\} x^1 + \left\{ \frac{s_2}{s_1 + s_2} \right\} x^2 \right]$ .

This vector is clearly a positive multiple of a convex combination of  $x^1$  and  $x^2$ , and hence it is in  $K$  when  $x^1, x^2 \in K$ . q.e.d.

Nonnegative linear combinations of more than two points occur in a natural way in the study of convex cones. Given an integer  $p > 2$ , each nonnegative linear combination  $\sum_{i=1}^p s_i x^i$  of  $p$  points  $x^1, x^2, \dots, x^p \in \mathbb{R}^n$  (for which  $0 \leq s_1, s_2, \dots, s_p$ ) is obviously a nonnegative linear combination of one of these points with a nonnegative linear combination of the remaining  $p - 1$  points, and conversely. This observation provides the following useful corollary to the last theorem.

Corollary. Each nonempty convex cone  $K \subseteq \mathbb{R}^n$  is closed under all finite nonnegative linear combinations (i.e.,  $\sum_{i=1}^p s_i x^i \in K$  when  $x^1, x^2, \dots, x^p \in K$  and  $0 \leq s_1, s_2, \dots, s_p$ ).

Proof. Simply use the last theorem and preceding observation with mathematical induction on  $p$ .

q.e.d.

Affine sets, convex sets and convex cones have certain important properties in common. The following definition leads to such properties.

Definition. Given a nonempty set  $S \subseteq \mathbb{R}^n$ ,

- (i) its affine hull,  $(\text{aff } S)$ , is the intersect of all affine sets containing  $S$ ,
- (ii) its convex hull,  $(\text{conv } S)$ , is the intersect of all convex sets containing  $S$ ,
- (iii) its convex conical hull,  $(\text{cone } S)$ , is the intersect of all convex cones containing  $S$ .

Since we already know from a previous exercise that an intersect of affine sets is affine,  $(\text{aff } S)$  is affine (even when  $S$  is not affine). Moreover, being an intersect of all affine sets containing  $S$  clearly implies that  $(\text{aff } S)$  contains  $S$  and is itself contained within every other affine set containing  $S$ . In essence then,  $(\text{aff } S)$  is the "smallest affine set containing  $S$ ." Furthermore, similar arguments (left to the reader) show that  $(\text{conv } S)$  is the "smallest convex set containing  $S$ ," and  $(\text{cone } S)$  is the "smallest convex cone containing  $S$ ." These observations help to establish the following important theorem.

Theorem. Given a nonempty set  $S \subseteq \mathbb{R}^n$ ,

- (i) its affine hull,  $(\text{aff } S)$ , is affine, and is identical to the set of all affine combinations of points in  $S$ .
- (ii) its convex hull,  $(\text{conv } S)$ , is convex, and is identical to the set of all convex combinations of points in  $S$ .
- (iii) its convex conical hull,  $(\text{cone } S)$ , is convex conical, and is identical to the set of all convex conical combinations of points in  $S$  (i.e., the set of all nonnegative linear combinations of points in  $S$ ).

Proof. In view of our previous observations, we need only establish the last assertions in parts (i), (ii) and (iii).

To establish the last assertion in part (i), first note that being an affine set containing  $S$  implies that  $(\text{aff } S)$  contains the set  $A(S)$  of all affine combinations of points in  $S$ . Consequently, we need only show that  $(\text{aff } S) \subseteq A(S)$  to establish the equation  $(\text{aff } S) = A(S)$  and hence complete a proof of part (i). Toward that end, note that  $A(S)$  obviously contains  $S$  and hence would also contain  $(\text{aff } S)$  if it happened to be affine.

Lemma. Given a nonempty set  $S \subseteq \mathbb{R}^n$ , the set  $A(S)$  of all affine combinations of points in  $S$  is an affine set.

Proof. Given  $x^1, x^2 \in A(S)$ , there are vectors  $y^i, z^i \in S$  and corresponding affine coefficients  $\alpha_i, \beta_i$  such that

$$x^1 = \sum_1^p \alpha_i y^i \text{ and } x^2 = \sum_1^q \beta_i z^i.$$

Given additional affine coefficients  $\delta_1, \delta_2$ , the affine combination

$$\delta_1 x^1 + \delta_2 x^2 = \delta_1 \sum_1^p \alpha_i y^i + \delta_2 \sum_1^q \beta_i z^i = \sum_1^p \delta_1 \alpha_i y^i + \sum_1^q \delta_2 \beta_i z^i$$

is in  $A(S)$  because  $p + q$  is finite and because

$$\sum_1^p \delta_1 \alpha_i + \sum_1^q \delta_2 \beta_i = \delta_1 \sum_1^p \alpha_i + \delta_2 \sum_1^q \beta_i = \delta_1 + \delta_2 = 1.$$

Proofs for the last assertions in parts (ii) and (iii) are analogous to the preceding proof of the last assertion in part (i). q.e.d

The following corollary is an elementary consequence of the preceding theorem.

Corollary A nonempty set  $S \subseteq \mathbb{R}^n$  is

- (i) affine if, and only if, it is identical to its affine hull, ( $\text{aff } S$ )
- (ii) convex if, and only if, it is identical to its convex hull, ( $\text{conv } S$ )
- (iii) convex conical if, and only if, it is identical to its convex conical hull, ( $\text{cone } S$ ).

The following Theorem is implied by the preceding Corollary and the "separation theory for convex sets". Since we do not have time to cover the separation theory for convex sets, we simply illustrate (in class) the interpretation of the following Theorem in  $\mathbb{R}^2$  (rather than present its formal proof).

Theorem The feasible solution set for a given linear optimization problem with only restricted variables is a convex polyhedral set whose elements  $x$  have the representation

$$x = \sum_k \delta_k v^k + \sum_k t_k d^k \text{ where } \delta_k \geq 0, \sum_k \delta_k = 1, \text{ and } t_k \geq 0,$$

where the vectors  $v^k$  and  $d^k$  are obtained from an extension of phase I of the simplex method applied to the given problem. In particular, the vectors  $v^k$  are all the basic feasible solutions, and the vectors  $d^k$  are all the "basic recession directions" -- generated by attempting to pivot in all columns of the terminal schema for phase I, as well as all columns of all subsequently generated feasible schemas. Such a feasible solution set is bounded (and hence a polytope) if, and only if, there are no basic recession directions  $d^k$ .

The preceding theorem provides the following important reformulation of the most general canonical problem P formulated earlier in these notes.

Problem P (reformulated)  $\phi = \min_{\delta, t} c(\sum_k \delta_k v^k + \sum_k t_k d^k) - d$  subject to

$$\sum_k \delta_k = 1,$$

$$\delta_k \geq 0, \text{ and } t_k \geq 0$$

The following theorem gives a solution to the preceding formulation of Problem P -- without direct recourse to phase II of the simplex method.

Theorem.  $\phi$  is finite (in that P is bounded) if, and only if,

$$c d^k \geq 0 \text{ for each } k;$$

in which case

$$\phi = \min_k c v^k - d$$

and

$$S^* = \{ \sum_k \delta_k v^k + \sum_k t_k d^k \mid \delta_k = 0 \text{ for each } k \text{ for which } c v^k - d > \phi, \text{ and } t_k = 0 \text{ for each } k \text{ for which } c v^k - d > 0 \}.$$

Proof. The nature of the constraints along with elementary linear algebra implies that

$$\phi = \min_{\delta, t} c(\sum_k \delta_k v^k + \sum_k t_k d^k) - d = (\min_{\delta} \sum_k \delta_k c v^k) + (\sum_k \min_{t_k} t_k c d^k) - d.$$

Moreover, the expression  $(\sum_k \min_{d^k} t_k c d^k)$  is clearly finite only if  $c d^k \geq 0$  for each  $k$  (because of the constraints  $t_k \geq 0$  for each  $k$ ); in which event it is clear that

$(\sum_k \min_{d^k} t_k c d^k) = 0$  and hence that

$$(\min_{\delta} \sum_k \delta_k c v^k) + (\sum_k \min_{d^k} t_k c d^k) - d = (\min_{\delta} \sum_k \delta_k c v^k) - d.$$

Furthermore, the expression  $(\min_{\delta} \sum_k \delta_k c v^k)$  is clearly always finite (because of the constraints  $\delta_k \geq 0$  for each  $k$  and  $\sum_k \delta_k = 1$ ); and it is also clear that

$$(\min_{\delta} \sum_k \delta_k c v^k) = \min_k c v^k.$$

Now, the preceding five displayed equations show that  $\phi = \min_k c v^k - d$  while the verification of the given equation for  $S^*$  is left as an exercise for the reader. q.e.d.

Since the preceding solution to Problem P uses all of its basic feasible solutions  $v^k$  and all of its basic recession directions  $d^k$  (generated by an extension of phase I of the simplex method), this solution is not more efficient than phase II of the simplex method -- which uses, and hence requires, the computation of only a relatively small number of the basic feasible solutions  $v^k$  and/or the basic recession directions  $d^k$ . However, this solution does provide additional insight into the nature of linear optimization, as well as a vehicle for establishing some of the more subtle characteristics of linear optimization.

A deeper study of parametric programming, post optimality analysis and post optimal sensitivity analysis requires the following extension of the concept of convexity from sets to functions.

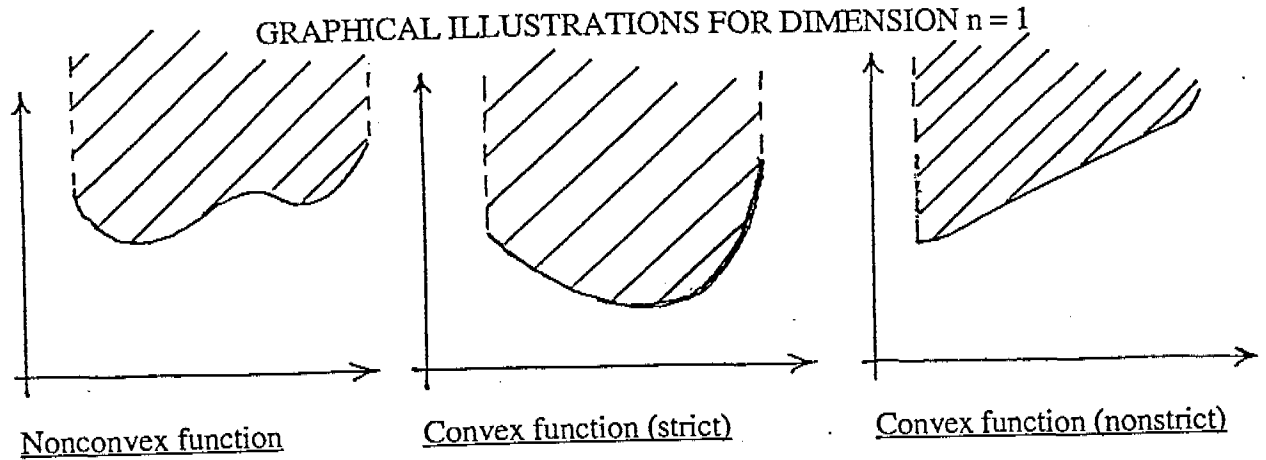
Definition. A function  $f$  with a domain  $S \subseteq \mathbb{R}^n$  is convex if its "epigraph"

$$(\text{epi } f) = \{(x, r) \in \mathbb{R}^{n+1} \mid x \in S \text{ and } f(x) \leq r\}$$

is a convex subset of  $\mathbb{R}^{n+1}$ . A convex function  $f$  is strictly convex if every point on its "graph"

$$(\text{gra } f) = \{(x, r) \in \mathbb{R}^{n+1} \mid x \in S \text{ and } f(x) = r\}$$

is an "extreme point" of its epigraph ( $\text{epi } f$ ) -- in that every point on its graph is not a convex combination of two other points in its epigraph.



Another important characterization of convex functions is given by the following theorem.

Theorem. A function  $f$  with a convex domain  $S \subseteq \mathbb{R}^n$  is convex if, and only if,

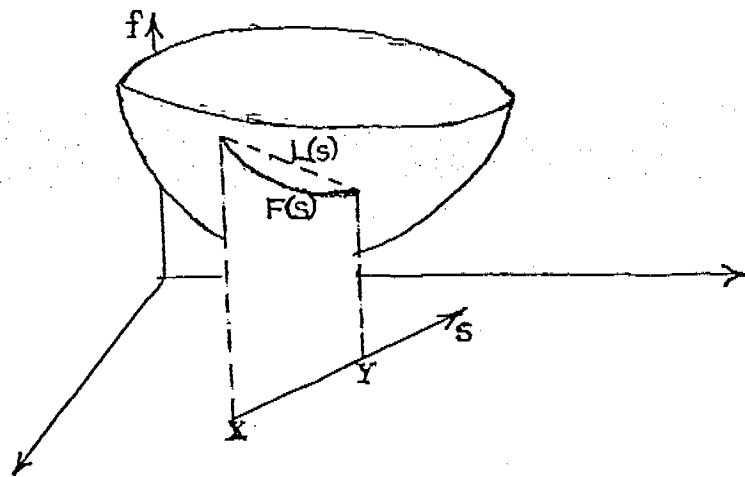
$$f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y) \quad \text{"Jensen's inequality"}$$

for each convex combination  $\alpha x + \beta y$  of points  $x, y \in S$  (for which  $\alpha, \beta \geq 0$  and  $\alpha + \beta = 1$ ). Moreover, a convex function  $f$  is strictly convex if Jensen's inequality is strictly satisfied when  $x \neq y$  and  $\alpha, \beta \neq 0$ .

Proofs for both the preceding theorem and the remaining theorems of this section will not be given, but their geometric interpretations in  $\mathbb{R}^2$  and  $\mathbb{R}^3$  will be illustrated, and their significance in linear optimization will be described. The following graphical illustration uses the scalar parameter  $s$  to describe the "convex weights"  $\alpha, \beta$  in the form  $\alpha = [1 - s]$  and  $\beta = s$ .

GRAPHICAL ILLUSTRATION FOR DIMENSION  $n = 2$

$$F(s) = f((1-s)x + sy) \leq [1-s]f(x) + sf(y) = L(s)$$

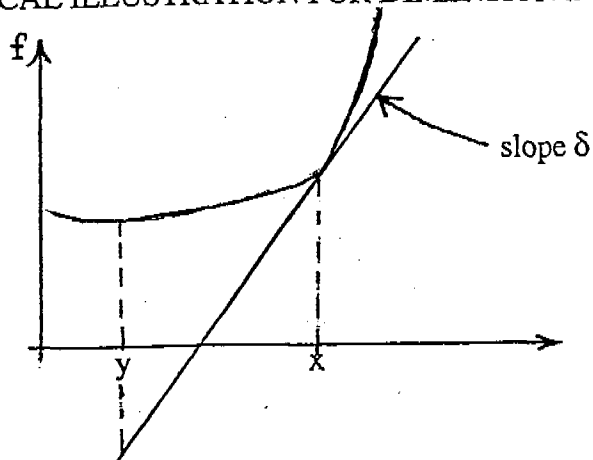


Observations and terminology. Since each affine function is obviously convex (why?) and since we already know that each polyhedral set is convex, each linear optimization problem can be viewed as an example of the minimization of a convex function over a convex feasible solution set. Such minimization problems are termed convex optimization problems.

A vector  $\delta \in \mathbb{R}^n$  is a subgradient for a given function  $f$  at a point  $x$  in its domain  $S \subseteq \mathbb{R}^n$  if

$$f(y) \geq f(x) + \delta(y - x) \text{ for each } y \in S.$$

GRAPHICAL ILLUSTRATION FOR DIMENSION  $n = 1$



Some important facts (made plausible by the preceding graphical illustration). A function  $f$  with a convex domain  $S \subseteq \mathbb{R}^n$  is convex if, and only if,  $f$  has at least one subgradient at each point  $x$  in the "relative interior" of  $S$ . If a function  $f$  is convex on its convex domain  $S \subseteq \mathbb{R}^n$  and if the relative

interior of  $S$  is actually the "interior" of  $S$ , then  $f$  is "differentiable" at a given point  $x$  in the interior of  $S$  if, and only if,  $f$  has exactly one subgradient at  $x$ ; in which case that subgradient of  $f$  is also the "gradient" of  $f$  at  $x$ .

The rest of this section gives additional facts about parametric programming and post-optimality analysis, as well as their relationships to duality, "equilibrium prices" and "Lagrange multipliers". This is done in the context of the family of linear optimization problems  $P(u)$  originally used in these notes to motivate economically the dual problem via "perturbations"  $u$  of the "resources"  $b$  available to a corporation for "production" of a "product mix"  $x$ . A concise restatement of that problem  $P(u)$  with a slack vector variable  $\eta$  is as follows.

Problem  $P(u)$  Determine

$$\phi(u) = \min_{x \in S(u)} c x - d$$

where

$$S(u) = \{x \mid Ax + \eta = b - u \text{ and } x, \eta \geq 0\}.$$

Observations This problem  $P(u)$  is clearly consistent if, and only if,  $u$  is in the "feasible perturbation set"

$$U = \{u \mid S(u) \neq \emptyset\} = \{b - Ax - \eta \mid x, \eta \geq 0\},$$

which is obviously a "finitely generated" (polyhedral) cone "with vertex at  $b$ ".

Moreover, the duality theory previously established clearly implies that either

$\phi(u) \equiv -\infty$  for each  $u \in U$  or  $\phi$  is finite on  $U$ , with the latter being the case if, and only if, the dual  $Q$  of problem  $P(0)$  is consistent.

The following theorem gives the nature of the function  $\phi$  with domain  $U$ , when  $\phi$  is finite on  $U$ .

Theorem If dual problem  $Q$  is consistent, its feasible solution set  $T$  has distinct basic feasible solutions  $y^1, y^2, \dots, y^t$  with corresponding slack vectors  $\xi^1, \xi^2, \dots, \xi^t$ , and  $U$  is the union of the

following  $t$  nontrivial, nonoverlapping, finitely generated, polyhedral cones  $U^q$  with common vertices at  $b$ :

$$U^q = \{b - Ax - \eta \mid x, \eta \geq 0; x_j = 0 \text{ when } \xi_j^q > 0; \eta_i = 0 \text{ when } y_i^q > 0\} \quad q = 1, 2, \dots, t.$$

Moreover,

$$\phi(u) = y^q(u - b) - d \text{ for } u \in U^q \quad q = 1, 2, \dots, t,$$

with  $\phi$  being a convex function on  $U$ .

The following definitions provide the concepts needed to thoroughly understand the relationship between the unperturbed problem  $P(0)$  and its perturbed versions  $P(u)$  for  $u \in U$ .

Definitions. A vector  $\alpha$  is an equilibrium price vector (or shadow price vector) for problem  $P(0)$  if

$$\alpha u - \phi(u) \leq \alpha 0 - \phi(0) \text{ for } u \in U.$$

A vector  $\lambda$  is a Lagrange multiplier vector for problem  $P(0)$  if

$$\min_{x, \eta \geq 0} \{cx + \lambda(Ax + \eta - b)\} = \phi(0).$$

A vector  $\delta$  is a subgradient vector for the function  $\phi$  at  $u = 0$  if

$$\phi(u) \geq \phi(0) + \delta u \text{ for } u \in U.$$

Theorem. The preceding three statements are equivalent when  $\alpha = \lambda = \delta$ . Moreover, the set of all such points  $\alpha$  (or  $\lambda$  or  $\delta$ ) is identical to the dual optimal solution set  $T^*$ .

Some important facts and implications. From the differential calculus, recall that if the function  $\phi$  is "differentiable" at a point  $0$  that is in the "interior" of its domain  $U$ , then its "directional derivative"

$$D_d \phi(0) = \lim_{s \rightarrow 0} [\phi(sd) - \phi(0)]/s$$

in the direction  $d$  (a "unit vector") at that point  $0$  is given by the formula

$$D_d \phi(0) = d \delta,$$

where  $\delta$  is the "gradient"  $\nabla \phi(0)$  of  $\phi$  at  $0$ . More generally, since  $\phi$  is a convex function on  $U$ , the corresponding "single-sided directional derivative"

$$D_d^+ \phi(0) = \lim_{s \rightarrow 0^+} [\phi(sd) - \phi(0)]/s$$

(in which  $s$  approaches  $0$  only from the "positive side"  $s > 0$ ) is given by the formula

$$D_d^+ \phi(0) = \max_{\delta \in \partial \phi(0)} d \delta,$$

where the "subgradient set"

$$\partial \phi(0) = \{ \delta \mid \phi(u) \geq \phi(0) + \delta u \text{ for } u \in U \}.$$

Finally, since the preceding theorem asserts that this subgradient set

$$\{ \delta \mid \phi(u) \geq \phi(0) + \delta u \text{ for } u \in U \} = T^*,$$

we conclude that

$$D_d^+ \phi(0) = \max_{\delta \in T^*} d \delta,$$

which shows that post-optimal sensitivity analysis in linear optimization is itself just linear optimization (because the dual optimal solution set  $T^*$  is polyhedral, and because  $T^*$  is easily obtained from the terminal schema or tableau of phase II of the simplex method).