2.3 Convex sets

In the subsequent discussion it will be helpful to use some ideas from the study of convex sets. A subset Ω of a real vector space is *convex* if $U, V \in \Omega$ implies $wU + (1-w)V \in \Omega$ for all $w \in [0,1]$. If $w_j \geq 0$ and $\sum_{j=1}^J w_j = 1$, say that $V = \sum_{j=1}^J w_j U_j \in \Omega$ is a *convex combination* of the points U_j. Figure 2.d shows several convex subsets of the $x-y$ plane.

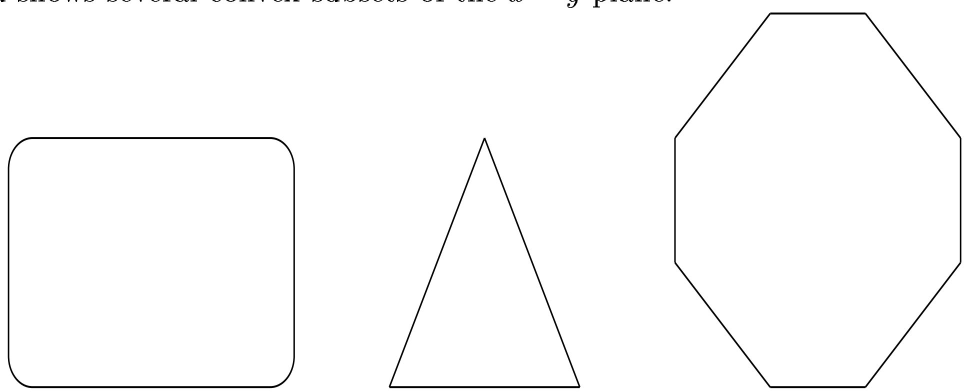

Figure 2d: Some convex sets

Lemma 2.4. *Suppose that $\Omega \subset \mathbb{R}^N$ is convex and $U_1, \ldots, U_J \in \Omega$. If $w_j \geq 0$ and $\sum_{j=1}^J w_j = 1$, then $\sum_{j=1}^J w_j U_j \in \Omega$.*

Proof. Without loss of generality, assume that each $w_j > 0$. The proof is by induction on J, the case $J = 1$ being trivial. For $J > 1$ write

$$\sum_{j=1}^J w_j U_j = w_1 U_1 + \sum_{j=2}^J w_j U_j = w_1 U_1 + [\sum_{j=2}^J w_j]\Big[\Big(\sum_{j=2}^J \frac{w_j}{\sum_{j=2}^J w_j}\Big) U_j\Big].$$

Since $\sum_{j=2}^J (w_j/[\sum_{j=2}^J w_j]) = 1$, the term $\sum_{j=2}^J (w_j/[\sum_{j=2}^J w_j]) U_j$ is a convex combination of $J-1$ elements of Ω, hence is in Ω by the induction hypothesis. By the definition of convexity

$$\sum_{j=1}^J w_j U_j = w_1 U_1 + \sum_{j=2}^J w_j U_j \in \Omega.$$

$\square$

Proposition 2.5. *The intersection of two or more convex sets is convex.*

Proof. Suppose the sets $\Omega_\alpha \subset \mathbb{R}^N$ are convex, and the vectors U and V are in all the sets Ω_α. Then for each w with $0 \le w \le 1$ the vectors $wU + (1-w)V$ are in each Ω_α, and so in the intersection. □

Recall that if $\mathcal{V}$ is a subspace of $\mathbb{R}^N$ and $X \in \mathbb{R}^N$ is a vector, then a set $\mathcal{W} = X + \mathcal{V} = \{X + V | V \in \mathcal{V}\}$ is called an *affine subspace* of $\mathbb{R}^N$. The dimension of $\mathcal{W}$ is defined to be the dimension of $\mathcal{V}$. If $dim(\mathcal{V}) = N - 1$, then $\mathcal{W}$ is called a *hyperplane.*

Proposition 2.6. *Affine subspaces of $\mathbb{R}^N$ are convex.*

Proof. Suppose $U_1 = X + V_1$ and $U_2 = X + V_2$ are vectors in the affine subspace $\mathcal{W} = X + \mathcal{V}$. For $0 \le w \le 1$ we have

$$wU_1 + (1-w)U_2 = X + wV_1 + (1-w)V_2 \in \mathcal{W}.$$

□

Proposition 2.7. *The feasible region for a linear program in ELP form is a convex set.*

Proof. We've just seen that an affine subspace is convex, and that the intersection of convex sets is convex. It only remains to show that the sector P consisting of vectors all of whose components are nonnegative is convex, If $U \in P$, $V \in P$, and $0 \le w \le 1$, then all components of $wU + (1-w)V$ are nonnegative, so P is convex. □

There is one more important notion needed from the theory of convex sets. Suppose $\Omega \subset \mathbb{R}^N$ is convex. Say that a point $Z \in \Omega$ is an *extreme point* if

$$Z = tU + (1-t)V, \quad 0 < t < 1, \quad U, V \in \Omega$$

implies $Z = U = V$.

As the next theorem shows, a closed and bounded convex set can be recovered from its extreme points. We won't give a proof of Theorem 2.8. Look at a picture in the plane to see that the result is plausible.

Theorem 2.8. *If $\Omega \subset \mathbb{R}^N$ is a closed and bounded convex set, which is contained in an affine subspace of dimension M, then every $X \in \Omega$ may be written as the convex combination of $M+1$ or fewer extreme points of Ω.*

2.4 Back to Linear Programming

In general there is no guarantee that the linear function we want to maximize in (ELP) actually has a maximum. The exercises provide a simple example with no maximum. Since the objective function g for an ELP is continuous, there is a simple test from calculus (or analysis) that guarantees the existence of a maximum.

Recall that a set $F \subset \mathbb{R}^N$ is bounded if there is some number $R > 0$ such that each $X \in F$ satisfies $\|X\| \le R$. A set F is said to be closed if F contains all its limit points. This means that if $X_n \in F$, $n = 1, 2, 3, \ldots$, and if $X_0 = \lim_{n\to\infty} X_n$, then $X_0 \in F$. If $h : \mathbb{R}^N \to \mathbb{R}$ is a continuous function, and $c \in \mathbb{R}$ then

$$F = h^{-1}(c) = \{X \in \mathbb{R}^N \mid h(X) = c\}$$

is closed. Also the positive sector P consisting of all vectors in $\mathbb{R}^N$ with all components nonnegative is a closed set. Moreover the intersection of closed sets is closed.

The result we need is that every continuous function f defined on a closed and bounded set F has a maximum (and minimum) on F. The next observation is an easy exercise. Some generalizations are sketched in the exercises.

Proposition 2.9. *Suppose one of the constraints*

$$\sum_{n=1}^{N} a_{mn} x_n = b_m,$$

in the ELP has all coefficients $a_{mn} > 0$. Then the feasible set F is closed and bounded.

For the next theorem we can relax the structural requirements of the feasible set.

Theorem 2.10. *If $K \subset \mathbb{R}^N$ is closed, bounded, and convex, and if $g : \mathbb{R}^N \to R$ is linear (or affine), then the maximum of g on K occurs at an extreme point.*

Proof. An induction proof will work. Let the maximum of g on K be achieved at X_0. By Theorem 2.8 any $X_0 \in K$ can be written as the convex combination of extreme points,

$$X_0 = \sum_{j=1}^{J} w_j U_j.$$

If X_0 is an extreme point, there is nothing to show. Suppose that the statement is true whenever X_0 is a convex combination of at most $J-1$ extreme points. Write

$$X_0 = w_1 U_1 + (1-w_1)\sum_{j=2}^{J} \frac{w_j}{1-w_1} U_j.$$

For $0 \le t \le 1$ let

$$X(t) = tU_1 + (1-t)\sum_{j=2}^{J} \frac{w_j}{1-w_1} U_j.$$

The function $G(t) = g(X(t))$ is the restriction of an affine function to a line, so $G(t)$ is affine. An affine function on $[0,1]$ must take its maximum at an endpoint, $t = 0, 1$. (Of course the function could be constant, thereby hitting the maximum on the whole segment.) Thus the maximum of g either occurs at the extreme point U_1, or at a point which is the convex combination of at most $J-1$ extreme points. □

With a little more thought we can see that for (ELP), any maximum which does exist must occur at an extreme point. As we will see now, a special role is played by the solutions of (ELP) with at least $N-L$ of the x_n equal to 0. Such a vector is called a *basic solution*, and if it also satisfies $x_k \ge 0$ for all $k = 1, \dots, N$, it is called a *basic feasible solution.*

Here is a helpful observation about points X in the feasible set F for an *ELP*.

Lemma 2.11. *Suppose X in the feasible set F for an ELP, and*

$$X = tU + (1-t)V, \quad 0 < t < 1, \quad U, V \in F.$$

If $X = (x_1, \dots, x_N)$ and $x_n = 0$, then $u_n = v_n = 0$.

Proof. Suppose $u_n \ne 0$. Since $U \in F$, we must then have $u_n > 0$. But then

$$0 = x_n = tu_n + (1-t)v_n, \quad 0 < t < 1,$$

forces $v_n < 0$, contradicting $V \in F$. □

Theorem 2.12. *If the objective function g for (ELP) has a maximum on the feasible set F, then this maximum occurs at an extreme point.*

Proof. Suppose the maximum for g on the feasible set F occurs at X_0. We claim that if X_0 is not an extreme point, then there is another point $Z \in F$, with $g(X_0) = g(Z)$, which has at least one more coordinate equal to 0 than X_0 does. To see this, write

$$X_0 = t_0 U + (1 - t_0)V, \quad 0 < t_0 < 1, \quad U, V \in F, \quad U \neq V.$$

Consider the affine function $G(t) = g(tU + (1 - t)V)$. Since $G(t)$ is maximal at $G(t_0) = g(X_0)$ and affine, it must be constant.

By the previous lemma, any coordinates which are 0 for X_0 must also be 0 for U and V. Since $U \neq V$, there must be at least one coordinate of $tU + (1 - t)V)$ which is positive at $t = t_0$, and which decreases as t increases (or decreases). Increase t until a coordinate which was not zero for X_0 is zero. Along this extension g remains constant, and we still satisfy $AX = B$.

Thus we can assume that X_0 has a maximal number of coordinates equal to 0. But this X_0 must be an extreme point. □

Finally, we want to use the ideas of this last proof to try to characterize the extreme points of F. The next result will show that the set of extreme points of F is in one to one correspondence with a subset of the set of choices of K components from N components. Thus the extreme points are finite in number and the only points we need to check to find the solution of (ELP).

Theorem 2.13. *Suppose X_0 satisfies $AX = B$ and has K components*

$$x_{n(1)}, \dots, x_{n(K)}$$

equal to zero, the rest being positive. Then X_0 is an extreme point for F if and only if the coefficient matrix for the system of equations

$$\begin{aligned} AX &= B, \\ x_{n(1)} &= 0, \\ &\vdots \\ x_{n(K)} &= 0, \end{aligned}$$

has rank N. Of course this means $K \geq N - L$.

Proof. Suppose first that X_0 has exactly K components equal to zero, and the coefficient matrix for the system of equations has rank N. Then the augmented system of equations has a unique solution, X_0. If $X_0 = tU + (1 -$

$t)V$ with $U, V \in F$, and $0 < t < 1$, then since U and V have components $n(1), \ldots, n(K)$ equal to 0, U, V satisfy the augmented system. By uniqueness $U = V = X_0$, so X_0 is an extreme point.

Suppose conversely that the coefficient matrix for the system of equations has rank less than N, (this will happen if $K < N - L$) and X_0 has exactly K coordinates equal to 0. Since X_0 is a solution of $AX_0 = B$, there is a nontrivial affine space of solutions to this system of equations. Write $X_0 = t_0 U + (1 - t_0)V$ for $U, V \neq X_0$ in this affine solution space. Since the remaining coordinates of X_0 are positive, they will remain positive for t near t_0. But then X_0 is not extreme. □

Here is an example to illustrate the use of this theorem.

Maximize $g = x_1 + 2x_2 + x_3 + 2x_4$ *subject to the constraints*

$$x_1 + 2x_2 + x_3 + x_4 = 4,$$

$$x_1 + 4x_2 + 3x_3 + x_4 = 7,$$

and all $x_k \geq 0$.

Since $N = 4$ and $L = 2$, extreme points have at least $N - L = 2$ coordinates equal to 0, and all coordinates nonnegative. Suppose $x_3 = x_4 = 0$. The remaining system of two equations has the solution $x_1 = 1$, $x_2 = 3/2$. This point is feasible, with the value of g equal to 4. Continue in this fashion, testing the $\binom{4}{2} = 6$ ways of choosing 2 coordinates equal to 0.