CHAPTER 10

Convex Analysis

This book is mostly about linear programming. However, this subject, important as it is, is just a subset of a larger subject called convex analysis. In this chapter, we shall give a brief introduction to this broader subject. In particular, we shall prove a few of the fundamental results of convex analysis and see that their proofs depend on some of the theory of linear programming that we have already developed.

1. Convex Sets

Given a finite set of points, $z_1, z_2, \ldots, z_n$, in $\mathbb{R}^m$, a point z in $\mathbb{R}^m$ is called a *convex combination* of these points if[1]

$$z = \sum_{j=1}^{n} t_j z_j,$$

where $t_j \geq 0$ for each j and $\sum_{j=1}^{n} t_j = 1$. It is called a *strict convex combination* if none of the t_j's vanish. For $n = 2$, the set of all convex combinations of two points is simply the line segment connecting them.

A subset S of $\mathbb{R}^m$ is called *convex* if, for every x and y in S, S also contains all points on the line segment connecting x and y. That is, $tx + (1-t)y \in S$, for every $0 < t < 1$. See Figure 10.1.

Certain elementary properties of convex sets are trivial to prove. For example, the intersection of an arbitrary collection of convex sets is convex. Indeed, let S_α, $\alpha \in I$, denote a collection of convex sets indexed by some set I. Then the claim is that $\cap_{\alpha \in I} S_\alpha$ is convex. To see this, consider an arbitrary pair of points x and y in the intersection. It follows that x and y are in each S_α. By the convexity of S_α it follows that S_α contains the line segment connecting x and y. Since each of these sets contains the line segment, so does their intersection. Hence, the intersection is convex.

Here is another easy one:

THEOREM 10.1. *A set C is convex if and only if it contains all convex combinations of points in C.*

[1] Until now we've used subscripts for the components of a vector. In this chapter, subscripts will be used to list sequences of vectors. Hopefully, this will cause no confusion.

R.J. Vanderbei, *Linear Programming*, International Series in Operations Research & Management Science 196, DOI 10.1007/978-1-4614-7630-6_10,

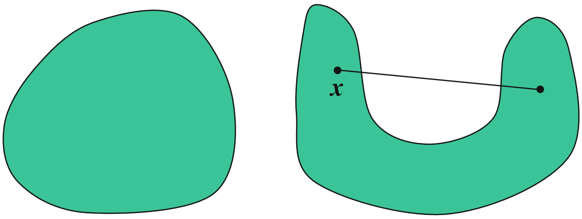

FIGURE 10.1. The set on the left is *convex*—for any pair of points in the set, the line segment connecting the two points is also contained in the set. The set on the right is *not convex*—there exists pairs of points, such as the x and y shown, for which the connecting line segment is not entirely in the set.

PROOF. Let C be a convex set. By definition, C contains all convex combinations of pairs of points in C. The first nontrivial step is to show that C contains all convex combinations of triples of points in C. To see this, fix z_1, z_2, and z_3 in C and consider

$$z = t_1 z_1 + t_2 z_2 + t_3 z_3,$$

where $t_j \geq 0$ for each j and $\sum_{j=1}^{3} t_j = 1$. If any of the t_j's vanish, then z is really just a convex combination of two points and so belongs to C. Hence, suppose that each of the t_j's is strictly positive. Rewrite z as follows:

$$\begin{aligned} z &= (1 - t_3)\left(\frac{t_1}{1 - t_3} z_1 + \frac{t_2}{1 - t_3} z_2\right) + t_3 z_3 \\ &= (1 - t_3)\left(\frac{t_1}{t_1 + t_2} z_1 + \frac{t_2}{t_1 + t_2} z_2\right) + t_3 z_3. \end{aligned}$$

Since C contains all convex combinations of pairs of points, it follows that

$$\frac{t_1}{t_1 + t_2} z_1 + \frac{t_2}{t_1 + t_2} z_2 \in C.$$

Now, since z is a convex combination of the two points $\frac{t_1}{t_1+t_2} z_1 + \frac{t_2}{t_1+t_2} z_2$ and z_3, both of which belong to C, it follows that z is in C. It is easy to see (pun intended) that this argument can be extended to an inductive proof that C contains all convex combinations of finite collections of points in C. Indeed, one must simply show that the fact that C contains all convex combinations of n points from C implies that it contains all convex combinations of $n + 1$ points from C. We leave the details to the reader.

Of course, proving that a set is convex if it contains every convex combination of its points is trivial: simply take convex combinations of pairs to get that it is convex. □

For each set S in $\mathbb{R}^m$ (not necessarily convex), there exists a smallest convex set, which we shall denote by conv(S), containing S. It is defined, quite simply, as the intersection of all convex sets containing S. From our discussion about intersections, it follows that this set is convex. The set conv(S) is called the *convex hull* of S. This definition can be thought of as a definition from the "outside," since it involves forming the intersection of a collection of sets that contain S. Our next theorem gives a characterization of convex sets from the "inside":

THEOREM 10.2. *The convex hull* conv(S) *of a set* S *in* $\mathbb{R}^m$ *consists precisely of the set of all convex combinations of finite collections of points from* S.

PROOF. Let H denote the set of all convex combinations of finite sets of points in S:

$$H = \left\{ z = \sum_{j=1}^{n} t_j z_j : n \geq 1, z_j \in S \text{ and } t_j > 0 \text{ for all } j, \text{ and } \sum_{j=1}^{n} t_j = 1 \right\}.$$

It suffices to show that (1) H contains S, (2) H is convex, and (3) every convex set containing S also contains H.

To see that H contains S, just take $n = 1$ in the definition of H.

To see that H is convex, fix two points x and y in H and a real number $0 < t < 1$. We must show that $z = tx + (1-t)y \in H$. The fact that $x \in H$ implies that $x = \sum_{j=1}^{r} p_j x_j$, for some $r \geq 1$, where $p_j > 0$ for $j = 1, 2, \ldots, r$, $\sum_{j=1}^{r} p_j = 1$, and $x_j \in S$ for $j = 1, 2, \ldots, r$. Similarly, the fact that y is in H implies that $y = \sum_{j=1}^{s} q_j y_j$, for some $s \geq 1$, where $q_j > 0$ for $j = 1, 2, \ldots, s$, $\sum_{j=1}^{s} q_j = 1$, and $y_j \in S$ for $j = 1, 2, \ldots, s$. Hence,

$$z = tx + (1-t)y = \sum_{j=1}^{r} t p_j x_j + \sum_{j=1}^{s} (1-t) q_j y_j.$$

Since the coefficients $(tp_1, \ldots, tp_r, (1-t)q_1, \ldots, (1-t)q_s)$ are all positive and sum to one, it follows that this last expression for z is a convex combination of $r + s$ points from S. Hence, z is in H. Since x and y were arbitrary points in H and t was an arbitrary real number between zero and one, the fact that $z \in H$ implies that H is convex.

It remains simply to show that H is contained in every convex set containing S. Let C be such a set (i.e., convex and containing S). From Theorem 10.1 and the fact that C contains S, it follows that C contains all convex combinations of points in S. Hence, C contains H. □

2. Carathéodory's Theorem

In the previous section, we showed that the convex hull of a set S can be constructed by forming all convex combinations of finite sets of points from S. In 1907, Carathéodory showed that it is not necessary to use all finite sets. Instead, $m + 1$ points suffice:

THEOREM 10.3. *The convex hull* conv(S) *of a set* S *in* $\mathbb{R}^m$ *consists of all convex combinations of* $m + 1$ *points from* S*:*

$$\text{conv}(S) = \left\{ z = \sum_{j=1}^{m+1} t_j z_j : z_j \in S \textit{ and } t_j \geq 0 \textit{ for all } j, \textit{ and } \sum_{j} t_j = 1 \right\}.$$

PROOF. Let H denote the set on the right. From Theorem 10.2, we see that H is contained in conv(S). Therefore, it suffices to show that every point in conv(S) belongs to H. To this end, fix a point z in conv(S). By Theorem 10.2, there exists a collection of, say, n points $z_1, z_2, \ldots, z_n$ in S and associated nonnegative multipliers $t_1, t_2, \ldots, t_n$ summing to one such that

$$z = \sum_{j=1}^{n} t_j z_j. \tag{10.1}$$

Let A denote the matrix consisting of the points $z_1, z_2, \ldots, z_n$ as the columns of A:

$$A = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}.$$

Also, let x^* denote the vector consisting of the multipliers $t_1, t_2, \ldots, t_n$:

$$x^* = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}.$$

Finally, let $b = z$. Then from (10.1), we see that x^* is feasible for the following linear programming problem:

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Ax = b \\ & e^T x = 1 \\ & x \geq 0 . \end{array} \tag{10.2}$$

The fundamental theorem of linear programming (Theorem 3.4) tells us that every feasible linear program has a basic feasible solution. For such a solution, only the basic variables can be nonzero. The number of basic variables in (10.2) coincides with the number of equality constraints; that is, there are at most $m + 1$ variables that are nonzero. Hence, this basic feasible solution corresponds to a convex combination of just $m + 1$ of the original n points. (See Exercise 10.5.) □

It is easy to see that the number $m + 1$ is the best possible. For example, the point $(1/(m+1), 1/(m+1), \ldots, 1/(m+1))$ in $\mathbb{R}^m$ belongs to the convex hull of the $m + 1$ points $e_1, e_2, \ldots, e_m, 0$ but is not a convex combination of any subset of them.

3. The Separation Theorem

We shall define a *halfspace* of $\mathbb{R}^n$ to be any set given by a single (nontrivial) linear inequality:

$$\{x \in \mathbb{R}^n : \sum_{j=1}^{n} a_j x_j \leq b\}, \qquad (a_1, a_2, \ldots, a_n) \neq 0. \tag{10.3}$$

Every halfspace is convex. To see this, suppose that $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ both satisfy the linear inequality in (10.3). Fix t between zero and one. Then both t and $1 - t$ are nonnegative, and so multiplying by them preserves the direction of inequality. Therefore, multiplying $\sum_j a_j x_j \le b$ by t and $\sum_j a_j y_j \le b$ by $1-t$ and then adding, we get

$$\sum_j a_j \left(t x_j + (1-t) y_j\right) \le b.$$

That is, $tx + (1-t)y$ also satisfies the inequality defining the halfspace.

If we allow the vector of coefficients $(a_1, a_2, \ldots, a_n)$ in the definition of a halfspace to vanish, then we call the set so defined a *generalized halfspace*. It is easy to see that every generalized halfspace is simply a halfspace, all of $\mathbb{R}^n$, or the empty set. Also, every generalized halfspace is clearly convex.

A *polyhedron* is defined as the intersection of a finite collection of generalized halfspaces. That is, a polyhedron is any set of the form

$$\left\{ x \in \mathbb{R}^n : \sum_{j=1}^n a_{ij} x_j \le b_i, i = 1, 2, \ldots, m \right\}.$$

Every polyhedron, being the intersection of a collection of convex sets, is convex.

The following theorem is called the *Separation Theorem* for polyhedra.

THEOREM 10.4. *Let P and $\tilde{P}$ be two disjoint nonempty polyhedra in $\mathbb{R}^n$. Then there exist disjoint halfspaces H and $\tilde{H}$ such that $P \subset H$ and $\tilde{P} \subset \tilde{H}$.*

PROOF. Suppose that P and $\tilde{P}$ are given by the following systems of inequalities:

$$P = \{x : Ax \le b\},$$
$$\tilde{P} = \{x : \tilde{A}x \le \tilde{b}\}.$$

The disjointness of P and $\tilde{P}$ implies that there is no solution to the system

$$\begin{bmatrix} A \\ \tilde{A} \end{bmatrix} x \le \begin{bmatrix} b \\ \tilde{b} \end{bmatrix}. \tag{10.4}$$

To continue the proof, we need a result known as Farkas' Lemma, which says that $Ax \le b$ has no solutions if and only if there is an m-vector y such that

$$\begin{aligned} A^T y &= 0 \\ y &\ge 0 \\ b^T y &< 0. \end{aligned}$$

We shall prove this result in the next section. For now, let us apply Farkas' Lemma to the situation at hand. Indeed, the fact that there are no solutions to (10.4) implies that there exists a vector, which we shall write in block form as

$$\begin{bmatrix} y \\ \tilde{y} \end{bmatrix},$$

such that

$$\begin{bmatrix} A^T & \tilde{A}^T \end{bmatrix} \begin{bmatrix} y \\ \tilde{y} \end{bmatrix} = A^T y + \tilde{A}^T \tilde{y} = 0 \tag{10.5}$$

$$\begin{bmatrix} y \\ \tilde{y} \end{bmatrix} \geq 0 \tag{10.6}$$

$$\begin{bmatrix} b^T & \tilde{b}^T \end{bmatrix} \begin{bmatrix} y \\ \tilde{y} \end{bmatrix} = b^T y + \tilde{b}^T \tilde{y} < 0. \tag{10.7}$$

From the last condition, we see that either $b^T y < 0$ or $\tilde{b}^T \tilde{y} < 0$ (or both). Without loss of generality, we may assume that

$$b^T y < 0.$$

Farkas' Lemma (this time applied in the other direction) together with the nonemptiness of P now implies that

$$A^T y \neq 0.$$

Put

$$H = \left\{ x : (A^T y)^T x \leq b^T y \right\} \qquad \text{and} \qquad \tilde{H} = \left\{ x : (A^T y)^T x \geq -\tilde{b}^T \tilde{y} \right\}.$$

These sets are clearly halfspaces. To finish the proof, we must show that they are disjoint and contain their corresponding polyhedra.

First of all, it follows from (10.7) that H and $\tilde{H}$ are disjoint. Indeed, suppose that $x \in H$. Then $(A^T y)^T x \leq b^T y < -\tilde{b}^T \tilde{y}$, which implies that x is not in $\tilde{H}$.

To show that $P \subset H$, fix x in P. Then $Ax \leq b$. Since $y \geq 0$ (as we know from (10.6)), it follows then that $y^T Ax \leq y^T b$. But this is exactly the condition that says that x belongs to H. Since x was an arbitrary point in P, it follows that $P \subset H$.

Showing that $\tilde{P}$ is a subset of $\tilde{H}$ is similar. Indeed, suppose that $x \in \tilde{P}$. Then $\tilde{A}x \leq \tilde{b}$. Multiplying on the left by $-\tilde{y}^T$ and noting that $\tilde{y} \geq 0$, we see that $-\tilde{y}^T \tilde{A}x \geq -\tilde{y}^T \tilde{b}$. But from (10.5) we see that $-\tilde{y}^T \tilde{A}x = y^T Ax$, and so this last inequality is exactly the condition that $x \in \tilde{H}$. Again, the arbitrariness of $x \in \tilde{P}$ implies that $\tilde{P} \subset \tilde{H}$, and the proof is complete. □

4. Farkas' Lemma

The following result, known as Farkas' Lemma, played a fundamental role in the proof of the separation theorem of the preceding section (Theorem 10.4). In this section, we state it formally as a lemma and give its proof.

LEMMA 10.5. *The system $Ax \leq b$ has no solutions if and only if there is a y such that*

$$\begin{aligned} A^T y &= 0 \\ y &\geq 0 \\ b^T y &< 0. \end{aligned} \tag{10.8}$$

PROOF. Consider the linear program

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & Ax \;\leq\; b \end{array}$$

and its dual

$$\begin{array}{lrcl} \text{minimize} & b^T y & & \\ \text{subject to} & A^T y & = & 0 \\ & y & \geq & 0\,. \end{array}$$

Clearly, the dual is feasible (just take $y = 0$). So if the primal is feasible, then the dual is bounded. Also, if the primal is infeasible, the dual must be unbounded. That is, the primal is infeasible if and only if the dual is unbounded. To finish the proof, we claim that the dual is unbounded if and only if there exists a solution to (10.8). Indeed, suppose that the dual is unbounded. The dual simplex method is guaranteed to prove that it is unbounded, and it does so as follows. At the last iteration, a step direction Δy is computed that preserves feasibility, i.e.,

$$A^T \Delta y = 0,$$

is a descent direction for the objective function, i.e.,

$$b^T \Delta y < 0,$$

and is a direction for which the step length is unbounded, i.e.,

$$\Delta y \geq 0.$$

But these three properties show that Δy is the solution to (10.8) that we were looking for. Conversely, suppose that there is a solution to (10.8). Call it Δy. It is easy to see that starting from $y = 0$, this step direction provides an unbounded decrease in the objective function. This completes the proof. $\square$

5. Strict Complementarity

In this section, we consider the usual inequality-form linear programming problem, which we write with its slack variables shown explicitly:

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Ax + w = b \\ & \quad x, w \geq 0. \end{array} \tag{10.9}$$

As we know, the dual can be written as follows:

$$\begin{array}{ll} \text{minimize} & b^T y \\ \text{subject to} & A^T y - z = c \\ & \quad y, z \geq 0. \end{array} \tag{10.10}$$

In our study of duality theory in Chapter 5, we saw that every pair of optimal solutions to these two problems possesses a property called complementary slackness. If (x^*, w^*) denotes an optimal solution to the primal and (y^*, z^*) denotes an optimal solution to the dual, then the complementary slackness theorem says that, for each $j = 1, 2, \ldots, n$, either $x_j^* = 0$ or $z_j^* = 0$ (or both) and, for each $i = 1, 2, \ldots, m$, either $y_i^* = 0$ or $w_i^* = 0$ (or both). In this section, we shall prove that there are optimal pairs of solutions for which the parenthetical "or both" statements don't happen. That is, there are optimal solutions for which exactly one member of each pair (x_j^*, z_j^*) vanishes and exactly one member from each pair (y_i^*, w_i^*) vanishes. In such cases, we say that the optimal solutions are *strictly complementary* to each other. The strictness of the complementary slackness is often expressed by saying that $x^* + z^* > 0$ and $y^* + w^* > 0$.[2]

As a warm-up, we prove the following theorem.

THEOREM 10.6. *If both the primal and the dual have feasible solutions, then there exists a primal feasible solution $(\bar{x}, \bar{w})$ and a dual feasible solution $(\bar{y}, \bar{z})$ such that $\bar{x} + \bar{z} > 0$ and $\bar{y} + \bar{w} > 0$.*

PROOF. If there is a feasible primal solution $\bar{x}$ for which $\bar{x}_j > 0$, then it doesn't matter whether there is a feasible dual solution whose jth slack variable is strictly positive. But what about indices j for which $x_j = 0$ for every feasible solution? Let j be such an index. Consider the following linear programming problem:

$$\begin{array}{ll} \text{maximize} & x_j \\ \text{subject to} & Ax \leq b \\ & x \geq 0. \end{array} \tag{10.11}$$

This problem is feasible, since its constraints are the same as for the original primal problem (10.9). Furthermore, it has an optimal solution (the corresponding objective function value is zero). The dual of (10.11) is:

$$\begin{array}{ll} \text{minimize} & b^T y \\ \text{subject to} & A^T y \geq e_j \\ & y \geq 0. \end{array}$$

By the strong duality theorem, the dual has an optimal solution, say y'. Letting z' denote the corresponding slack variable, we have that

$$\begin{aligned} A^T y' - z' &= e_j \\ y', z' &\geq 0. \end{aligned}$$

Now, let y be any feasible solution to (10.10) and let z be the corresponding slack variable. Then the above properties of y' and z' imply that $y + y'$ is feasible for (10.10) and its slack is $z + z' + e_j$. Clearly, for this dual feasible solution we have that the jth component of its vector of slack variables is at least 1. To summarize, we have shown that, for each j, there exists a primal feasible solution, call it $(x^{(j)}, w^{(j)})$, and a dual feasible solution, call it $(y^{(j)}, z^{(j)})$, such that $x_j^{(j)} + z_j^{(j)} > 0$. In the same way, one can exhibit primal and dual feasible solutions for which each individual dual variable and its corresponding primal slack add to a positive number. To complete the proof, we now form a strict convex combination of these $n + m$ feasible solutions. Since the feasible region for a linear programming problem is convex, these convex combinations preserve primal and dual feasibility. Since the convex combination is strict, it follows that every primal variable and its dual slack add to a strictly positive number as does every dual variable and its primal slack. ☐

[2]Given any vector ξ, we use the notation $\xi > 0$ to indicate that every component of ξ is strictly positive: $\xi_j > 0$ for all j.

A variable x_j that must vanish in order for a linear programming problem to be feasible is called a *null variable*. The previous theorem says that if a variable is null, then its dual slack is not null.

The following theorem is called the *Strict Complementary Slackness Theorem*

THEOREM 10.7. *If a linear programming problem has an optimal solution, then there is an optimal solution* (x^*, w^*) *and an optimal dual solution* (y^*, z^*) *such that* $x^* + z^* > 0$ *and* $y^* + w^* > 0$.

We already know from the complementary slackness theorem (Theorem 5.1) that x^* and z^* are complementary to each other as are y^* and w^*. This theorem then asserts that the complementary slackness is strict.

PROOF. The proof is much the same as the proof of Theorem 10.6 except this time we look at an index j for which x_j vanishes in every *optimal* solution. We then consider the following problem:

$$\begin{array}{ll} \text{maximize} & x_j \\ \text{subject to} & Ax \leq b \\ & c^T x \geq \zeta^* \\ & x \geq 0, \end{array} \tag{10.12}$$

where ζ^* denotes the objective value of the optimal solution to the original problem. In addition to the dual variables y corresponding to the $Ax \leq b$ constraints, there is one more dual variable, call it t, associated with the constraint $c^T x \geq \zeta^*$. The analysis of problem (10.12) is similar to the analysis given in Theorem 10.6 except that one must now consider two cases: (a) the optimal value of t is strictly positive and (b) the optimal value of t vanishes. The details are left as an exercise (see Exercise 10.6). □

Exercises

10.1 Is $\mathbb{R}^n$ a polyhedron?

10.2 For each $b \in \mathbb{R}^m$, let $\xi^*(b)$ denote the optimal objective function value for the following linear program:

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Ax \leq b \\ & x \geq 0. \end{array}$$

Suppose that $\xi^*(b) < \infty$ for all b. Show that the function $\xi^*(b)$ is concave (a function f on $\mathbb{R}^m$ is called *concave* if $f(tx+(1-t)y) \geq tf(x)+(1-t)f(y)$ for all x and y in $\mathbb{R}^m$ and all $0 < t < 1$). *Hint: Consider the dual problem.*

10.3 Describe how one needs to modify the proof of Theorem 10.4 to get a proof of the following result:

Let P and $\tilde{P}$ be two disjoint polyhedra in $\mathbb{R}^n$. Then there exist disjoint generalized halfspaces H and $\tilde{H}$ such that $P \subset H$ and $\tilde{P} \subset \tilde{H}$.

10.4 Find a strictly complementary solution to the following linear programming problem and its dual:

$$\begin{array}{lrcl} \text{maximize} & 2x_1 + x_2 & & \\ \text{subject to} & 4x_1 + 2x_2 & \leq & 6 \\ & x_2 & \leq & 1 \\ & 2x_1 + x_2 & \leq & 3 \\ & x_1,\ x_2 & \geq & 0\,. \end{array}$$

10.5 There is a slight oversimplification in the proof of Theorem 10.3. Can you spot it? Can you fix it?

10.6 Complete the proof of Theorem 10.7.

10.7 *Interior solutions*. Prove the following statement: If a linear programming problem has feasible solutions and the set of feasible solutions is bounded, then there is a strictly positive dual feasible solution: $y > 0$ and $z > 0$. *Hint. It is easier to prove the equivalent statement: if a linear programming problem has feasible solutions and the dual has null variables, then the set of primal feasible solutions is an unbounded set.*

Notes

Carathéodory (1907) proved Theorem 10.3. Farkas (1902) proved Lemma 10.5. Several similar results were discovered by many others, including Gordan (1873), Stiemke (1915), Ville (1938), and Tucker (1956). The standard reference on convex analysis is Rockafellar (1970).