0.1 Linear Programming

0.1.1 Some examples

Here are some specific examples (borrowed in part from T. Ferguson's UCLA web site).

The Diet Problem

We have a possible diet consisting of M foods $F_1, \ldots, F_M$, which together contain N essential nutrients $E_1, \ldots, E_N$. Assume that c_j is the minimum daily requirement of N_j, and that b_i is the price of a unit of food F_i. Let a_{ij} be the amount of nutrient N_j in a unit of food F_i. Finally, let y_i be the daily intake of units of food F_i. The problem is to minimize

$$C = \sum_i b_i y_i$$

subject to

$$\sum_i a_{ij} y_i \geq c_j, \quad y_i \geq 0.$$

Transportation problem

We have ports or factories $P_1, \ldots, P_I$ which are the source of a particular commodity or product. There are J markets $M - 1, \ldots, M_J$ needing the product. Assume P_i produces amount s_i of product, and M_j needs amount r_j. Let y_{ij} be the amount of product sent from i to j, and let b_{ij} be the unit cost of shipping from i to j. Then the cost of a port to market plan is

$$C = \sum_{ij} b_{ij} y_{ij}.$$

We want to minimize cost C subject to the constraints

$$y_{ij} \geq 0, \quad \sum_i y_{ij} \geq r_j.$$

0.1.2 General discussion

For functions of several variables some of the simplest minimization problems for convex functions are linear programs. The typical linear programming problem asks for the min or max of a linear (affine) function

$$g(x_1, \dots, x_N) = \sum_{n=1}^{N} c_n x_n$$

subject to the positivity constraints

$$x_n \ge 0, \quad n = 1, \dots, N, \tag{0.1.1}$$

and linear equation or inequality constraints

$$\sum_{n=1}^{N} a_{mn} x_n = b_m, \quad m = 1, \dots, M_1, \quad \sum_{n=1}^{N} \alpha_{mn} x_n \le \beta_m, \quad m = 1, \dots, M_2, \tag{0.1.2}$$
$$\sum_{n=1}^{N} \gamma_{mn} x_n \le \delta_m, \quad m = 1, \dots, M_3.$$

The points in $\mathbb{R}^N$ satisfying the constraints define a convex set Ω for the problem. The set Ω is often called the *feasible region* or *feasilbe set*.

Notice that $\nabla g = (c_1, \dots, c_N)$ is not zero in any interesting case. Thus we never see the extreme value of g in the interior of the feasible region. In fact, in well posed problems the extreme values must occur at one of a finite number of vertices of Ω.

Linear programming problems are often computationally challenging. Consider the simple case where $\Omega \subset \mathbb{R}^N$ is the hypercube

$$\Omega = \{(x_1, \dots, x_N) \ |0 \le x_n \le 1\}.$$

The vertices are the points with each x_n equal to either 0 or 1. There are 2^N such points, so if $N \simeq 100$ there would be about 10^{30} vertices. Even for modest values of N, one needs a computer to tackle such problems. Perhaps that explains why the theory of linear programming did not develop until the WW2 period, when electronic computers were developed.

An approach called the *simplex algorithm* was developed almost immediately and remains popular. The simplex algorithm repeatedly tries to move from a vertex to its neighbors while improving $g(x_1, \ldots, x_N)$. In practice the algorithm is almost always efficient, but in some special cases this method can get stuck or run for times that grow exponentially in N. Because many practical optimization problems can be posed as linear programming problems, this method is very important for applications. As recently as 2001 (Dan Spielman and Shang-Hua Teng) there were breakthroughs in understanding why the simplex algorithm is so effective in practice.

Another response to the computational complexity issues was to look for alternative algorithms guaranteed to be efficient. A method developed by Khachiyan (1979) provided a theoretical breakthrough, although it was not considered practical. A subsequent approach by Karmarkar (1984), *Interior point methods*, is considered a practical competitor to the simplex method, and offers new approaches to more general problems. There is also an extensive literature. The book Linear Programming, by H. Karloff (Birkhäuser 1991) looks like a good reference on the subject. It appears to be concise, well written, and reasonably up to date. In addition to covering the commonly used simplex algorithm, this book also treats interior point methods.

0.1.3 More sample problems

Function approximation

There are several applications of linear programming that are not immediately obvious. One interesting one has to do with approximation of functions. These problems arise frequently in engineering. Two types of problems will be considered.

First, suppose we are trying to approximate a function $f(x)$ by a polynomial,

$$p(x) = \sum_{n=1}^{N} c_n x^n,$$

for $0 \le x \le 1$. We would like the function $p(x)$ to be close to f for all $0 \le x \le 1$. To make the problem into a linear program, we first replace the condition that the approximation be good for all x by the condition that it be good for a large 'dense' collection x_j. Thus we want

$$|f(x_j) - p(x_j)| \le \epsilon$$

where ϵ is as small as possible.

First this collection of inequalities is rewritten as

$$f(x_j) - p(x_j) \leq \epsilon,$$

$$f(x_j) - p(x_j) \geq -\epsilon.$$

To finish the statement of the problem, suppose we fix N. Our variables are $(\epsilon, c_1, \ldots, c_N)$, and the function we want to minimize is ϵ with the sign constraint $\epsilon \geq 0$. Notice that an inequality of the form

$$f(x_j) - p(x_j) \leq \epsilon$$

can be rewritten as

$$\sum_n c_n x_j^n \geq \epsilon + f(x_j),$$

which is a linear inequality on the coefficients c_n.

The second problem is similar, except that we try to minimize the error in one region, subject to the constraint that the error not be too great in another region. Suppose that $0 < R_1 < R_2 < 1$. In this case δ is given, and we have the constraints

$$f(x_j) - p(x_j) \leq \delta,$$

$$f(x_j) - p(x_j) \geq -\delta,$$

for $0 \leq x_j \leq R_1$. Typically in these problems we are not concerned with the values of $p(x_j)$ for $R_1 \leq x < R_2$, but then as in the previous problem we try to find ϵ which is nonnegative and minimal so that

$$f(x_j) - p(x_j) \leq \epsilon,$$

$$f(x_j) - p(x_j) \geq -\epsilon,$$

for $R_1 \leq x_j \leq R_2$.

Linear Network Flow

We are given a directed graph $\mathcal{G}$ with a set of N nodes and a set A of arcs. Each arc is an ordered pair of nodes (i, j). There is a flow f_{ij} from node i to node j; the flow can be water, manufactured products, people, etc. Each arc has an associated cost coefficient a_{ij}. The problem is to minimize the total cost

$$C = \sum_{(i,j) \in A} a_{ij} f_{ij}$$

subject to two types of constraints.

Often the inflow and outflow at node i must balance for all nodes.

$$\sum_{j|(i,j)\in A} f_{ij} - \sum_{j|(j,i)\in A} f_{ij} = 0, \quad i = 1, \ldots, N, \tag{0.1.3}$$

It is also possible to allow some nodes to be sources or sinks, in which case these constraints take the form

$$\sum_{j|(i,j)\in A} f_{ij} - \sum_{j|(j,i)\in A} f_{ij} = s_i, \quad i = 1, \ldots, N.$$

In addition there may be flow bound constraints of the form

$$b_{ij} \le f_{ij} \le c_{ij}, \quad (i,j) \in A.$$

In case $s_i \neq 0$ for some i, the problem requires $\sum_i s_i = 0$, as we can see by summing the sides of (0.1.3) over j.

0.1.4 ELP

There are a variety of essentially equivalent forms for these problems. The popular *simplex algorithm* takes advantage of the form called the *equality linear program* (ELP). In this form, the constraints are positivity of the variables and a set of equality constraints,

$$\sum_{n=1}^{N} a_{mn} x_n = b_m, \quad m = 1, \ldots, M.$$

Suppose we start with a problem with inequality constraints, as originally posed. Rewrite the inequality constraints with new *slack variables*

$$\sum_{n=1}^{N} a_{mn} x_n + x_{N+m} = b_m, \quad m = 1, \ldots, M,$$

where the slack variables $x_{m+1}, \ldots, x_{m+N}$ are also nonnegative.

It is easy to check that if we minimize g subject to these equality constraints, then the first N coordinates will be feasible, and must also be minimizing for the inequality constraints. Conversely, if we minimize g subject to the original inequality constraints, then we can choose values for the slack variables by taking

$$x_{m+N} = b_m - \sum_{n=1}^{N} a_{mn} x_n, \quad m = 1, \ldots, M,$$

and the vectors $(x_1, \ldots, x_{M+N})$ will optimize the ELP.

For most of the remaining discussion, the linear programs are considered in the (ELP) form: minimize

$$g(x_1, \ldots, x_N) = \sum_{n=1}^{N} c_n x_n$$

subject to the constraints

$$\sum_{n=1}^{N} a_{mn} x_n = b_m, \quad m = 1, \ldots, M,$$

$$x_n \geq 0, \quad n = 1, \ldots, N.$$

Notice that the matrix A and the other variables do not have to come from an inequality linear program, despite the use of the same notation. For notational convenience we define rows $A_1, \ldots, A_M$ for our constraint matrix $A = (a_{mn})$. To avoid problems due to redundancy in the constraints, we assume that $rank(A) = M$, that is the vectors $A_1, \ldots, A_M$ are linearly independent. Define

$$B = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}.$$

As we will see shortly, a special role is played by the solutions of (ELP) with at least $N - M$ of the x_n equal to 0. Such a vector is called a *basic solution*, and if it also satisfies $x_k \geq 0$ for all $k = 1, \ldots, N$, it is called a *basic feasible solution*.

Before turning to the important geometric aspects of linear programming, consider the homogeneous system

$$AY = 0, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}.$$

Since the rows of A are linearly independent, this system of equations has a solution space of dimension $N-M$. Also, A will have an $M \times M$ submatrix $\tilde{A}$ which is invertible (exercise 4). Thus we can find a particular solution X_p of $AX = B$ by setting at least one group of $N-M$ of the x_n equal to 0, and getting the other components of X_p via

$$\tilde{X} = \tilde{A}^{-1}B.$$

The general solution of $AX = B$ always has the form $X = X_p + Y$, where Y is any solution of $AY = 0$. A set of the form $X_p + V$, where V is a vector subspace (of R^N), is called an affine subspace. The feasible region for our linear program, in ELP form, is thus the intersection of the affine subspace of solutions to $AX = B$ with the positive sector $x_n \geq 0$ in R^N.

0.1.5 Geometry of Linear Programs

Hyperplanes

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

Suppose $\mathcal{S}$ is an $N-1$ dimensional subspace and that $\mathcal{T} = X + \mathcal{S}$ is a hyperplane. Let $\beta \in R^N$ be a nonzero vector which is orthogonal to $\mathcal{S}$, so that $\mathcal{S} = \{Y \in R^N | \langle \beta, Y \rangle = 0\}$. Then if $X + S \in \mathcal{T}$ we have $\langle X + S, \beta \rangle = \langle X, \beta \rangle$. Conversely, suppose that $\beta \neq 0$. Define $\mathcal{S} = \{Y \in R^N | \langle \beta, Y \rangle = 0\}$, and consider the set of $Y \in R^N$ such that $\langle Y, \beta \rangle = \alpha$. Choose a particular Y_1 with $\langle Y_1, \beta \rangle = \alpha$. Then $\langle Y - Y_1, \beta \rangle = 0$ so that $Y - Y_1 \in \mathcal{S}$ and $Y \in Y_1 + \mathcal{S}$. Consequently the hyperplanes are exactly the level sets of the linear functions obtained by taking the inner product with some nonzero $\beta \in R^N$.

Suppose that $X_1, X_2 \in R^N$, $X_1 \neq X_2$. If $\beta = X_2 - X_1$, then $\langle X_2 - X_1, \beta \rangle \neq 0$, so that $\langle X_2, \beta \rangle \neq \langle X_1, \beta \rangle$. Thus there is a linear function given by inner product with β separating X_1 and X_2, or, geometrically, there is a hyperplane containing X_1 but not X_2.

Lemma 0.1.1. *If $\Omega \subset R^N$ is convex, then either Ω lies in a hyperplane or Ω contains an open subset of R^N. If $\Omega \subset R^N$ is convex, is closed under multiplication by -1, and does not lie in a hyperplane, then Ω contains an open neighborhood of 0.*

Proof. For the first part, it will be convenient to have $0 \in \Omega$; this can always be achieved by subtracting a single vector from every element of Ω, and this translation process will not affect the result. If Ω does not lie in a proper subspace of R^N then there must be $U_1, \dots, U_N \in \Omega$ which are linearly independent. There is an invertible linear map L taking U_j to the standard basis vector E_j. The function L is continuous, so the inverse image of any open set is open. Take the inverse image of the set K of points X in R^N (called the standard open simplex) of the form

$$X = \sum_{n=1}^N w_n E_n, \quad 0 < w_n < 1, \quad \sum_{n=1}^N w_n < 1.$$

The set $L^{-1}(K)$ will be a open subset of the convex combinations of $0, U_1, \dots, U_N$, which gives an open set contained in Ω.

For the second part, the linear map L will take the convex combinations of

$$\pm U_1, \dots, \pm U_N \in \Omega$$

to $\pm E_n$, where E_n are the standard basis vectors. The convex hull of $\{L^{-1}(\pm E_n)\}$ will have the desired neighborhood of 0. □

Lemma 0.1.2. *If Ω is a convex set which is not equal to R^N, then there is a vector $\beta \in R^N$ with $\|\beta\| = 1$, and a real number α, such that every element $X \in \Omega$ satisfies $\langle \beta, X \rangle \le \alpha$. For each $Y \in R^N$ not in the closure of Ω, β and α may be chosen so that in addition $\langle \beta, Y \rangle > \alpha$.*

Proof. First observe that there is a nonempty open set V disjoint from Ω. If a translate of Ω lies in a proper subspace of R^N this is easy. If Ω does not lie in a hyperplane, then by the previous lemma Ω contains an open set W. Let $Y_0 \in R^N$ with $Y_0 \notin \Omega$. Consider all the line segments

$$(1 - t)Y_0 + tP, \quad -1 \le t \le 1,$$

with midpoint Y_0 having one endpoint P in W. Since $Y_0 \notin \Omega$ the second endpoint $2Y_0 - P$ of each segment is not an element of Ω. This collection of endpoints forms an open set $\mathcal{P}$ disjoint from Ω. As a consequence of this first observation, there is a point Y which is not in the closure of Ω; the closure of Ω is also convex.

Let Z be a point in the closure of Ω which minimizes the distance to Y. The proof will be finished if all points in Ω lie on one side of the hyperplane

$$H = \{X \in R^N | \langle X - Z, Y - Z \rangle = 0.$$

Let $W \in \Omega$ be different from Z, and consider the function

$$\phi(t) = \langle tW + (1-t)Z - Y, tW + (1-t)Z - Y \rangle.$$

Since this function has a minimum at $t = 0$, it follows that $\phi'(0) \geq 0$. But $\phi'(0) = 2\langle W - Z, Z - Y \rangle$ so that $\langle W, Y - Z \rangle \leq \langle Z, Y - Z \rangle$.

Since Y is not in the closure of Ω, the vector $Y - Z$ satisfies $\langle Y - Z, Y - Z \rangle > 0$ and

$$\langle Y, Y - Z \rangle > \langle Z, Y - Z \rangle \geq \langle W, Y - Z \rangle.$$

Choose

$$\beta = \frac{Y - Z}{\|Y - Z\|}, \quad \alpha = \langle Z, \beta \rangle.$$

□

Recall that Z is a boundary point of a set S in R^N if every ϵ ball centered at Z contains points of S and the complement of S.

Lemma 0.1.3. *Suppose that Z is a boundary point of the closed convex set Ω. Then there is a vector $\beta \in R^N$ with $\|\beta\| = 1$ such that for all $X \in \Omega$*

$$\langle \beta, X \rangle \leq \langle \beta, Z \rangle.$$

Proof. Since Z is a boundary point of Ω there is a sequence of points $\{Y_k\}$ in the complement of Ω with $\|Y_k - Z\| \leq 1/k$, for $k = 1, 2, 3, \dots$. By Lemma 0.1.2 there are vectors β_k with $\|\beta_k\| = 1$, and numbers α_k, such that

$$\langle \beta_k, X \rangle \leq \alpha_k < \langle \beta_k, Y_k \rangle, \quad X \in \Omega.$$

Since $\langle \beta_k, Z \rangle \leq \alpha_k$, the Cauchy-Schwarz inequality implies that the sequence $\{\alpha_k\}$ is bounded below by $-\|Z\|$. Similarly, the inequalities

$$\alpha_k < \langle \beta_k, Y_k \rangle, \quad \|Y_k\| \leq \|Z\| + 1/k,$$

imply that the sequence $\{\alpha_k\}$ is bounded above by $\|Z\| + 1$. In addition, the set of vectors in R^N with norm 1 is compact. Thus the sequence $\{\beta_k\}$ has a subsequence $\{\beta_j\}$ which converges to $\beta \in R^N$ with $\|\beta\| = 1$, and such that $\lim_{j\to\infty} \alpha_j = \alpha$.

For every $X \in \Omega$,

$$\langle \beta, X \rangle = \lim_{k\to\infty} \langle \beta_k, X \rangle \le \lim_j \alpha_j = \alpha.$$

On the other hand

$$\langle \beta, Z \rangle = \lim_{j\to\infty} \langle \beta_j, Y_j \rangle \ge \lim_j \alpha_j = \alpha.$$

Since $Z \in \Omega$,

$$\langle \beta, Z \rangle = \alpha.$$

□

Extreme points

If Ω is convex then we say that U is an *extreme point* of Ω if

$$U = wV + (1-w)Z, \quad V, Z \in \Omega, \quad w \in (0,1)$$

implies $U = V = Z$.

Lemma 0.1.4. *If $\Omega \subset R^N$ is a compact convex set, which is contained in an affine subspace of dimension M, then every $X \in \Omega$ is the convex combination of at most $M+1$ extreme points of Ω.*

Proof. If Ω has only one point, or if X is an extreme point, there is nothing to show.

The proof proceeds by induction on the dimension M. If $M = 1$ then a compact convex set is a point or a compact interval, so the result is immediate. Suppose the result is true for $M < K$ and let Ω be a compact convex subset contained in an affine subspace Aff of dimension K.

It will be convenient to work in R^K instead of Aff. First (see Exercises 1 and 2), there is a $Y_p \in \Omega$ and an $N \times K$ matrix $\mathbf{L}$ such that the function $L : X \to \mathbf{L}X + Y_p$ maps R^K one to one and onto Aff. The set $\Lambda = L^{-1}(\Omega) \subset R^K$ is convex, and Z is an extreme point of Ω if and only if $L^{-1}(Z)$ is an extreme point of Λ.

We first show that Λ has an extreme point. If Λ has at least two points, X_1 and X_2, then there is a vector $\beta \in R^K$ such that $\langle X_1, \beta \rangle < \langle X_2, \beta \rangle$. Since this function is continuous and Λ is compact, there is a subset of Λ where this function takes its maximum value. The set of all $X \in R^K$ where the function takes on this maximum is an affine subspace C of dimension $K-1$. By the induction hypothesis $\Lambda \cap C$ has at least one extreme point Y, and because $\langle Y, \beta \rangle$ is maximal, Y must also be an extreme point of Λ (Exercise 3).

(Exercise 3: Suppose $\beta \in R^N$, $\Omega \subset R^N$ is convex, and $\alpha = \max_{X\in\Omega}\langle \beta, X\rangle$. If Z is an extreme point of $\Omega \cap \{X \in \Omega \mid \langle \beta, X\rangle = \alpha\}$, then Z is an extreme point of Ω.)

For $0 \le t \le 1$ and $X \in \Lambda$ the segment $(1-t)Y + tX$ lies in Λ. Since Λ is compact it has a boundary point $Z = (1-t_1)Y + t_1 X$, for some $t_1 \ge 1$. Since Λ is a closed set, Lemma 0.1.3 guarantees that there is a $\beta \in R^K$ with $\|\beta\| = 1$, and the function

$$X \to \langle X, \beta \rangle, \quad X \in \Lambda$$

is maximal at Z. Again using the induction hypothesis, Z may be written as a convex combination of K extreme points $Z_1, \dots, Z_K$, and then X may be written as a convex combination of the extreme points $Z_1, \dots, Z_K, Y$.

Finally, $\mathbf{L}X$ is a convex combination

$$\mathbf{L}X = w_{K+1}\mathbf{L}Y + \sum_{j=1}^{K} w_j \mathbf{L}Z_j.$$

□

This previous lemma is sometimes stated, a bit less precisely, but in a form that generalizes quite a bit, as

Every compact, convex set in R^N is the convex hull of its extreme points.

0.1.6 Back to Linear Programming

Start with a simple lemma.

Lemma 0.1.5. *Suppose $g : \mathbb{R}^N \to \mathbb{R}$ is an affine function,*

$$g(x_1, \dots, x_N) = c_1x_1 + \cdots + c_Nx_N + b.$$

Assume $V, W \in \mathbb{R}^N$, and let

$$X(t) = tV + (1-t)W, \quad t \in \mathbb{R}.$$

Then the function $G(t) = g(X(t))$ is an affine function of one variable.

Proof. A direct substitution gives

$$G(t) = c_1(tv_1 + (1-t)w_1) + \cdots + c_N(tv_N + (1-t)w_N)$$
$$= t(\sum_{k=1}^{N} c_k(v_k - w_k)) + \sum_{k=1}^{N} c_k w_k.$$

□

Now there is in general no guarantee that the linear function we want to minimize in (ELP) actually has a minimum. Examples are easy to construct. However, if the feasible set Ω is compact, then since g is continuous, the min must be achieved on Ω.

Theorem 0.1.6. *If $\Omega \subset R^N$ is compact and convex, and if $g : R^N \to R$ is linear (or affine), then the minimum of g on Ω occurs at an extreme point.*

Proof. An induction proof will work. Let the minimum of g on Ω be achieved at X_0. We know that any $X_0 \in \Omega$ 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 ($J = 1$), 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 $t \in \mathbb{R}$ let

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

By Lemma 0.1.5 the function $G(t) = g(X(t))$ is affine, and $X_0 = X(w_1)$, with $0 \le w_1 \le 1$. An affine function on $[0, 1]$ must take its minimum at an endpoint, either $t = 0$ or $t = 1$. (Of course the function could be constant, thereby hitting the min on the whole segment.) Thus the min 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, in which case the induction hypothesis finishes the proof. □

The feasible set Ω for an ELP need not be compact, but with a little more thought we can see that for ELP, any minimum which does exist must occur at an extreme point.

Theorem 0.1.7. *If the objective function g for (ELP) has a minimum on the feasible set Ω, then this minimum occurs at an extreme point.*

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

$$X_0 = tU + (1-t)V, \quad 0 < t < 1, \quad U, V \in \Omega, \quad U \neq V.$$

Notice that if any coordinate of X_0 is zero, then the same coordinate must be 0 for both U and V, otherwise they would not both be in Ω, where all coordinates are nonnegative.

Since g is minimal at X_0 and affine on the segment, it must be constant on the segment. Extend the segment to $X(t) = tU + (1-t)V$ for $t \in \mathbb{R}$. Since $U \neq V$ there is a value of t where $X(t)$ has nonnegative coordinates, but a coordinate which was positive for X_0 becomes 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 Ω. The next result will show that the set of extreme points of Ω 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).

While that is good news, there is also bad news. For many linear programming problems the set of extreme points is too large for an exhaustive computer search. The analysis of the simplex method continues by characterizing extreme points where a minimum is achieved so that an exhaustive search is generally not required.

Theorem 0.1.8. *Suppose X_0 satisfies $AX = B$, the components of X_0 are nonnegative, and the components of X_0 which are zero are the K components $x_{n(1)}, \dots, x_{n(K)}$. Then X_0 is an extreme point for Ω if and only if the coefficient matrix for the system of equations*

$$\begin{aligned} AX &= B, \\ x_{n(1)} &= 0, \\ &\vdots \\ x_{n(K)} &= 0, \end{aligned} \tag{0.1.4}$$

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

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 $U, V \in \Omega$, then since U and V have components $n(1), \dots, n(K)$ 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 (0.1.4) has rank less than N, (Of course this will happen if $K < N - M$.) and X_0 has exactly K coordinates equal to 0. Since X_0 is a solution of (0.1.4), there is a nontrivial affine space of solutions S, and $X_0 \in S$.

Pick $U \neq X_0$ in the affine solution space S. Then $X_0 - U$ satisfies the homogeneous version of (0.1.4) and

$$X(t) = X_0 + t(X_0 - U)$$

lies in S. Since U satisfy the equations (0.1.4), U and $X(t)$ have coordinates $n(1), \dots, n(K)$ equal to 0. Since the remaining coordinates of X_0 are positive by the hypotheses, these coordinates will remain positive for $X(t)$ if t is in some interval $-\epsilon \leq t \leq \epsilon$ with $\epsilon > 0$.

Now

$$X_0 = \frac{1}{2}X(-\epsilon) + \frac{1}{2}X(\epsilon),$$

and

$$X(\epsilon) - X(-\epsilon) = 2\epsilon(X_0 - U).$$

That is, $X(\epsilon)$ and $X(-\epsilon)$ are distinct, and X_0 is not extreme. □

In most cases the number K in the previous theorem is just the difference between the number of columns and rows in the matrix A. Occasionally the rank of the (augmented) coefficient matrix may be less than N, in which case additional zero components should be added.

Here are two simple examples illustrating the ideas of the last theorem.

Example LP1: Minimize $g(w, x, y, z) = 3w - x + 4y + z$ subject to the constraints

$$w + x + y + z = 5,$$

$$w + 2x - 3y + z = 6,$$

with $w, x, y, z \geq 0$.

Notice that the first constraint equation together with the requirement that all components be nonnegative means the feasible set Ω is bounded and closed, so compact. There is a solution. Look for the minimum at extreme points, which are solutions of the constraint equations with at least 2 components equal to 0.

There are $\frac{4 \cdot 3}{2!} = 6$ ways of picking two coordinates to be zero. These choices give the following solutions of the equations.

$$(4, 1, 0, 0), (0, 21/5, 4/5, 0), (0, 1, 0, 4),$$

$$(5.25, 0, -.25, 0), (0, 0, -1/4, 5.25), (w, 0, 0, z).$$

The first three choices correspond to extreme points. The fourth and fifth don't satisfy the positivity constraints. In the last case we find equations

$$w + z = 5, \quad w + z = 6.$$

Of course these are contradictory; the rank of the augmented coefficient matrix is not 4.

At the extreme points the values of g are respectively $11, -1, 3$, so the minimum occurs at $w = 0$, $x = 21/5$, $y = 4/5$, $z = 0$.

Example LP2: Minimize $g(w, x, y, z) = 3w - x + 4y + z$ subject to the constraints

$$w + x + y + z = 5,$$

$$2w + x + y + 2z = 5,$$

$$3w + x + y + 2z = 5,$$

with $w, x, y, z \geq 0$.

Notice that the first constraint equation together with the requirement that all components be nonnegative means the feasible set Ω is bounded and closed, so compact. There is a solution. Look for the minimum at extreme points, which are solutions of the constraint equations with at least 1 component equal to 0.

There are 4 ways of picking one coordinate to be zero. If we add the equation $z = 0$, the equations force $w = 0$ and we find a line of solutions $x + y = 5$. The extreme points $(0, 5, 0, 0)$ and $(0, 0, 5, 0)$ are found by adding equations forcing additional coordinates to be 0 so the rank of the coefficient matrix is 4.