The Mathematical Foundations of

OPTIMIZATION

Notes for use with the course MATH314

Chapter 1 Introduction

Chapter 2 Analytic preliminaries

Chapter 3 Convexity

Chapter 4 Linear Programming

Chapter 5 Non Linear Optimization

The University of Newcastle

INTRODUCTION

Problems of optimization are among the most ubiquitious in mathematics and its many applications. Consequently, the subject of optimization is extensive and necessarily the selection of material for this course is an eclectic one. The emphasis throughout will be on the underlying theory, from which we rigorously develop basic approaches to several important classes of optimization problems. While we motivate and illustrate the problems with simple instances of how they arise in practice, we will not be overly concerned with applications. Nor will we concern ourselves with important but more practical aspects of the topic such as the development of computationally **efficient** algorithms.

The subject provides a natural vehicle for introducing several fundamental areas of Pure Mathematics, in particular, the rudiments of functional analysis and convexity theory.

Chapter 1 is largely concerned with examples. On a first reading you should not worry too much about the precise details of a problem or its formulation. Rather you should attempt to gain the flavour of what is being discussed. You should return to each problem for a more detailed examination when it is discussed later in the course.

Chapter 2 introduces those aspects of functional analysis (largely in $\mathbb{R}^n$) which underlie much of the theory of optimization. Throughout this and subsequent chapters we will freely identify any n-dimensional vector space X with $\mathbb{R}^n$. If X is to be equipped with an inner-product (or norm) is it tacitly assumed to be that inherited under the identification from the standard *dot* product for $\mathbb{R}^n$.

Chapter 3 develops the theory of convex sets and convex functions. A revision of first and second year work on vectors and linear algebra might be appropriate at this stage.

Chapter 4 is concerned with linear programming. The simplex algorithm in tableau form is formulated and justified.

Chapter 5 concerns non-linear problems. A geometric approach is developed which leads naturally to the Karush, Kuhn-Tucker condition for convex problems and to the John Multiplier Rule in the differentiable case. Besides the material of Chapter 3, it is assumed that you are familiar with the basic multivariable calculus developed in second year.

Exercises are provided at the end of each chapter. You should attempt to do as many of these as possible as you work through the notes. The more difficult ones are marked with a $*$.

Please accept my apology for any obscurities in the presentation and for the minor errors which undoubtedly occur. They are certainly my most original contribution to the subject.

I relied heavily on the book 'Convex Functions' by Roberts and Varberg (Academic Press, 1973) for the presentation in Chapters 3 and 4.

Chapter 5 was developed largely from the article, 'Modern Multiplier Rules' by B.H. Pourciau (American Mathematical Monthly, Vol.87, no.6, 1980, pp.433-452), the book 'Optimization Theory: the Finite Dimensional Case' by M. Hestenes (Wiley, 1975), and the survey 'Non Convex Optimization Problems' by Ivar Ekeland (Bulletin of the American Mathematical Society, Vol.1, No.3, May 1979).

Unfortunately the constraints imposed by a one semester introductory course have prevent the inclusion of many important aspects of the subject. For example, it has not been possible to consider; the fundamental notion of Duality for non-linear problems, the use of penalty functions, important techniques such as Bellman's method of Dynamic Programming, or various other algorithms for handling non-linear problems.

Brailey Sims

CHAPTER 1

The problem central to our study, and one which arises naturally in a great variety of situations is the following **optimization problem:**

minimize: $f(x)$
subject to: $x \in C$

That is, we seek a point $x_0 \in C$ such that $f(x_0) \leq f(x)$ for all $x \in C$. Here $f : X \to \mathbb{R}$ is a given function, referred to as the **objective function**, defined on some 'natural' domain X (usually $X = \mathbb{R}^n$, or some other vector space) and C is a specified subset of X known as the **constraint set**, or the set of **feasible** (= possible) **solutions.** If $C = X$ we refer to the problem as an *unconstrained problem.*

We may think of the problem as first determining

$$m := \text{ infimum } \{f(x) : x \in C\}$$

and then, if it exists, a point $x_0 \in C$ for which $f(x_0) = m$.

When such an x_0 exists we refer to it as an **optimal solution**, or simply a *solution* to the problem. m is then referred to as the **minimum of** f on C, and we say f *attains its minimum* on C at x_0.

NOTE: The problem of **maximizing** $f(x)$ can always be rephrased as that of minimizing $-f(x)$. So maximization problems are also encompassed by the general optimization problem set out above.

Possible strategies which we will pursue include:

• Finding **necessary conditions** for an optimal solution. All optimal solutions will satisfy such conditions, however not all points satisfying them are necessarily optimal. For example, the condition

$$f'(x_0) = 0$$

is a necessary condition for the differentiable function $f : \mathbb{R} \to \mathbb{R}$ to be a minimum at x_0. In general necessary conditions serve to narrow down our search for optimal solutions. If however, we are lucky and there is only one point satisfying them and we are sure that an optimal solution exists then that point is the optimal solution.

- Determining **sufficient conditions** for an optimal solution. Any point satisfying such conditions is optimal, however not all (indeed any) optimal solutions need satisfy the conditions. Useful sufficient conditions are the most dificult to find. Such conditions cannot be used to search for solutions — they may miss them. At best they may serve to verify that a suspected solution is indeed optimal, an example of a sufficient condition for $f : \mathbb{R} \to \mathbb{R}$ is $f''(x) \geq 0$ for all $x \in \mathbb{R}$ and $f'(x_0) = 0$.

Only rarely is it possible to derive the elusive and highly prized **necessary** and **sufficient** condition.

- **Solve the optimal programming problem**; that is, develop effective algorithms for determining (or at least approximating) $m = \text{infimum } \{f(x) : x \in C\}$, and if possible also locating an optimal solution x_0. Usually such algorithms are of an **iterative** nature: starting with some initial point x_1 the algorithm generates an 'improved' point x_2 with $f(x_2) < f(x_1)$. Reapplying the algorithm with x_2 as starting point we obtain a better *approximate solution* x_3. Continuing in this fashion a sequence of ever-improving points $x_1, x_2, x_3, x_4, \ldots$ is generated which hopefully, either terminates at an optimal solution, or converges to one (in practice the process is terminated when an adequately close approximation is obtained).

The following considerations are of vital importance for the successful application of these strategies.

- **Establishing the existence of an optimal solution**. Without an assurance that the problem has a solution large amounts of computation may be wasted before the infeasibility of the problem is realized, or even worse one may be led to accept an "approximation" to the *non-existent answer*.

REMEMBER: "One good theory is worth a thousand computer runs".

• **Verifying that the algorithm will indeed generate a sequence of approximations which converge and converge to a solution.** Often this can be achieved only if one can **verify that the problem has a unique optimal solution**, algorithms often behave chaotically when there is more than one point to which they might converge.

• **Obtain reliable estimates on the rate at which the sequence of approximations converges to a solution.** Such estimates allow the iterative procedure to be terminated at a sensible stage.

• **Determine bounds on the size of the problem and the complexity of the algorithm.** There is little point embarking on a lengthy computation without some assurance that it can fit on the machine and be completed in realistic time.

We conclude this chapter by presenting a number of examples which llustrate simple optimization problems and some of the circumstances from which they arise.

EXAMPLE 1. (a) **Diet Problem**

A University cafeteria intends to serve chilli, the basic ingredients being meat and beans for which the following nutritional information is available.

constituents	*units per ounce of ingredient*		*max units per serve*	*min units per serve*
	meat	*bean*		
indigestional	1	1	3	-
flatulate	1	5	10	-
sustainol	2	1	-	1

Suppose that the *cost per ounce* of meat is 50 cents and per ounce of beans is 10 cents.

We wish to know x_1, the number of ounces of meat, and x_2, the number of ounces of bean to include per serving, so that the cost $50x_1 + 10x_2$ is a minimum subject to the constraints

indigestinal:	$x_1 + x_2 \leq 3$
units flatulate:	$x_1 + 5x_2 \leq 10$
units of sustainol:	$2x_1 + x_2 \geq 1$

and the positivity constraints

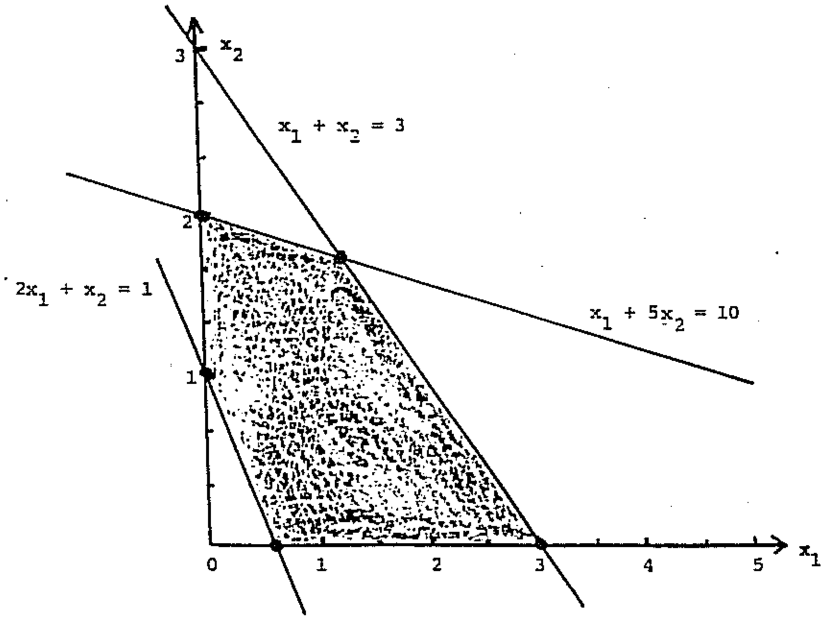
$x_1 \geq 0$	(It is hard to take meat out of the customer)
$x_2 \geq 0$	(and even harder to take beans).

Thus, we have the problem

minimize: $50x_1 + 10x_2$

subject to:
$$x_1 + x_2 \leq 3$$
$$x_1 + 5x_2 \leq 10$$
$$2x_1 + x_2 \geq 1$$
$$x_1,\ x_2 \geq 0$$

Here the *feasible region* is the convex polygon illustrated below.



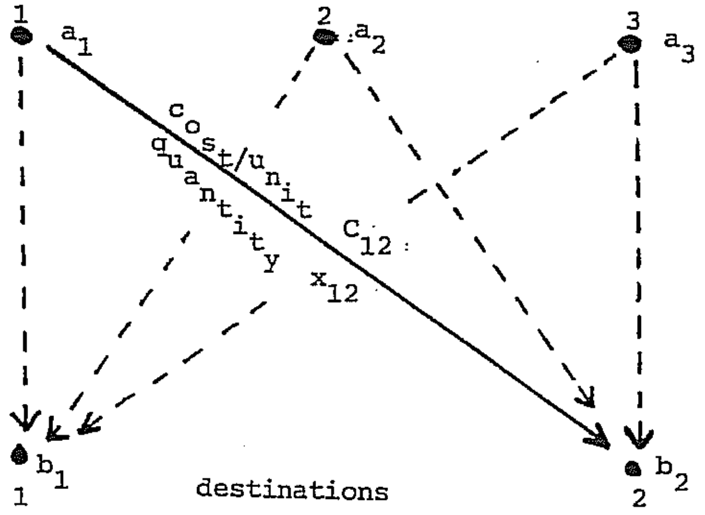
From which we see that at least feasible solutions exist.

This is of course a trivialized version of a serious problem, although in reality both the numbers of ingredients and constituents would be much larger. Because the objective function and each of the expressions in the constraints are linear in the variables (x_1 and x_2) the problem is known as a **Linear programming problem.**

Linear programming problems constitute perhaps the largest single class of problems encountered in practice, particularly from the field of economics and industry. Another example of a linear programming problem is:

EXAMPLE 1. (b) Transportation problem:
Quantities $b_1, b_2, \ldots, b_n$ of a certain good are to be supplied to n different destinations. Suppose there are m sources of supply (warehouses/factories) and that the i'th source is able to supply an amount a_i
(of course we require

$$\sum_{j=1}^{n} b_j \leq \sum_{i=1}^{m} a_i \ ) .$$



If the cost of shipping one unit of the product from the i'th source to the j'th destination is c_{ij} and x_{ij} denotes the amount shipped from i to j, then we have the problem:

Find x_{ij} $\quad i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ so that

$$\sum_{i,j} c_{ij} x_{ij} \quad \text{(mn unknowns)}$$

is a minimum, subject to the constraints;

(amount leaving i'th source) $\quad \sum_j x_{ij} \leq a_i \quad i = 1, 2, \ldots, m$

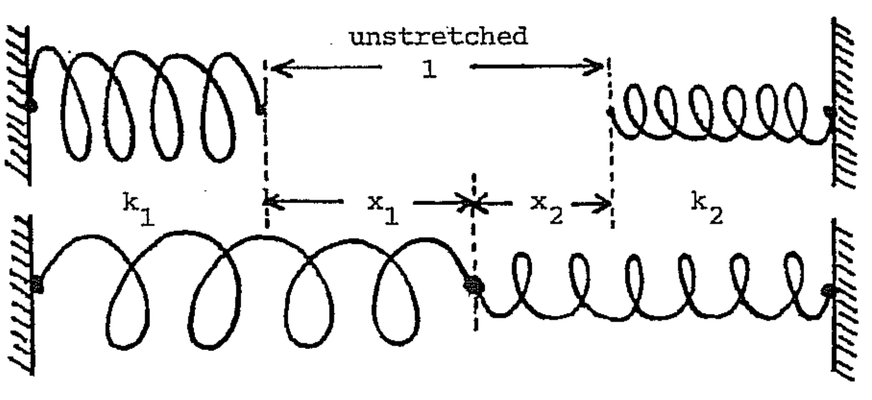
(amount arriving at $j'th$ destination) $\quad \sum_i x_{ij} = b_j \quad j = 1, 2, \ldots, n$

(positivity) $\quad x_{ij} \geq 0 \quad i = 1, 2, ..., m; j = 1, 2, \ldots, n.$
a total of $(m + n + mn$ constraints).

Similar problems arise in Manufacturing (to schedule production, minimize wastage, design plant layouts), agriculture, engineering, Military operations and many other areas including computing, statistics and mathematics itself.

EXAMPLE 2. Mechanics Stable equilibrium configurations are those for which the energy of a system is minimal (more generally, Lagrangian mechanics tells us that physical systems act in such a way as to minimize their *action*).

For example consider the equilibrium of two springs attached as illustrated



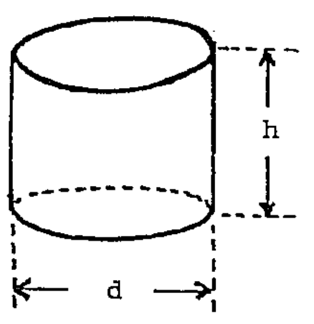
Since the energy in a spring with constant k when extended from equilibrium by an amount x is $\frac{1}{2}kx^2$, to determine the equilibrium configuration for our two springs we seek x_1 and x_2 for which $\frac{1}{2}k_1x_1^2 + \frac{1}{2}k_2x_2^2$ is a minimum, subject to the constraints

$$x_1 + x = 1$$
$$x_1 \geq 0$$
$$x_2 \geq 0.$$

(Verify that the solution to this is $k_1x_1 = k_2x_2$, i.e. the forces balance.)

Because of the nature of the objective function this is an example of a **quadratic programming problem** - a special case of **convex programming.**

EXAMPLE 3. **Production cost for a metal can of prescribed volume V.**



Cost of production is the sum of

(i) Cost of material $\propto$ surface area $\pi dh + \frac{1}{2}\pi d^2$

and

(ii) Cost of cutting and seaming $\propto$ length of 'edges', $2\pi d + h$.

Thus our problem is:

Minimize:

$$f(d,h) = c_1(\pi dh + \frac{1}{2}\pi d^2) + c_2(2\pi d + h)$$

subject to the constraints;

$$\frac{1}{4}\pi d^2 h = V$$

and limitations set by useage requirements such as

$$0 < d_m \leq d \leq d_M$$

$$0 < h_m \leq h \leq h_M$$

(It would be hard to drink out of a 1 foot diameter half inch high beer can!)

EXAMPLE 4. Games Theory

To illustrate, consider the following (not very exciting) *2 person game* between players A. Hog and B. Greedy.

A and *B* each have a large supply of 5 cent and 10 cent coins. At a given signal, each displays a coin. If the sum of the displayed coins is odd A wins B's coin; if it is even, B wins A's coin.

We can summarize by representing A's winnings with the **pay off matrix**

$P = [p_{ij}] :=$

A \ *B*	5	10
5	-5	10
10	5	-10

Not knowing how B will play, player A reasons to maximize his winnings (minimize his losses) as follows.

If I choose 5 cents ($i = 1$) the least I can win is $\min_j p_{1j} = -5$.

If I choose 10 cents ($i = 2$) the least I can win is $\min_j p_{2j} = -10$.

I should choose to maximize this; that is, to achieve

$$\max_i \min_j p_{ij} = -5$$

so A plays 5 cents (worst when B also plays 5 cents).

Similarly, B chooses to minimize his maximum loss, that is

$$\min_j \max_i p_{ij} = \min \begin{Bmatrix} j=1 & j=2 \\ 5 & 10 \end{Bmatrix} \\ = 5$$

so B also plays 5 cents (worst when A plays a 5 cent).

However, if B were to continue on this strategy, A would soon see the advantage of changing to 10 cents, which presumably would be followed shortly by a change of B to 10 cents and, then A back to 5 cents and so on.

This unstable situation is different if we change P to a new pay off matrix, such as

$Q = [q_{ij}] :=$

A \ B	5	10
5	-5	5
10	1	5

Note that $q_{21} = 1$ is the maximum entry of its column and the minimum entry of its row. Such a point is known as a **saddle point** for the matrix Q.

In this situation we have

$$\max_i \min_j q_{ij} = 1 = \min_j \max_i q_{ij}$$

(at A playing 10 cents which is worst if B plays 5 cents).

(at B plays 5 cents, which is worst if A plays 10 cents).

In this case, if A continues to play 10 cents nothing B can do will improve his situation, and while ever B plays 5 cents, nothing A can do will improve his lot either. The strategy; A plays 10 cents, B plays 5 cents, is **stable.**

These examples illustrate the situation in general two person games, which lead to a pay-off function $\phi(x, y)$ and the optimization problems:

> Find x_0 at which $\max_x \min_y \phi(x, y)$ occurs and y_0 at which $\min_y \max_x \phi(x, y)$ which occurs.

For obvious reasons such a problem is known as a **minimax problem.**

Note that one always has the inequality

$$\max_x \min_y \phi(x, y) \leq \min_y \max_x \phi(x, y) \qquad \ldots \quad (\star)$$

and that if there exists a saddle point (x_0, y_0) for ϕ; that is $\phi(x, y_0) \leq \phi(x_0, y_0) \leq \phi(x_0, y)$ for all x and y, then one has equality in $(\star)$ and the solution (x_0, y_0) corresponds to a stable strategy. (See exercise 2).

Returning to our original pay-off matrix, how might A and B proceed to play the game when there is no stable strategy? Here we could seek a **mixed strategy** for A (and B), one where A plays the various options available to him at random (so as to 'confuse' B) but with probabilities chosen to maximize his expected gain. And similarly for B.

More specifically: if A plays option i with probability a_i and B option j with probability b_j (thus $\sum_i a_i = \sum_j b_j = 1$ and $a_i \geq 0, b_j \geq 0$) then A's expected gain is

$$E(\mathbf{a}, \mathbf{b}) := \sum_{j=1}^{n} b_j \sum_{i=1}^{n} a_i p_{ij} = \sum_{j=1}^{n} \sum_{i=1}^{n} a_i b_j p_{ij}$$

where we are using the vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ to represent the respective probability distributions for A and B which we wish to determine.

We then seek $\mathbf{a}_0$ so that

$$\max_{\mathbf{a}} \min_{\mathbf{b}} E(\mathbf{a}, \mathbf{b}) \text{ occurs at } \mathbf{a} = \mathbf{a}_0$$

Note: In this case the objective function is $\min_{\mathbf{b}} E(\mathbf{a}, \mathbf{b})$, a function of $\mathbf{a}$ only while the constraint set corresponds to $\mathbf{a}$ being a probability distribution.

In our particular game

$$E(\mathbf{a}, \mathbf{b} = -5a_a b_1 + 10a_1 b_2 + 5a_2 b_1 - 10a_2 b_2$$

and we seek a probability distribution (a_1, a_2) which maximizes

$$f(a_1, a_2) = \min\{-5a_1 b_1 + 10a_1 b_2 + 5a_2 b_1 - 10a_2 b_2 : b_1 + b_2 = 1, b_1 \geq 0, b_2 \geq 0\}$$

Similarly we seek $\mathbf{b}_0$ so that

$$\min_{\mathbf{b}} \max_{\mathbf{a}} E(\mathbf{a}, \mathbf{b}) \text{ occurs at } \mathbf{b} = \mathbf{b}_0.$$

As we shall see, such a pair $(\mathbf{a}_0, \mathbf{b}_0)$ always exists and is a saddle point for $E(\mathbf{a}, \mathbf{b})$; a result known as the **fundamental theorem for matrix games.** The problem of finding $\mathbf{a}_0$ (and $\mathbf{b}_0$) will be seen to reduce to a linear programming problem (and its dual).

EXAMPLE 5. **Closest point problems** (approximation theory).

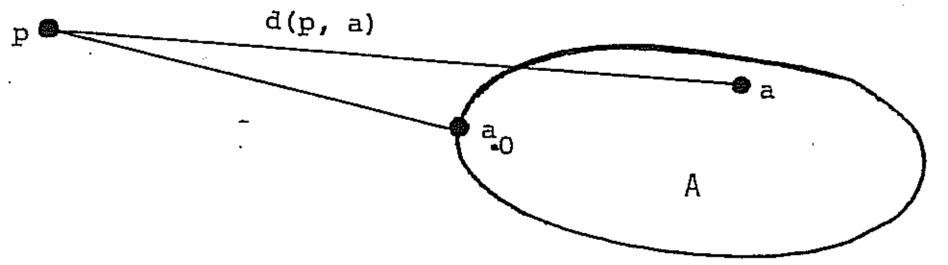
Given a set A, an element $p \notin A$ and some measure $d(p, a)$ of the *distance* from p to each point $a \in A$, we seek to find the distance from p to A; that is,

$$\inf_{a \in A} d(p, a),$$

and if it exists an element $a_0 \in A$ for which

$$d(p, a_0) = \min_{a \in A} d(p, a).$$

Such an element a_0 is a **best approximation** to p from A (or a **closest point** of A to p).



As an illustration, consider the statistical problem of **linear regression** (or, determining the straight line of **least squares best fit** to a set of data).

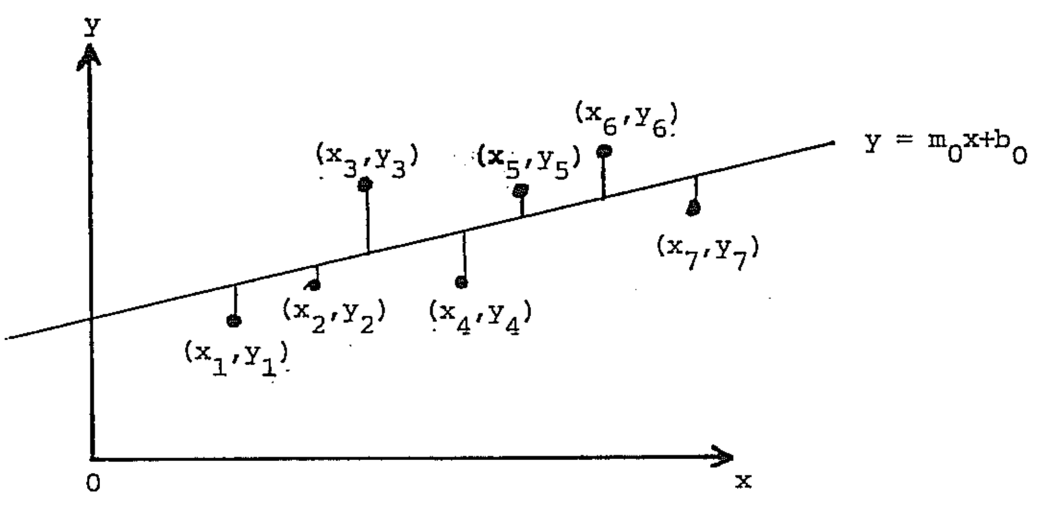
Here we have n data points;

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)$$

and seek m_0 and b_0 at which the function

$$f(m, b) := \sum_{i=1}^{n} (y_i - [mx_i + b])^2$$

is a minimum. Then $y = m_0 x + b_0$ is the line of best fit to the data (in the sense that it is the line which minimizes the sum of the squares of the residual errors at each data point).



To realize this as a closest point problem, let

$$\mathbf{y} := (y_1, y_2, \ldots, y_n)$$

$$\mathrm{x} := (x_1, x_2, \ldots, x_n)$$

and

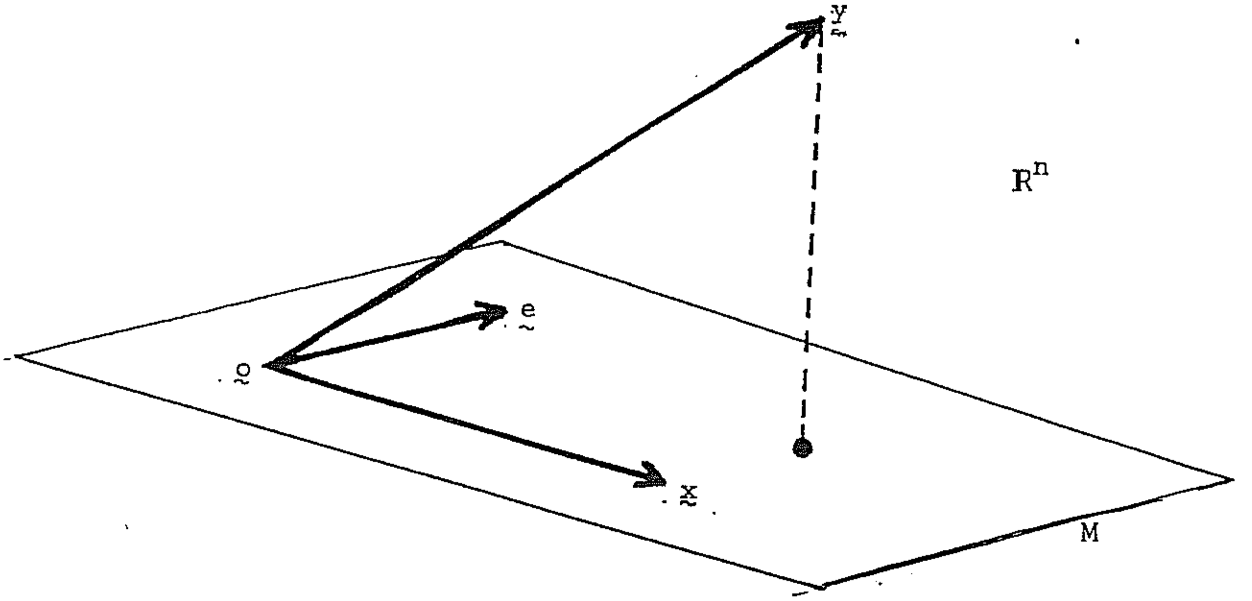
$$\mathrm{e} := (1, 1, \ \ldots, \ 1) \in \mathbb{R}^n$$

then for any m and b;

$$f(m, b) = \|\mathrm{y} - [m\mathrm{x} + b\mathrm{e}]\|^2$$

the square of the distance from y to the point $m\mathrm{x} + b\mathrm{e}$ in $\mathbb{R}^n$.

Thus, what we are seeking is the point $m_0\mathrm{x} + b_0\mathrm{e}$ from the subspace M spanned by x and e which is closest to the point y; that is, the best approximation from M to y.



Exercises

(1) Express the following *Manufacturing problem* as a linear programming problem. (Note, at this stage you are not being asked to solve the problem.)
Suppose we own two mines each of which can be operated at any level (x_1 tons per year, $i = 1, 2$) to yield ore containing lead, silver and zinc. When the mines are operated at the unit level (one ton per year) we have the following information.

	Yields in lbs			Cost of Operation
	lead	silver	zinc	
Mine 1	25	1/4	1	\$600
Mine 2	100	1/3	3/2	\$400

Assuming linearity in the production facilities, if it is necessary to produce at least 5000 lbs of lead, 60 lbs of silver and 150 lbs of zinc per year at what level should each mine be operated if the overall cost of operation is to be minimized?

(2) (i) Assuming that all necessary maxima and minima exist, show that for a function $f(x, y)$ of two variables we always have

$$\max_x \min_y f(x, y) \leq \min_y \max_x f(x, y) .$$

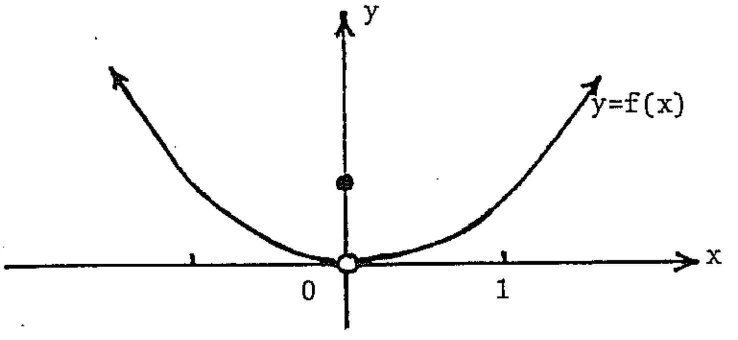
(ii) Show that if f has a saddle point at (x_0, y_0) then the inequality in (i) is an equality and the common value is $f(x_0, y_0)$.

CHAPTER 2 - Analytic Preliminaries

In this chapter we develop certain aspects of functional analysis relevant to optimization. In particular, we will derive basic criteria for the existence of optimal solutions.

To motivate the topics to be discussed, consider the function

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} x^2, & x \neq 0, \\ 1, & x = 0 . \end{cases}$$



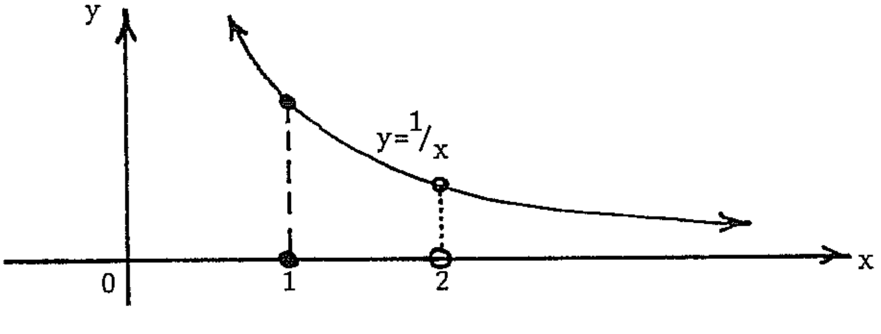
f fails to attain a minimum on $\mathbb{R}$ because of the discontinuity at $x = 0$. This shows that a continuity requirement on the objective function is essential for the existence of an optimal solution. A consideration we take up shortly for functions $f : \mathbb{R}^n \to \mathbb{R}$.

Next consider the optimization problems

minimize: $f(x) = \frac{1}{x}$
subject to: $x \geq 1$

and

minimize: $f(x) = \frac{1}{x}$
subject to: $1 \leq x < 2$; that is $x \in [1, 2)$



Even though the objective function is continuous, both problems fail to have optimal solutions. In the first case because the constraint set is unbounded and in the second because the end point $x = 2$ is excluded.

On the other hand

$$\begin{array}{ll} \textit{minimize:} & f(x) = \frac{1}{x} \\ \textit{subject to} & x \in [1, 2] \end{array}$$

has the optimal solution $x = 2$.

Here f is continuous and the constraint set is a *closed* and *bounded* interval.

Thus, we need to study subsets of $\mathbb{R}^n$ to identify restrictions which might be placed on a constraint set to ensure the existence of optimal solutions. From the above examples being both closed and bounded is necessary.

Throughout we will be concerned with $\mathbb{R}^n$ equipped with the *Euclidean norm.*

$$\begin{aligned} \|\mathbf{x}\| &:= \sqrt{\mathbf{x} \cdot \mathbf{x}} \\ &:= \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \end{aligned}$$

where $\mathbf{x} := (x_1, x_2, \ldots, x_n)$.

Much of what we do is however valid for an arbitrary *normed linear space*, that is a vector space X together with a **norm** function $\|\cdot\|$ satisfying:

1. $\|\mathbf{x}\| > 0$ if $\mathbf{x} \in X$ and $\mathbf{x} \neq \mathbf{0}$
2. $\|\lambda \mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$ for all $\mathbf{x} \in X$ and scalars λ .
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in X$ – *the triangle inequality*

That the Euclidean norm on $\mathbb{R}^n$ satisfies the triangle inequality follows from its definition and the Cauchy-Schwarz inequality

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$$

(refer to your linear algebra course for detail).

By the distance between two points $\mathbf{x}$ and $\mathbf{y}$ of $\mathbb{R}^n$ we shall understand

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| \; .$$

The **open ball** of radius r and centre $\mathbf{x}$ is

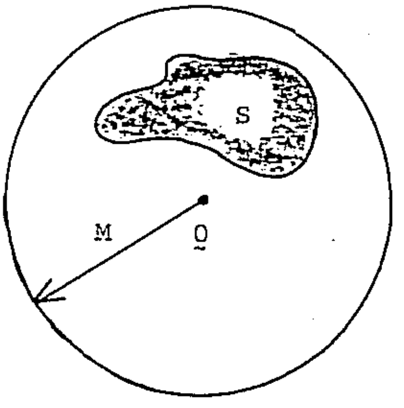
$$B_r(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| < r\}.$$

The **closed ball** of radius r and centre $\mathbf{x}$ is

$$B_r[\mathbf{x}] := \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| \le r\}.$$

A subset S of $\mathbb{R}^n$ is **bounded** if it is contained in some ball centred on $\mathbf{0}$. Equivalently S is bounded if

$$\|\mathbf{x}\| \le M \quad \text{for all} \quad \mathbf{x} \in S \quad \text{and some} \quad M > 0 .$$



Sequences and Convergence

Extending the definition in $\mathbb{R}$ we will say that a sequence of vectors $(\mathbf{x}_n)_{n=1}^{\infty}$ in R^n **converges** to $\mathbf{x}$ if, given $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\|\mathbf{x}_n - \mathbf{x}\| < \epsilon \quad \text{whenever } n \ge N.$$

When this is the case we write $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$, or simply $\mathbf{x}_n \to \mathbf{x}$.

Thus, $\mathbf{x}_n \to \mathbf{x}$ if and only if $\|\mathbf{x}_n - \mathbf{x}\| \to 0$.

Let A be a subset of $\mathbb{R}^n$, we define the **closure** of A, denoted by $\bar{A}$, to be the set of those points in $\mathbb{R}^n$ which are the limit of some sequence of points from A (such points are called *limit points* of A).

That is,

$$\bar{A} := \{\mathbf{x} \in \mathbb{R}^n : \exists (\mathbf{x}_n) \subseteq A \quad \text{with} \quad \mathbf{x}_n \to \mathbf{x}\} .$$

Always $A \subseteq \bar{A}$. To see this, for any $\mathbf{x} \in A$ take $\mathbf{x}_n = \mathbf{x}$ for all n to obtain $\mathbf{x} \in \bar{A}$.

We say that the set is **closed** if $A = \bar{A}$. Thus, A is closed if and only if whenever $(\mathbf{x}_n)$ is a sequence of points of A with $\mathbf{x}_n \to \mathbf{x}$ we have $\mathbf{x} \in A$ (that is, A contains all its limit points).

Lemma: $\mathbf{x} \in \bar{A}$ *if and only if for each* $\epsilon > 0$ *there exists an* $\mathbf{a} \in A$ *with* $\|\mathbf{x} - \mathbf{a}\| < \epsilon$.

Proof: ($\Rightarrow$) If $\mathbf{x} \in \bar{A}$ there exists $(\mathbf{a}_n) \subsetneq A$ with $\mathbf{a}_n \to \mathbf{x}$. Taking $\mathbf{a} = \mathbf{a}_n$ for any suficiently large value of n we then have $\|\mathbf{x} - \mathbf{a}\| < \epsilon$.

($\Leftarrow$) For each $n \in \mathbb{N}$, taking $\epsilon = \frac{1}{n}$ we may choose a point of A, call it $\mathbf{a}_n$, with $\|\mathbf{x} - \mathbf{a}_n\| < 1/n$. The resulting sequence $(\mathbf{a}_n)$ is contained in A and converges to $\mathbf{x}$, as $\|\mathbf{x} - \mathbf{a}_n\| < 1/n \to 0$. Thus $\mathbf{x} \in \bar{A}$.

□

Theorem: $\bar{A}$ *is the smallest closed set containing* A.

Proof: We must prove two things

(i) $\bar{A}$ is a closed set; that is, $\overline{(\bar{A})} = \bar{A}$

and

(ii) if B is any closed set containing A, then $\bar{A} \subseteq B$.

(i) Let $\mathbf{x} \in \overline{(\bar{A})}$, then for any $\epsilon > 0$ there exists by the lemma a point $\mathbf{y} \in \bar{A}$ with $\|\mathbf{x} - \mathbf{y}\| < \epsilon/2$.

Again by the Lemma, since $\mathbf{y} \in \bar{A}$ there exists $\mathbf{a} \in A$ with $\|\mathbf{y} - \mathbf{a}\| < \epsilon/2$.

By the triangle inequality

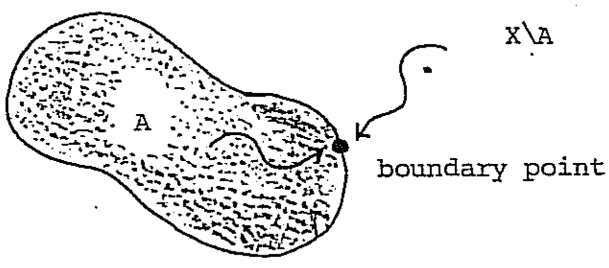
$$\begin{aligned} \|\mathbf{x} - \mathbf{a}\| &= \|\mathbf{x} - \mathbf{y} + \mathbf{y} - \mathbf{a}\| \\ &\leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathbf{a}\| \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} . \end{aligned}$$

That is, $\|\mathbf{x} - \mathbf{a}\| < \epsilon$ and so by the lemma $\mathbf{x} \in \bar{A}$.

(ii) Let $\mathbf{x} \in \bar{A}$, then there exists a sequence $(\mathbf{a}_n) \subset A$ with $\mathbf{a}_n \to \mathbf{x}$.

But $A \subseteq B$ so $(\mathbf{a}_n)$ is a sequence of elements of B converging to $\mathbf{x}$ and so, since B is closed, we must have $\mathbf{x} \in B$. Thus $\bar{A} \subseteq B$.

□



The boundary of $A \subseteq \mathbb{R}^n$ is defined to be

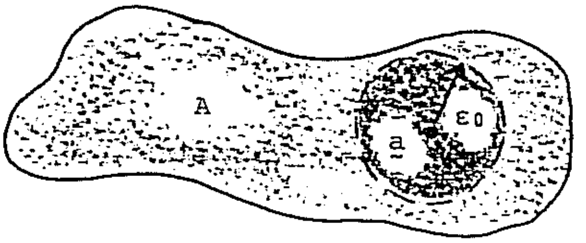
$$\text{bdry } A := \bar{A} \bigcap \overline{(\mathbb{R}^n \backslash A)}$$

That is, x $\in$ bdry A if and only if there exist sequences of points from both A and the complement of A which converge to x.

A point of A which is not in the boundary of A is said to be an **interior point** of A. The set of all interior points of A, denoted by int$A$, is referred to as the **interior** of A. A set all of whose points are interior is said to be an **open set.**

The previous lemma leads to an important characterization of interior points.

Theorem: *If* a *is an interior point of* $A \subset \mathbb{R}^n$ *then there exists* $\epsilon_0 > 0$ *such that* $B_{\epsilon_0}(\mathrm{a}) \subseteq A$. *That is, there is a ball centred on* a *lying entirely in* A.



[The converse is also true, and is left as an exercise.]

Proof: Suppose no such ball exists. Then for each $\epsilon > 0$ we must have $B_\epsilon(\mathrm{a}) \not\subseteq A$; that is, for each $\epsilon > 0$ there exists $\mathrm{x} \notin A$ with $\mathrm{x} \in B_\epsilon(\mathrm{a})$. In other words, for each $\epsilon > 0$ there exists $\mathrm{x} \in X \backslash A$ with $\|\mathrm{a} - \mathrm{x}\| < \epsilon$, and so by the lemma $\mathrm{a} \in \overline{X \backslash A}$. Thus, since a $\in$ A, we have a is in bdry A contradicting the fact that a is an interior point of A.

□

Continuity

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **continuous** at x_0 if for each $\epsilon > 0$ there is a $\delta > 0$ so that

$$|f(\mathrm{x}) - f(\mathrm{x}_0)| < \epsilon \text{ whenever } \|\mathrm{x} - \mathrm{x}_0\| < \delta ,$$

or equivalently, so that

$$f(\mathrm{x}_0) - \epsilon < f(\mathrm{x}) < f(\mathrm{x}_0) + \epsilon \text{ whenever } \|\mathrm{x} - \mathrm{x}_0\| < \delta .$$

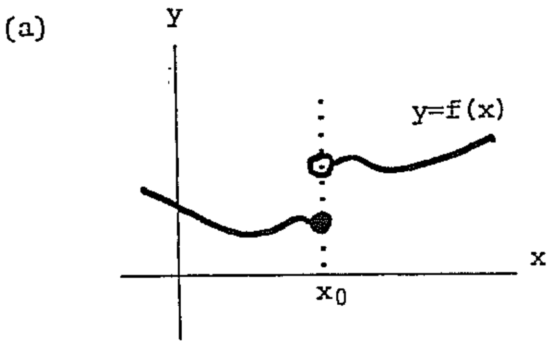
From this we see that continuity at x_0 in fact imposes two conditions on f :

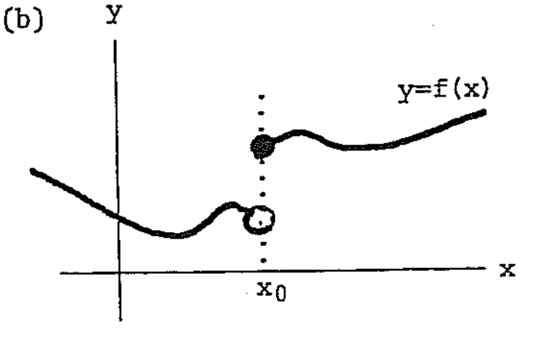
(1) For each $\epsilon > 0$ there is a $\delta > 0$ so that $f(\mathrm{x}_0) - \epsilon < f(\mathrm{x})$ whenever $\|\mathrm{x} - \mathrm{x}_0\| < \delta$.

(2) For each $\epsilon > 0$ there is a $\delta > 0$ so that $f(\mathrm{x}) < f(\mathrm{x}_0) + \epsilon$ whenever $\|\mathrm{x} - \mathrm{x}_0\| < \delta$.

We will say that a function f satisfying condition (1) is **lower semi-continuous** at x_0. Similarly, a function satisfying condition (2) is **upper semi-continuous** at x_0. In terms of these definitions f is continuous at x_0 if and only if it is both lower and upper semi-continuous at x_0 .

If you can see why the function illustrated in (a) below is lower semi-continuous, but not upper semi-continuous at x_0, while that illustrated in (b) is upper semi-continuous, but not lower semi-continuous at x_0, then you understand what the definitions are about.





A function which is lower semi-continuous, upper semi-continuous, or continuous at every point x_0 of its domain is said to be (globally) lower semi-continuous, upper semi-continuous, or continuous respectively.

Perhaps not surprisingly, lower (upper) semi-continuity will be an important condition if we want to ensure a function achieves a minimum (maximum) value.

Subsequences and Compactness

Let $(\mathbf{x}_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^n$ and let $(n_k)_{k=1}^{\infty}$ be a strictly increasing sequence of natural numbers, then the sequence

$$\mathbf{x}_{n_1}, \mathbf{x}_{n_2}, \ldots, \mathbf{x}_{n_k}, \ldots,$$

which we denote by $(\mathbf{x}_{n_k})_{k=1}^{\infty}$, is a **subsequence** of $(\mathbf{x}_n)$.

For example: If $x_n = \frac{1}{n}$ and we take $n_k = 2^k$ then we obtain the subsequence (x_{n_k}) ;

$$\frac{1}{2}, \ \frac{1}{4}, \ \frac{1}{8}, \ \ldots, \ \frac{1}{2^k}, \ \ldots,$$

of the sequence (x_n);

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \frac{1}{9}, \ldots, \ \frac{1}{n}, \ldots \qquad .$$

Intuitively, we obtain a subsequence by deleting terms from the sequence, the only restrictions are that the remaining terms stay in their original order and that there be an infinite number of them.

Note that a subsequence of a subsequence of $(\mathbf{x}_n)$ is itself a sub-sequence of $(\mathbf{x}_n)$.

Proposition: $(\mathbf{x}_n)$ converges to $\mathbf{x}$ if and only if every subsequence of $(\mathbf{x}_n)$ converges to $\mathbf{x}$.

Proof: ($\Leftarrow$) If every subsequence converges to $\mathbf{x}$, then since $(\mathbf{x}_n)$ is a subsequence of itself we have that $\mathbf{x}_n \to \mathbf{x}$.

($\Rightarrow$) Since $\mathbf{x}_n \to \mathbf{x}$, given $\epsilon > 0$ there is $N \in \mathbb{N}$ with $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$ whenever $n \geq N$. Now since n_k is a strictly increasing sequence of natural numbers we have $n_k \geq k$, and so $\|\mathbf{x}_{n_k} - \mathbf{x}\| < \epsilon$ whenever $k \geq N$. That is, $\mathbf{x}_{n_k} \to \mathbf{x}$.

□

Definition: A subset K of $\mathbb{R}^n$ is **compact** if *every* sequence of points in K has a subsequence which converges to a point of K.

In order to help understand compactness try to see why $\mathbb{R}$ is not a compact set.

The importance for optimization of compactness and the other concepts developed so far is illustrated by the following fundamental existence theorem for optimal solutions.

Theorem: *A lower semi-continuous function f defined on a nonempty compact set K achieves its minimum on K; that is, there exists $\mathbf{x}_0 \in K$ with $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for all $\mathbf{x} \in K$.*

Proof: Let $m := \inf\{f(\mathbf{x}) \ : \ \mathbf{x} \in K\}$ Then there exists a sequence of function values $f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n), \ldots$ which converges to m from above. The underlying sequence $(\mathbf{x}_n)$, being a sequence of points in the compact set K, has a subsequence $(\mathbf{x}_{n_k})$ which converges to a point $\mathbf{x}_0$ of K.

We will show that $f(\mathbf{x}_0) = m$ thereby establishing the result.

From the definition of m we already know that $m \leq f(\mathbf{x}_0)$. On the other hand, f is lower semi-continuous and $\mathbf{x}_{n_k} \to \mathbf{x}_0$. Thus given any $\epsilon > 0$ for all sufficiently large k we will have $\mathbf{x}_{n_k}$ sufficiently near to $\mathbf{x}_0$ so that $f(\mathbf{x}_0) - \epsilon < f(\mathbf{x}_{n_k})$.

In particular then

$$f(\mathbf{x}_0) - \epsilon \leq \lim_k f(\mathbf{x}_{n_k}) = m \ .$$

But, this is true for all $\epsilon > 0$ and so we conclude that $f(\mathbf{x}_0) \leq m$. Hence $f(\mathbf{x}_0) = m$, as required to complete the proof.

□

Corollary (1): *If K is a nonempty compact set and $f : K \to \mathbb{R}$ is upper semi-continuous then f achieves its maximum on K.*

Proof: If f is upper semi-continuous then $-f$ is lower semi-continuous and a maximum for f is a minimum for $-f$.

□

Corollary (2): *If K is a nonempty compact set and $f : K \to \mathbb{R}$ is continuous then f achieves both its maximum and minimum on K.*

In order to apply these last results we must be able to identify compact sets. Accordingly the remainder of the chapter is devoted to characterizing the compact subsets of $\mathbb{R}^n$.

Some **necessary conditions** for compactness are provided by the next two results.

Proposition: *A compact set is closed.*

Proof: Let K be a compact set and let $\mathbf{a}_n$ be a sequence of points of K converging to $\mathbf{x}$. We must show $\mathbf{x} \in K$. Now K is compact so there exists a subsequence $(\mathbf{a}_{n_k})$ converging to a **point** of K.

But, since $(\mathbf{a}_n)$ is convergent to $\mathbf{x}$, every subsequence of $(\mathbf{a}_n)$ also converges to $\mathbf{x}$. In particular $(\mathbf{a}_{n_k})$ converges to $\mathbf{x}$ and so we conclude that $\mathbf{x} \in \overline{K} = K$.

□

Proposition: *A compact set is bounded.*

Proof: Let K be a compact set. Since the function $f(\mathbf{x}) := \|\mathbf{x}\|$ is continuous (see exercises), it achieves its maximum on K. Let M be the maximum value of f on K, then necessarily M is finite (it is achieved as a value of f) and $\|\mathbf{x}\| \leq M$ for all $\mathbf{x} \in K$. That is, K is bounded.

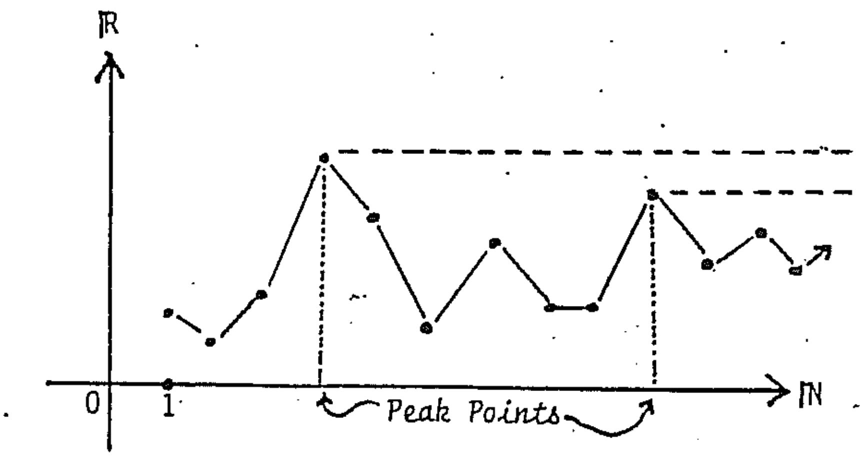
□

Results obtained up to this point in the chapter remain valid if $\mathbb{R}^n$ is replaced by any normed linear space X. This is not true for the remaining results. They depend in an essential way on X being finite dimensional.

We have shown so far that any compact set is closed and bounded. We now aim to show that in $\mathbb{R}^n$ the converse is true. This is, any closed bounded subset of $\mathbb{R}^n$ is compact. We begin by proving the result when $n = 1$; that is, in $\mathbb{R}$.

Lemma: *Any bounded sequence of real numbers has a convergent subsequence.*

Proof: (x_n) be a bounded sequence of points of $\mathbb{R}$, it suffices to show that there is a subsequence $(x_{n_k})_{k=1}^{\infty}$ which is either decreasing or increasing, for then, (x_{n_k}) being bounded both above and below is necessarily convergent.



all subsequent terms lie beneath the dotted lines

Call $n \in \mathbb{N}$ a 'peak point' of (x_n) if $x_n > x_m$ for all $m > n$, then we have two possibilities:

(1) (x_n) has an infinite number of peak points at $n_1 < n_2 < n_3 < \ldots$, in which case $x_{n_1} > x_{n_2} > x_{n_3} > \ldots$ and so $(x_{n_k})_{k=1}^{\infty}$ is the required decreasing subsequence.

(2) (x_n) has only finitely may peak points at $n_1 < n_2 < \ldots < n_m$ say, in this case let $N_1 > n_m$, then N_1 is not a peak point so there exists $N_2 > N_1$ with $x_{N_1} < x_{N_2}$, further

something missing

Theorem: *A subset of $\mathbb{R}^m$ is compact if and only if it is closed and bounded.*

Proof: ($\Rightarrow$) has already been proved.

($\Leftarrow$) Let K be a closed bounded subset of $\mathbb{R}^m$ with $\|\mathbf{x}\| \leq M$ for all $\mathbf{x} \in K$ and let $(\mathbf{x}_n)$ be a sequence of points of K, where $\mathbf{x}_n = (x_n(1), x_n(2), \ldots x_n(m))$. That is, $x_n(j)$ denotes the j'th component of the n'th vector in the sequence. For all n and j we have

$$|x_n(j)| \leq \sqrt{x_n(1)^2 + x_n(2)^2 + \ldots + x_n(m)^2} = \|\mathbf{x_n}\| \leq M .$$

Consequently, for each j the sequence of j'th components; $(x_n(j))_{n=1}^{\infty}$, is a bounded sequence of real numbers.

In particular then $(x_n(1))$ is a bounded sequence of real numbers and so, applying the previous lemma, there is a subsequence $(x_{n_k}(1))_{k=1}^{\infty}$ convergent to some real number x_1. The corresponding subsequence $(\mathbf{x}_{n_k})$ of $(\mathbf{x}_n)$ therefore has its sequence of first components converging to x_1. Similarly $x_{n_k}(2)_{k=1}^{\infty}$, the sequence of second components of $(\mathbf{x}_{n_k})$, is a bounded sequence of real numbers and so we may pass to a further subsequence $(\mathbf{x}_{n_{k_\ell}})_{\ell=1}^{\infty}$ with $x_{n_{k_\ell}}(1) \to x_1$ and $x_{n_{k_\ell}}(2) \to x_2$ for some real number x_2. Continuing in this way we eventually arrive at a subsequence of $(\mathbf{x}_n)$ which for simplicity we will denote by $(\mathbf{x}_{p_q})$ with $x_{p_q}(j) \to x_j$, for $j = 1, 2, \ldots, m$. (see diagram on next page).

The proof is completed by showing that $(\mathbf{x}_{p_q}) \to \mathbf{x} := (x_1, x_2, \ldots, x_m)$. (That $\mathbf{x} \in K$ follows automatically, since $(\mathbf{x}_{p_q})$ is a sequence in K and K is closed.)

Now,

$$\begin{aligned}\|\mathbf{x}_{p_q} - \mathbf{x}\| &= \sqrt{(x_{p_q}(1) - x_1)^2 + (x_{p_q}(2) - x_2)^2 + \ldots + (x_{p_q}(m) - x_m)^2} \\ &\to 0, \quad \text{as } x_{p_q}(j) - x_j \to 0 \text{ for } j = 1, 2, \ldots, m,\end{aligned}$$

and so $\mathbf{x}_{p_k}$ is a subsequence of $(\mathbf{x}_n)$ which converges to $\mathbf{x} \in K$ as required.

□

$\underline{x}_1 = (x_1(1), x_1(2), x_1(3), \ldots, x_1(m))$
$\underline{x}_2 = (x_2(1), x_2(2), x_2(3), \ldots, x_2(m))$
$\underline{x}_3 = (x_3(1), x_3(2), x_3(3), \ldots, x_3(m))$
$\underline{x}_4 = (x_4(1), x_4(2), x_4(3), \ldots, x_4(m))$
$\underline{x}_5 = (x_5(1), x_5(2), x_5(3), \ldots, x_5(m))$
$\underline{x}_6 = (x_6(1), x_6(2), x_6(3), \ldots, x_6(m))$
$\underline{x}_7 = (x_7(1), x_7(2), x_7(3), \ldots, x_7(m))$
... ...

$\underline{x}_1 = (x_1(1), x_1(2), x_1(3), \ldots, x_1(m))$
$\underline{x}_2 =$ ~~$(x_2(1)$~~$, x_2(2), x_2(3), \ldots, x_2(m))$
$\underline{x}_3 = (x_3(1), x_3(2), x_3(3), \ldots, x_3(m))$
$\underline{x}_4 = (x_4(1), x_4(2), x_4(3), \ldots, x_4(m))$
$\underline{x}_5 =$ ~~$(x_5(1)$~~$, x_5(2), x_5(3), \ldots, x_5(m))$
$\underline{x}_6 = (x_6(1), x_6(2), x_6(3), \ldots, x_6(m))$
$\underline{x}_7 = (x_7(1), x_7(2), x_7(3), \ldots, x_7(m))$
... ...
↓
x_1

$\underline{x}_1 = (x_1(1), x_1(2), x_1(3), \ldots, x_1(m))$
~~$\underline{x}_2 = (x_2(1), x_2(2), x_2(3), \ldots, x_2(m))$~~
$\underline{x}_3 = (x_3(1), x_3(2), x_3(3), \ldots, x_3(m))$
$\underline{x}_4 = (x_4(1), x_4(2), x_4(3), \ldots, x_4(m))$
~~$\underline{x}_5 = (x_5(1), x_5(2), x_5(3), \ldots, x_5(m))$~~
$\underline{x}_6 = (x_6(1), x_6(2), x_6(3), \ldots, x_6(m))$
$\underline{x}_7 = (x_7(1), x_7(2), x_7(3), \ldots, x_7(m))$
... ...
↓
x_1

$\underline{x}_1 = (x_1(1), x_1(2), x_1(3), \ldots, x_1(m))$
~~$\underline{x}_2 = (x_2(1), x_2(2), x_2(3), \ldots, x_2(m))$~~
$\underline{x}_3 = (x_3(1),$ ~~$x_3(2)$~~$, x_3(3), \ldots, x_3(m))$
$\underline{x}_4 = (x_4(1), x_4(2), x_4(3), \ldots, x_4(m))$
~~$\underline{x}_5 = (x_5(1), x_5(2), x_5(3), \ldots, x_5(m))$~~
$\underline{x}_6 = (x_6(1), x_6(2), x_6(3), \ldots, x_6(m))$
$\underline{x}_7 = (x_7(1),$ ~~$x_7(2)$~~$, x_7(3), \ldots, x_7(m))$
... ...
↓ ↓
x_1 x_2

$\underline{x}_1 = (x_1(1), x_1(2), x_1(3), \ldots, x_1(m))$
~~$\underline{x}_2 = (x_2(1), x_2(2), x_2(3), \ldots, x_2(m))$~~
~~$\underline{x}_3 = (x_3(1), x_3(2), x_3(3), \ldots, x_3(m))$~~
$\underline{x}_4 = (x_4(1), x_4(2), x_4(3), \ldots, x_4(m))$
~~$\underline{x}_5 = (x_5(1), x_5(2), x_5(3), \ldots, x_5(m))$~~
$\underline{x}_6 = (x_6(1), x_6(2), x_6(3), \ldots, x_6(m))$
~~$\underline{x}_7 = (x_7(1), x_7(2), x_7(3), \ldots, x_7(m))$~~
... ...
↓ ↓
x_1 x_2

$\underline{x}_1 = (x_1(1), x_1(2), x_1(3), \ldots, x_1(m))$
~~$\underline{x}_2 = (x_2(1), x_2(2), x_2(3), \ldots, x_2(m))$~~
~~$\underline{x}_3 = (x_3(1), x_3(2), x_3(3), \ldots, x_3(m))$~~
$\underline{x}_4 = (x_4(1), x_4(2),$ ~~$x_4(3)$~~$, \ldots, x_4(m))$
~~$\underline{x}_5 = (x_5(1), x_5(2), x_5(3), \ldots, x_5(m))$~~
$\underline{x}_6 = (x_6(1), x_6(2), x_6(3), \ldots, x_6(m))$
~~$\underline{x}_7 = (x_7(1), x_7(2), x_7(3), \ldots, x_7(m))$~~
... ...
↓ ↓ ↓
x_1 x_2 x_3

$\underline{x}_1 = (x_1(1), x_1(2), x_1(3), \ldots, x_1(m))$
~~$\underline{x}_2 = (x_2(1), x_2(2), x_2(3), \ldots, x_2(m))$~~
~~$\underline{x}_3 = (x_3(1), x_3(2), x_3(3), \ldots, x_3(m))$~~
~~$\underline{x}_4 = (x_4(1), x_4(2), x_4(3), \ldots, x_4(m))$~~
~~$\underline{x}_5 = (x_5(1), x_5(2), x_5(3), \ldots, x_5(m))$~~
$\underline{x}_6 = (x_6(1), x_6(2), x_6(3), \ldots, x_6(m))$
~~$\underline{x}_7 = (x_7(1), x_7(2), x_7(3), \ldots, x_7(m))$~~
... ...
↓ ↓ ↓
x_1 x_2 x_3

etc.

Exercises

(1) Using the 3 norm properties, show that
$|\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x} - \mathbf{y}\|$.

Hence deduce that the function $\mathbf{x} \mapsto \|\mathbf{x}\|$ is continuous.

(2) Show that $\|\mathbf{x}\| := \max\{|x_1|, |x_2|, \ldots, |x_n|\}$ and $\|\mathbf{x}\| := |x_1| + |x_2| + \ldots + |x_n|$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ are both norm functions on $\mathbb{R}^n$.

(3) Show that $B_r(\mathbf{x}) = \mathbf{x} + B_r(0) := \{\mathbf{x} + \mathbf{y} : \mathbf{y} \in B_r(0)\}$.

(4) Show that $S \subset \mathbb{R}^n$ is bounded if and only if it is contained in some ball not necessarily centred on the origin.

(5) Show that a convergent sequence has a unique limit; that is, if $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{x}_n \to \mathbf{y}$ then $\mathbf{x} = \mathbf{y}$.

*(6) Show that a convergent sequence $(\mathbf{x}_n)$ is bounded; that is, $\{\mathbf{x}_n : n \in \mathbb{N}\}$ is a bounded set.

(7) Show that the intersection of two closed sets is closed.

*(8) Show that the union of two closed sets is closed.

(9) Show that $\overline{B_r(\mathbf{x})} = B_r[\mathbf{x}]$, for any $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$.

*(10) For any set $A \subseteq \mathbb{R}^n$ show that
$$\bar{A} = \bigcap_n \, (A + B_{1/n}(0))$$

(11) If $\mathbf{a} \in A \subseteq \mathbb{R}^n$ is such that for some $r > 0$ we have $B_r(\mathbf{a}) \subseteq A$, show that $\mathbf{a}$ is an interior point of A.

(12) For any set $A \subseteq \mathbb{R}^n$ show that $\bar{A} = (\text{bdry}A) \cup (\text{int}A)$.

(13) Show that any finite subset of $\mathbb{R}^N$ is compact.

(14) Show that a closed subset of a compact set is itself compact.

(15) Show that in $\mathbb{R}^n$ the closest point problem to a compact set has a solution. That is, if K is a nonempty compact subset of $\mathbb{R}^n$ and $\mathbf{x} \notin K$ then there exists a point $\mathbf{k}_0$ of K with $\|\mathbf{x} - \mathbf{k}_0\| \leq \|\mathbf{x} - \mathbf{k}\|$, for all $\mathbf{k} \in K$.

*(16) Show that in $\mathbb{R}^n$ the conclusion of (15) remains valid if it is only assumed that K is nonempty and closed.

*(17) Show that the image of a compact set under a continuous function is compact. That is, if $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous and $K \subset \mathbb{R}^n$ is compact, then $f(K) := \{f(\mathrm{x}) : \mathrm{x} \in K\}$ is a compact subset of $\mathbb{R}^m$.

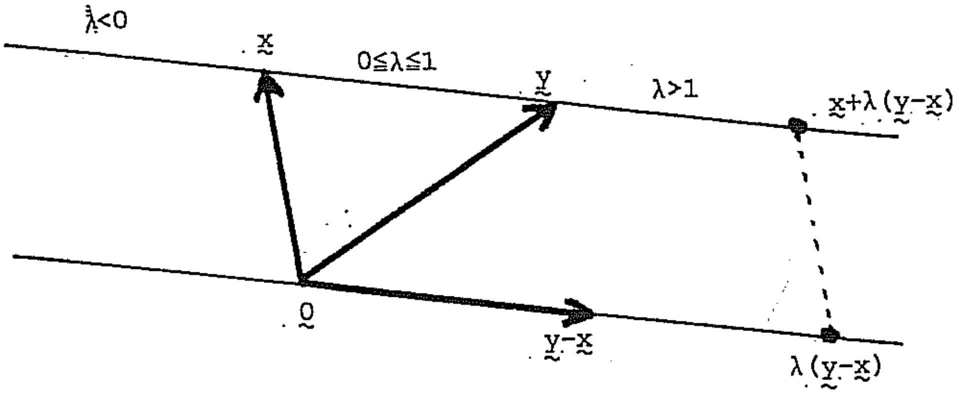
CHAPTER 3 - Convexity

3.1 Convex Sets

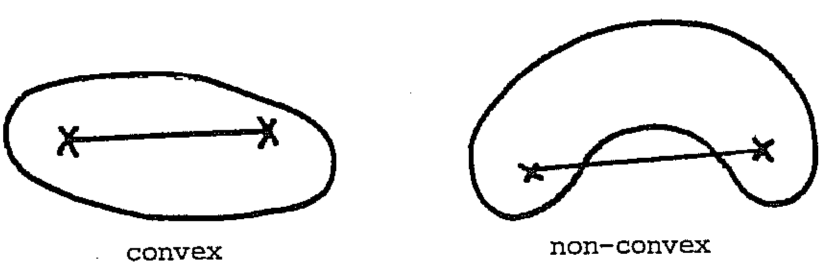
Let us begin by recalling that if $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are two points in $\mathbb{R}^n$ (or indeed any vector space), then a point on the line through x and y is of the form:

$$x + \lambda(y - x) = (1 - \lambda)x + \lambda y,$$

where λ is a real number. Points on the segment between x and y correspond to λ values in [0, 1]. See illustration below.



Some Definitions: A subset $C \subseteq \mathbb{R}^n$ is **convex** if it contains the line segment between every pair of points in it. That is, if $\lambda x + (1 - \lambda)y \in C$ whenever $x, y \in C$ and $\lambda \in [0, 1]$.

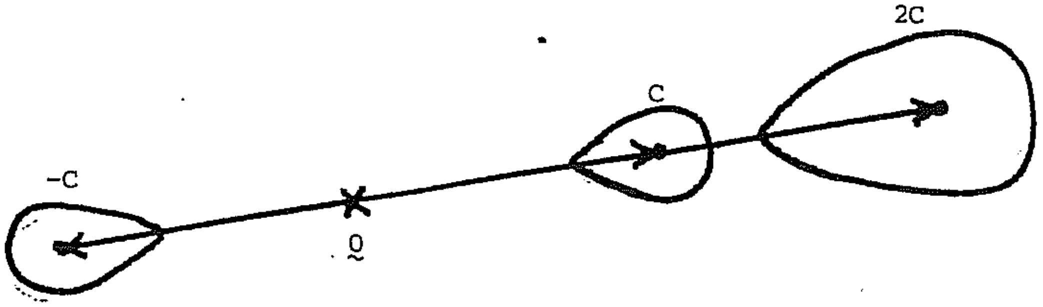


A subset $A \subseteq \mathbb{R}^n$ is **afine** (a **linear variety**) if it contains the line through every pair of points in it. That is, if $\lambda x + (1 - \lambda)y \in A$ whenever $x, y \in A$ and $\lambda \in \mathbb{R}$.

Clearly, affine $\Rightarrow$ convex

Facts:

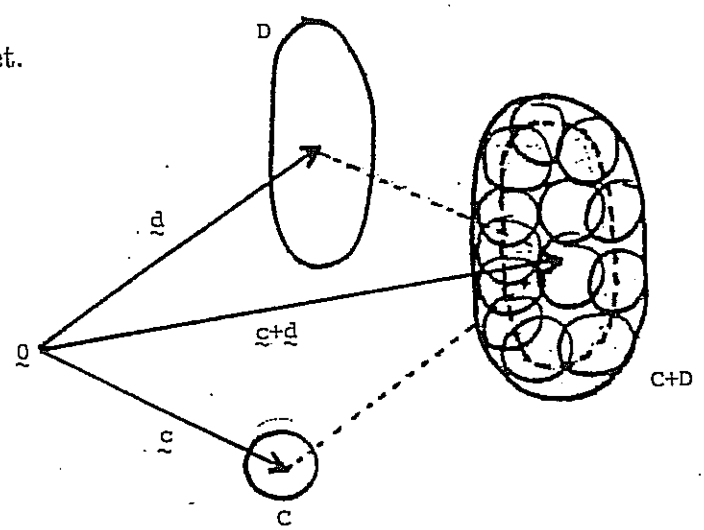
(1) If $\alpha \in \mathbb{R}$ and C is a convex set then $\alpha C := \{\alpha \mathbf{x} : \mathbf{x} \in C\}$ is convex.



(2) If C and D are convex sets then

$$C + D := \{\mathbf{c} + \mathbf{d} \;:\; \mathbf{c} \in C,\ \mathbf{d} \in D\}$$

is a convex set.



Proof: Let $\mathbf{x}_1, \mathbf{x}_2 \in C + D$ and $\lambda \in [0, 1]$, then $\mathbf{x}_1, = \mathbf{c}_1 + \mathbf{d}_1$ and $\mathbf{x}_2 = \mathbf{c}_2 + \mathbf{d}_2$ for some $\mathbf{c}_1, \mathbf{c}_2 \in C$ and $\mathbf{d}_1, \mathbf{d}_2 \in D$.

Hence

$$\begin{aligned} \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2 &= \lambda(\mathbf{c}_1 + \mathbf{d}_1) + (1 - \lambda)(\mathbf{c}_2 + \mathbf{d}_2) \\ &= (\lambda \mathbf{c}_1 + (1 - \lambda)\mathbf{c}_2) + (\lambda \mathbf{d}_1 + (1 - \lambda)\mathbf{d}_2) \\ &\in C + D \end{aligned}$$

as $\lambda \mathbf{c}_1 + (1 - \lambda)\mathbf{c}_2 \in C$ and $\lambda \mathbf{d}_1 + (1 - \lambda)\mathbf{d}_2 \in D$ since C and D are convex sets.

(3) Any intersection of convex sets is convex.

(4) A is an affine set if and only if it is the translate of some subspace M; that is

$$A = \mathbf{x}_0 + M$$

Proof: ($\Leftarrow$) See exercise.

($\Rightarrow$) Choose any $\mathbf{x}_0 \in A$ and let $M := A - \mathbf{x}_0$.

Clearly, $A = \mathbf{x}_0 + M$, so it suffices to show that M is a subspace.

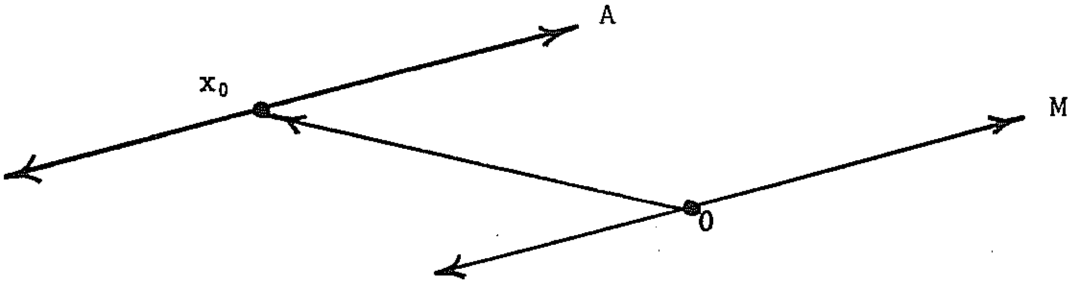
We will show M is closed under addition (that it is closed under scalar multiplication, and hence a subspace, follows similarly).

Let $\mathbf{m}_1, \mathbf{m}_2 \in M$ then there exists $\mathbf{a}_1, \mathbf{a}_2 \in A$ so that $\mathbf{m}_1 = \mathbf{a}_1 - \mathbf{x}_0$ and $\mathbf{m}_2 = \mathbf{a}_2 - \mathbf{x}_0$. Thus

$$\begin{aligned} \mathbf{m}_1 + \mathbf{m}_2 &= (\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{x}_0) - \mathbf{x}_0 \\ &= \underbrace{\underbrace{2\underbrace{\frac{\mathbf{a}_1 + \mathbf{a}_2}{2}}_{\in A} - \mathbf{x}_0}_{\in A} - \mathbf{x}_0}_{\in M} \end{aligned}$$

□

The subspace M defined in (4) is unique (that is, if $A = \mathbf{x}_0 + M$ and $A = \mathbf{y}_0 + M'$ then $M = M'$. See exercises). We refer to this unique subspace as the *subspace parallel to A*.



By the **dimension** of the affine set A we will mean the dimension of the unique subspace parallel to A; that is the dimension of $M = A - \underline{x}_0$ where $\underline{x}_0$ is any point of A.

More definitions: Let S be a subset of $\mathbb{R}^n$.

The **convex hull** of S, denoted by $\text{co}(S)$, is the smallest convex set containing S. $\text{co}(S)$ equals the intersection of all convex sets containing S.

The **closed convex hull** of S is the closure of $\text{co}(S)$ and will be denoted by $\overline{\text{co}}$ (S).

The **affine hull** of S, denoted by aff(S) is the smallest affine set containing S.

If $\mathbf{s}_0 \in S$ then

$$\text{aff}(S) = \mathbf{s}_0 + M$$

where M is the subspace spanned by the vectors in the set $S - \mathbf{s}_0$.

For a convex set C we take the **dimension of C** to be the dimension of aff(C). It equals the dimension of the subspace spanned by the vectors in $C - \mathbf{c}_0$ for any $\mathbf{c}_0 \in C$

An affine set which is the translate of an $(n-1)$-dimensional subspace in $\mathbb{R}^n$ is a **hyperplane**. Examples of hyperplanes are; a plane in $\mathbb{R}^3$, a line in $\mathbb{R}^2$.

Theorem: *H is a hyperplane in $\mathbb{R}^n$ if and only if $H = \{\mathbf{x} \in \mathbb{R}^n \ : \ \mathbf{a} \cdot \mathbf{x} = c\}$ for some non-zero vector $\mathbf{a} \in \mathbb{R}^n$ and some constant c.*

[Note: This result should come as no surprise. In two dimensions $\mathbf{a} \cdot \mathbf{x} = c$ is the general equation of a line, while in three dimensions it is the equation of a plane.]

Proof: ($\Leftarrow$) Let $H := \{\mathbf{x} : \mathbf{a} \cdot \mathbf{x} = c\}$, choose $\mathbf{x}_0 \in H$ and let $M = H - \mathbf{x}_0$ then

$$M = \{\mathbf{m} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{m} = 0\}$$

That is, M equals the collection of all vectors perpendicular to $\mathbf{a}$, which is a subspace of dimension $n-1$ (why?). Hence $H = \mathbf{x}_0 + M$ is a hyperplane.

($\Rightarrow$) Let H be a hyperplane and choose $\mathbf{x}_0 \in H$, then $M := H - \mathbf{x}_0$ is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. Choose $\mathbf{a}$ perpendicular to M then $\mathbf{m} \in M \Leftrightarrow \mathbf{a} \cdot \mathbf{m} = 0$ so

$$\begin{aligned} \mathbf{x} \in H &\Leftrightarrow \mathbf{x} - \mathbf{x}_0 \in M \\ &\Leftrightarrow \mathbf{a} \cdot (\mathbf{x} - \mathbf{x}_0) = 0 \\ &\Leftrightarrow \mathbf{a} \cdot \mathbf{x}_0 = \mathbf{a} \cdot \mathbf{x}_0 \end{aligned}$$

taking $c := \mathbf{a} \cdot \mathbf{x}_0$ we therefore have $H = \{\mathbf{x} : \mathbf{a} \cdot \mathbf{x} = c\}$ as required.

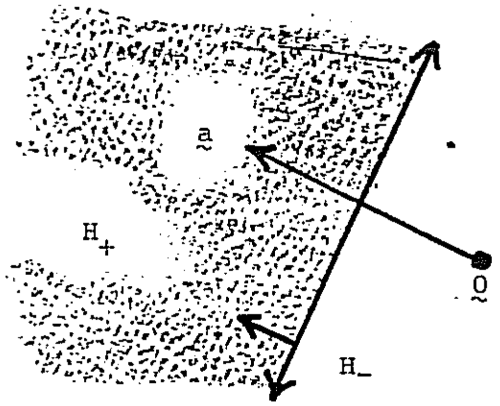
□

The hyperplane $H := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} = c\}$ defines two **closed half-spaces**

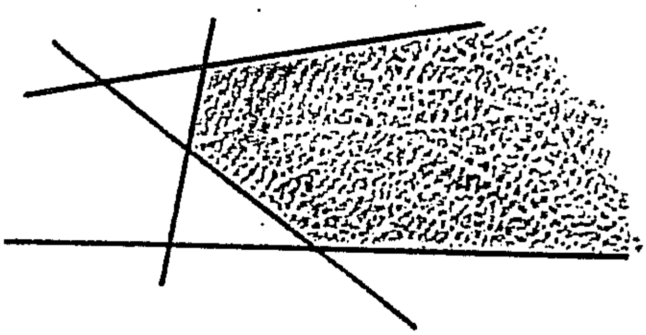
$$\begin{aligned} H_+ &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} \geq c\}, \\ H_- &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} \leq c\} \end{aligned}$$

and corresponding open half-spaces where the inequalities are required to be strict

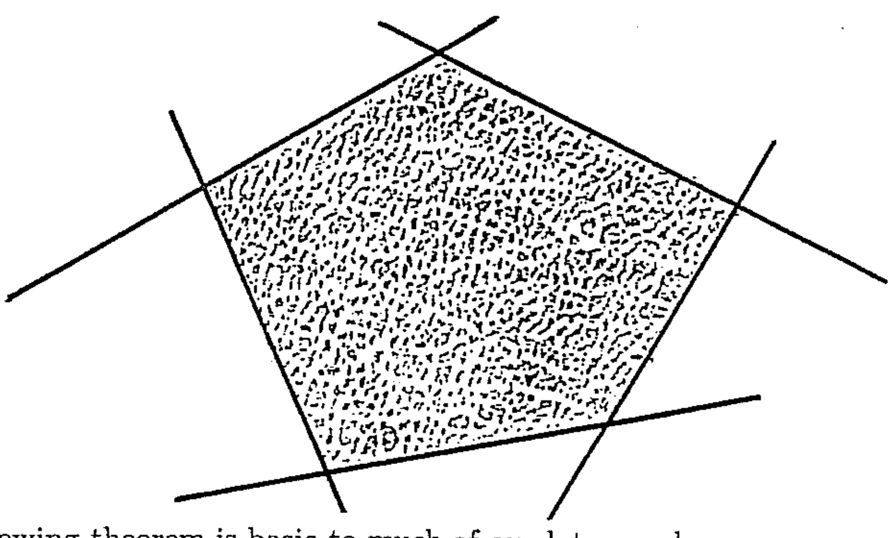
$$\begin{aligned} \overset{\circ}{H}_+ &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} > c\}, \\ \overset{\circ}{H}_- &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a} \cdot \mathbf{x} < c\} \end{aligned}$$



A convex set which is the intersection of a finite number of closed half-spaces is termed a **convex polytope**.



If in addition a polytope is bounded we call it a **polyhedron**.

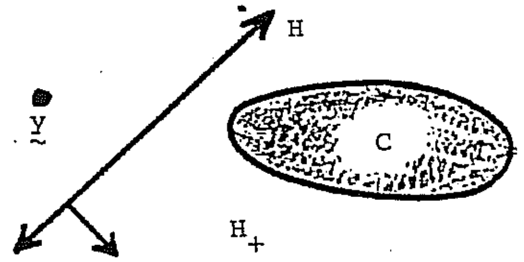


The following theorem is basic to much of our later work.

Separation Theorem: *Let C be a closed convex set in $\mathbb{R}^n$ and let y be a point outside of C. Then there exists a hyperplane $H = \{\mathrm{x} \in \mathbb{R}^n : \mathrm{a} \cdot \mathrm{x} = c\}$ with $\mathrm{y} \in \mathring{H}_-$ and $C \subseteq H_+$.*

That is,

$$\mathrm{a}\cdot\mathrm{y} < c \leq \inf_{\mathrm{x}\in C} \mathrm{a}\cdot\mathrm{x}$$



Proof: Let x_0 be a closest point of C to y (such a point exists by exercise 16 of chapter 2. Then for any $\mathrm{x}\in C$ the function

$$F(\lambda) = \|\mathrm{y} - [\lambda\mathrm{x} + (1-\lambda)\mathrm{x}_0]\|^2$$

achieves its minimum on $[0,1]$ at $\lambda = 0$.

Now

$$\begin{aligned}F(\lambda) &= \|\mathrm{y} - \lambda\mathrm{x} - (1-\lambda)\mathrm{x}_0\|^2 \\ &= (\mathrm{y} - \lambda\mathrm{x} - (1-\lambda)\mathrm{x}_0)\cdot((\mathrm{y} - \lambda\mathrm{x} - (1-\lambda)\mathrm{x}_0) \\ &= ((\mathrm{y}-\mathrm{x}_0) + \lambda(\mathrm{x}_0 - \mathrm{x}))\cdot(\mathrm{y}-\mathrm{x}_0) + \lambda(\mathrm{x}_0-\mathrm{x})) \\ &= (\mathrm{y}-\mathrm{x}_0)\cdot(\mathrm{y}-\mathrm{x}_0) + 2\lambda(\mathrm{y}-\mathrm{x}_0)\cdot(\mathrm{x}_0-\mathrm{x}) \\ &\qquad + \lambda^2(\mathrm{x}_0-\mathrm{x})\cdot(\mathrm{x}_0-\mathrm{x}) \\ &= \|\mathrm{y}-\mathrm{x}_0\|^2 + 2(\mathrm{y}-\mathrm{x}_0)\cdot(\mathrm{x}_0-\mathrm{x})\lambda + \|\mathrm{x}_0-\mathrm{x}\|^2\lambda^2\end{aligned}$$

So $F(\lambda)$ is a quadratic in λ. In particular it is differentiable on $[0,1]$ with

$$F'(\lambda) = 2(\mathrm{y}-\mathrm{x}_0)\cdot(\mathrm{x}_0-\mathrm{x}) + 2\lambda\|\mathrm{x}_0-\mathrm{x}\|^2,$$

and so we must have

$$2(\mathrm{y}-\mathrm{x}_0)\cdot(\mathrm{x}_0-\mathrm{x}) = F'(0) \geq 0\,.$$

Since this is true for all $\mathrm{x}\in C$, setting $\mathrm{a} = \mathrm{x}_0 - \mathrm{y}$ and $c = \mathrm{a}\cdot\mathrm{x}_0$ we have for all $\mathrm{x}\in C$ that

$$\begin{aligned}\mathrm{a}\cdot\mathrm{x} &\geq c \\ &= \mathrm{a}\cdot\mathrm{x}_0 \\ &= \mathrm{a}\cdot\mathrm{y} + \mathrm{a}\cdot(\mathrm{x}_0-\mathrm{y}) \\ &= \mathrm{a}\cdot\mathrm{y} + \|\mathrm{x}_0-\mathrm{y}\|^2 \\ &> \mathrm{a}\cdot\mathrm{y}\,.\end{aligned}$$

Hence $\mathrm{a}\cdot\mathrm{y} < c \leq \inf_{\mathrm{x}\in C} \mathrm{a}\cdot\mathrm{x}$ as required.

□

Corollary 1: If C is a convex subset of $\mathbb{R}^n$ and y is in the boundary of C, then there exists $\mathrm{a} \neq 0$ such that

$$\mathrm{a}\cdot\mathrm{y} \leq \mathrm{a}\cdot\mathrm{x} \text{ for all } \mathrm{x}\in C .$$

Proof: Since y is in the boundary of C we can find a sequence (y_n) with each y_n outside the closure of C and with $\mathrm{y}_n \to \mathrm{y}$. By the separation theorem, for each n there exists $\mathrm{a}_n \neq 0$ such that

$$\mathrm{a}_n\cdot\mathrm{y}_n < \inf_{\mathrm{x}\in\bar{C}} \mathrm{a}_n\cdot\mathrm{x} .$$

In particular then

$$\mathrm{a}_n\cdot\mathrm{y}_n < \mathrm{a}_n\cdot\mathrm{x} \quad \text{for all } \mathrm{x}\in C$$

Dividing both sides by $\|a_n\|$, we may without loss of generality assume that each a_n has $\|\mathrm{a}_n\| = 1$. By the theorem on p.24, there exists a subsequence (a_{n_k}) converging to some vector a. (Note, that $\|\mathrm{a}\| = \lim_k\|\mathrm{a}_{n_k}\| = 1$, so $\mathrm{a}\neq 0$.)

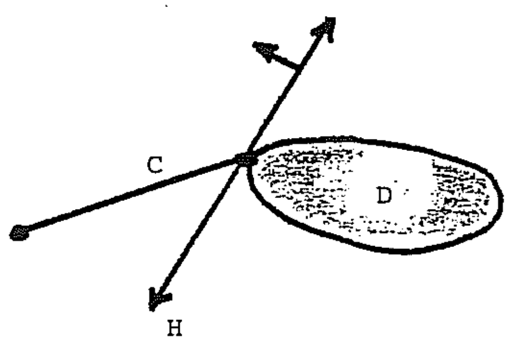
For each $\mathrm{x}\in C$ we then have

$$\mathrm{a}\cdot\mathrm{y} = \lim_k \mathrm{a}_{n_k}\cdot\mathrm{y}_{n_k} \leq \lim_k \mathrm{a}_{n_k}\cdot\mathrm{x} = \mathrm{a}\cdot\mathrm{x}$$

which establishes the result.

□

Corollary 2: Let D be a convex set with interior points and let C be a nonemptyconvex set which contains no interior points of D. Then there exists a hyperplane H with $C\subseteq H_+$ and $D\subseteq H_-$.



Proof

The interior of D is a non-empty convex set (see Exercise 8), thus $C - \mathrm{int}D$ is a nonempty convex set with $0\notin C-\mathrm{int}D$, and so either $0 \notin \overline{C-\mathrm{int}D}$ or $0\in \mathrm{bdry}\ (C-\mathrm{int}D)$.

By the separation theorem, or corollary 1, there is therefore an $a \neq 0$ so that

$$0 = a \cdot 0 \leq a \cdot x, \quad \text{whenever}$$

$x \in C - \text{int}D$.

Thus, for any $c \in C$ and $d \in \text{int}D$ we have

$$0 \leq a \cdot (c - d)$$

or

$$\underline{a} \cdot d \leq a \cdot c \; .$$

Let $c := \inf\{a \cdot c : c \in C\}$, then

then

$$a \cdot d \leq c \leq a \cdot c \; .$$

for all $c \in C$ and $d \in \text{int}D$.

That is,

$$C \subseteq H_+ \quad \text{and} \quad \text{int}D \subseteq H_- \, ,$$

where

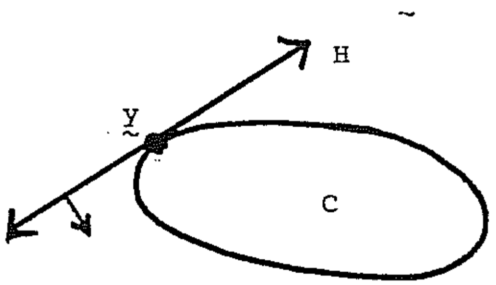
$$H := \{x : a \cdot x = c\} \; .$$

Since H_- is closed, it follows from exercise 8 and the theorem on page 16 that

$$D \subseteq \overline{\text{int D}} \subseteq H_- \; .$$

□

For a convex set C and y in the boundary of C a hyperplane H is a **support hyperplane for C at y** if $C \subseteq H_+$ and $y \in H$.



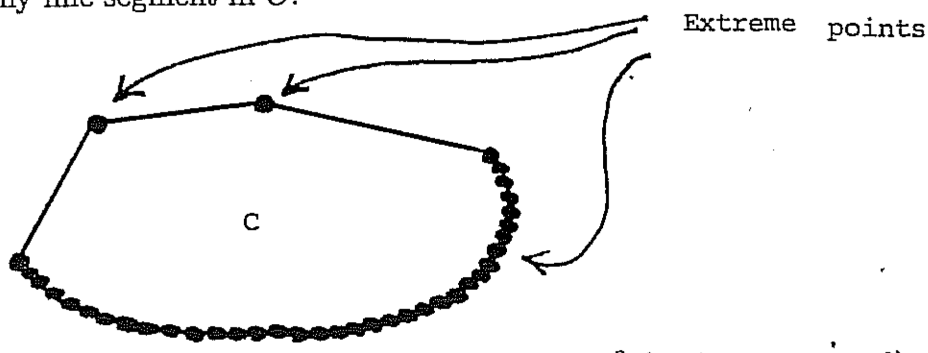
Corollary 1 ensures the existence of a support hyperplane at each boundary point y of C; namely:

$$H := \{x : a \cdot x \; c\}$$

with $c = \mathbf{a} \cdot \mathbf{y}$ where $\mathbf{a}$ is given by corollary 1.

Extremal Structure

Definition: A point $\mathbf{x}$ of the convex set C is an **extreme point** of C if whenever $\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ with $\lambda \in (0,1)$ and $\mathbf{x}_1, \mathbf{x}_2, \in C$ we have $\mathbf{x}_1 = \mathbf{x}_2$. That is $\mathbf{x}$ does not belong to the interior of any line segment in C.



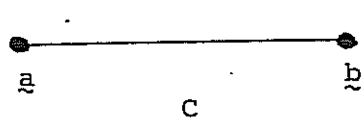
For example, the extreme points of a convex polyhedron in $\mathbb{R}^2$ (in fact in general) are its vertices.

Theorem: *A closed bounded convex set C of $\mathbb{R}^n$ is the convex hull of its extreme points.*

Proof: Our proof is by induction on n the dimension of C.

When $n = 1$ C is a closed bounded line segment

$$C = \{\mathbf{x} : \mathbf{x} = \lambda \mathbf{a} + (1-\lambda)\mathbf{b}, 0 \leq \lambda \leq 1\}$$

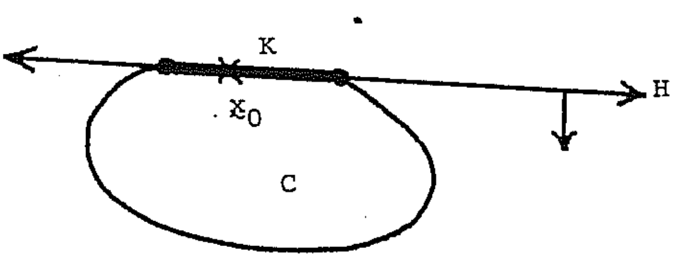


and the result is clearly true as $C = \text{co}\{\mathbf{a}, \mathbf{b}\}$. Now, suppose it is true for convex sets of dimension $n-1$ or less and let C be a convex set of dimension n.

Let $\mathbf{x}_0 \in C$. We consider two cases.

Case 1: $\mathbf{x}_0$ is a **boundary point** of C. By corollary 1 above there exists a hyperplane $H = \{\mathbf{x} : \mathbf{a} \cdot \mathbf{x} = c\}$ which supports C at $\mathbf{x}_0$. Let $K = C \cap H$, then $\mathbf{x}_0 \in K$ and K is a closed bounded convex set of dimension at most $n-1$ (as $K \subset H$).

So, by the induction hypothesis $\mathbf{x}_0$ is in the convex hull of the extreme points of K.



That $\mathbf{x}_0$ is therefore in the convex hull of the extreme points of C now follows from the following lemma for which we interrupt the proof.

Lemma: *For K and C as above, the extreme points of K are extreme points of C.*

Proof: (of lemma). Let $\mathbf{x}$ be an extreme point of K and suppose $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ where $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $\lambda \in (0,1)$. We show that $\mathbf{x}_1 = \mathbf{x}_2$ and hence that $\mathbf{x}$ is also an extreme point of C.

Since $\mathbf{x} \in K \cap H$, we have

$$c = \mathbf{a}\cdot\mathbf{x} = \lambda\mathbf{a}\cdot\mathbf{x}_1 + (1-\lambda)\mathbf{a}\cdot\mathbf{x}_2 \ ,$$

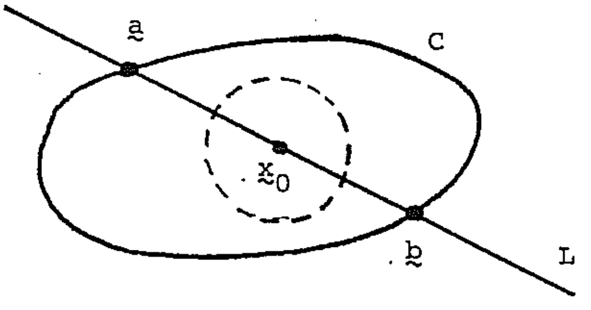
while

$$\mathbf{a}\cdot\mathbf{x}_i \geq c \text{ for } i = 1,2, \text{ as } C \subseteq H_+ \ .$$

It follows that $\mathbf{a}\cdot\mathbf{x}_1 = \mathbf{a}\cdot\mathbf{x}_2 = c$ and so both $\mathbf{x}_1$ and $\mathbf{x}_2$ are in H and therefore in K. Since $\mathbf{x}$ is an extreme point of K we must therefore have $\mathbf{x}_1 = \mathbf{x}_2$ as required.

We now return to the proof of the theorem..

Case 2: x_0 **is an interior point of** C. Choose any Line L through x_0, then $L \cap C$ is a line segment whose end points a and b are boundary points of C.



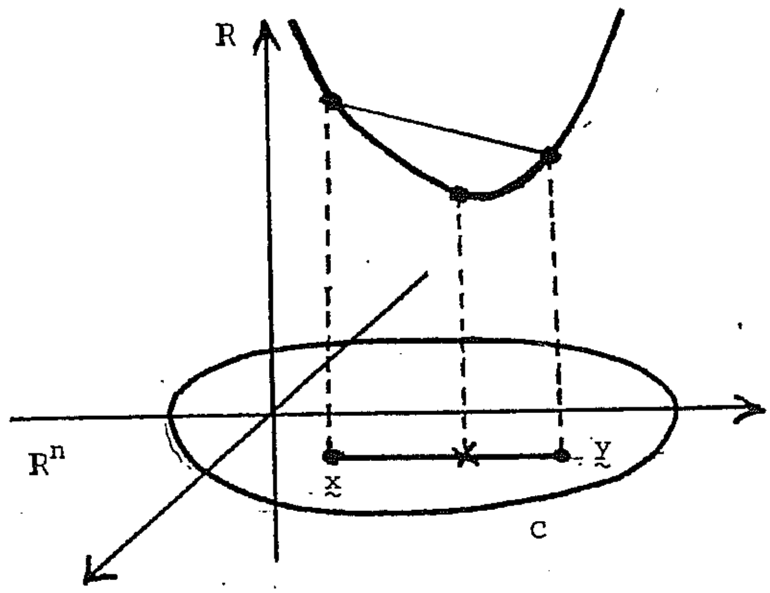
By Case 1, a and b are in the convex hull of the extreme points of C and x_0 is a convex combinaiton of a and b. Thus, since the convex hull of the extreme points of C is, by definition, a convex set, we must have that x_0 is in it.

□

3.2 CONVEX FUNCTIONS

Definition: For C a convex subset of $\mathbb{R}^n$, a function $f : C \to \mathbb{R}$ is **convex** if for $\mathrm{x}, \mathrm{y} \in C$ and $\lambda \in [0,1]$ we have

$$f(\lambda \mathrm{x} + (1-\lambda)\mathrm{y}) \leq \lambda f(\mathrm{x}) + (1-\lambda) f(\mathrm{y}) .$$



We say f is **strictly convex** if the inequality is strict whenever $\lambda \in (0,1)$ and $\mathrm{x} \neq \mathrm{y}$.

Convex functions are precisely those functions which satisfy **Jensen's inequality**

$$f\left(\sum_{i=1}^{k} \lambda_i \mathrm{x}_i \right) \leq \sum_{i=1}^{k} \lambda_i f(\mathrm{x}_i)$$

whenever

$$\sum_{i=1}^{k} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \text{ for } i = 1, 2, \ldots, k . \text{ (Prove this.)}$$

We say f is **concave** if $-f$ is convex.

Example: Any linear function $f : \mathbb{R}^n \to \mathbb{R}$ is both convex and concave.

Theorem: *Let C be a convex subset of $\mathbb{R}^n$ and let $f : C \to \mathbb{R}$ be a convex function. If f has a local minimum at $\mathrm{x}_0 \in C$, then f has a global minimum at x_0.*

Proof: Given $x \in C$ by choosing $\lambda > 0$ sufficiently small we can make $\lambda x + (1-\lambda)x_0$ as close to x_0 as we please, and so for a sufficiently small λ we have

$$\begin{aligned} f(x_0) &\leq f(\lambda x + (1-\lambda)x_0), \text{ as } f \text{ has a local minimum at } x_0 \\ &\leq \lambda f(x) + (1-\lambda)f(x_0), \text{ as } f \text{ is convex.} \end{aligned}$$

Hence

$$0 \leq \lambda[f(x) - f(x_0)]$$

or, since $\lambda > 0$

$$f(x_0) \leq f(x)$$

as required.

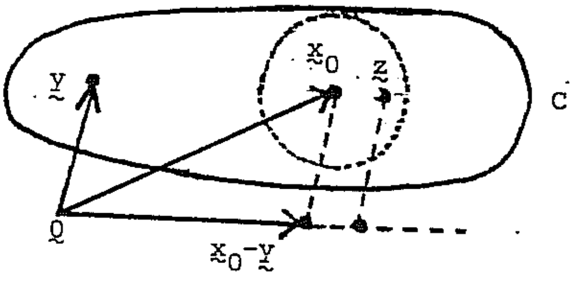
□

Theorem: *Let C be a convex subset of $\mathbb{R}^n$ and let $f : C \to \mathbb{R}$ be a convex function. If f attains a global maximum at some point x_0 in the interior of C, then f is constant on C.*

Proof: Suppose not, then there exists $y \in C$ with $f(y) < f(x_0)$. Now choosing $\mu > 1$ sufficiently near to 1 we have that

$$z := y + \mu(x_0 - y)$$

is arbitrarily close to x_0 and hence in C (x_0 is in the interior of C).



Rearranging we see that

$$x_0 = \frac{1}{\mu}z + \frac{\mu - 1}{\mu}y$$

is a convex combination of $\mathbf{z}$ and $\mathbf{y}$.

Hence

$$
\begin{aligned}
f(\mathbf{x}_0) &\leq \frac{1}{\mu} f(\mathbf{z}) + \frac{\mu - 1}{\mu} f(\mathbf{y}) \text{ , by the convexity of } f \\
&< \frac{1}{\mu} f(\mathbf{x}_0) + \frac{\mu - 1}{\mu} f(\mathbf{x}_0), \text{ as } f(\mathbf{z}) \leq f(\mathbf{x}_0), \text{ (maximum at } \mathbf{x}_0\text{), and } f(\mathbf{y}) < f(\mathbf{x}_0) \\
&= f(\mathbf{x}_0)
\end{aligned}
$$

a contradiction.

□

Theorem: *Let C be a closed bounded convex subset of $\mathbb{R}^n$ and let $f : C \to \mathbb{R}$ be a continuous convex function. Then f attains a global maximum at an extreme point of C.*

Proof: Since C is closed bounded, and hence compact, and f is continuous by the theorem on page 21 there exists $\mathbf{x}_0 \in C$ at which f attains its maximum.

Our proof now proceeds by induction on the dimension of C. Clearly the claim is true if $\dim(C) = 0$ as then $C = \{\mathbf{x}_0\}$.

Assume the result is true for all continuous convex functions defined on closed bounded convex sets of dimension $n-1$ or less, and suppose C has dimension n.

If $\mathbf{x}_0$ is an interior point of C, then f is constant on C (by the previous theorem) and so certainly f achieves its maximum (only) value at an extreme point of C. On the other hand, if $\mathbf{x}_0$ is a boundary point of C then there exists a hyperplane H supporting C at $\mathbf{x}_0$. Let $K = C \cap H$, then $\mathbf{x}_0 \in K$ and f restricted to K is a continuous convex function on a convex set of dimension at most $n-1$ with maximum value $f(\mathbf{x}_0)$. By the induction hypothesis, this maximum is attained at some extreme point $\mathbf{k}$ of K. But, by the previous lemma, $\mathbf{k}$ is an extreme point of C and

$$f(\mathbf{k}) = f(\mathbf{x}_0) = \text{ maximum of } f \text{ on } C \text{ .}$$

□

Application to Game Theory

Recall that in chapter 1 example 4 our problem was to find

$$\max_{\mathrm{a}} \ \min_{\mathrm{b}} \ E(\mathrm{a}, \mathrm{b}), \quad \text{where } E(\mathrm{a},\mathrm{b}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j p_{ij}$$

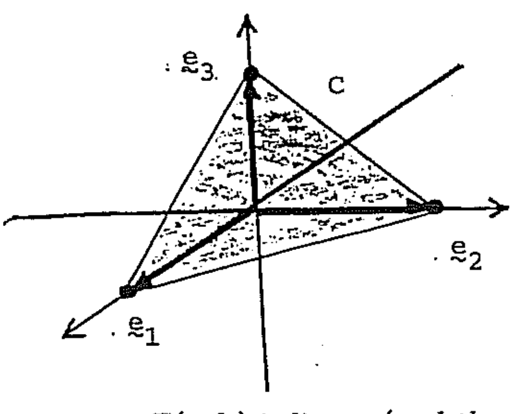
and a, b are probability densities, that is

$$\mathrm{a}, \mathrm{b} \in C := \{\mathrm{x} = (x_1, \ldots, x_n) : x_i \geq 0 \text{ and } \sum_{i=1}^{n} x_i = 1\}.$$

Now, C is a convex subset of $\mathbb{R}^n$ with the n standard unit vectors

$$\mathrm{e}_1 = (1, 0, \ldots 0), \ \mathrm{e}_2 = (0, 1, 0, \ \ldots, 0), \ \ldots, \ \mathrm{e}_n = (0, \ldots, 0, 1)$$

as its extreme points (see Exercise 17).



Also for a fixed a the function $E(\mathrm{a}, \mathrm{b})$ is linear (and therefore concave) in b. Thus by the previous theorem it attains its minimum at an extreme point of C.

That is,

$$\begin{aligned} f(\mathrm{a}) &:= \min_{\mathrm{b} \in C} E(\mathrm{a}, \mathrm{b}) \\ &= \min_{j=1,2,\ldots,n} E(\mathrm{a}, \mathrm{e_j}). \end{aligned}$$

To maximize f over C is therefore to find the largest z such that

$$z \leq E(\mathrm{a}, \mathrm{e}_1), \ \ z \leq E(\mathrm{a}, \mathrm{e}_2), \ \ldots, \ \ z \leq E(\mathrm{a}, \mathrm{e}_n)$$

for some a in C.

Since $E(\mathbf{a}, \mathbf{e}_j) = \sum_{i=1}^{n} a_i p_{ij}$ our problem is to find $a_1, a_2, \ldots, a_n$ and z which maximize $f(z) := z$ subject to the constraints

$$
\begin{aligned}
z - \sum a_i p_{i1} &= z - p_{11}a_1 - p_{21}a_2 - \ldots - p_{n1}a_n \leq 0 \\
z - \sum a_i p_{i2} &= z - p_{12}a_1 - p_{22}a_2 - \ldots - p_{n2}a_n \leq 0 \\
&\ldots \\
z - \sum a_i p_{in} &= z - p_{1n}a_1 - p_{2n}a_2 - \ldots - p_{nn}a_n \leq 0 \\
&a_1 + a_2 + \ldots a_n = 1 \\
&a_1 \geq 0 \\
&a_2 \geq 0 \\
&\ldots \\
&a_n \geq 0
\end{aligned}
$$

A problem which we recognise as a **linear programming problem.**

For example, in our game between A. Hog and B. Greedy the best mixed strategy for A is found by solving:

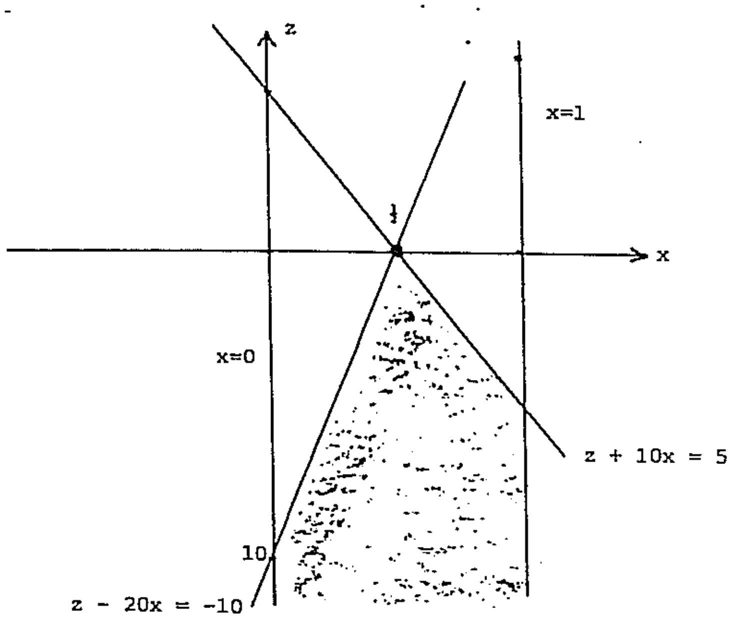
$$
\begin{aligned}
\textit{maximise: } & z \\
\textit{subject to: } & z + 5a_1 - 5a_2 \leq 0 \\
& z - 10a_1 + 10a_2 \leq 0 \\
& a_1 + a_2 = 1 \\
& a_1,\ a_2 \geq 0\ .
\end{aligned}
$$

We may solve this problem graphically as follows.

Let $x = a_1$, then $a_2 = 1 - x$ and the requirement $a_2 \geq 0$ becomes $x \leq 1$. Stated in terms of x our problem is

$$
\begin{aligned}
\textit{maximise: } & z \\
\textit{subject to: } & z + 5x - 5(1-x) \leq 0, \text{ or } z + 10x - 5 \leq 0 \\
& z - 10x + 10(1-x) \leq 0, \text{ or } z - 20x + 10 \leq 0 \\
& 0 \leq x \leq 1\ .
\end{aligned}
$$

From the graph of the constraint set depicted below



we readily see that the maximum feasible z value is 0 which occurs for $x = \frac{1}{2}$. That is, $a_1, a_2 = \frac{1}{2}$ and so A should play his 5 cent and 10 cent coin with equal frequency, in which case his expected gain (the value of z) is 0.

Exercises

(1) For $\alpha \in \mathbb{R}$ and C a convex subset of $\mathbb{R}^n$ show that αC is convex.

(2) Let $C_i,\ i \in I$ be a family of convex subsets of $\mathbb{R}^n$, show that $\bigcap_{i \in I} C_i$ is a (possibly empty) convex set.

(3) For any $\mathbf{x}_0 \in \mathbb{R}^n$ and subspace M of $\mathbb{R}^n$ show that $\mathbf{x}_0 + M$ is an affine subset of $\mathbb{R}^n$.

(4) Let A be an affine set and suppose that $A = \mathbf{x}_1 + M_1$ and $A = \mathbf{x}_2 + M_2$ where M_1 and M_2 are subspaces. Show that $M_1 = M_2$.

(5) Let C be a convex subset of $\mathbb{R}^n$. If $\mathbf{x}_1, \cdots, \mathbf{x}_m \in C$ and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are positive numbers with $\sum_{i=1}^{m} \lambda_i = 1$, show that $\sum_{i=1}^{m} \lambda_i \mathbf{x}_i \in C$. [Hint; use induction on m, by the definition of a convex set, it is true when $m = 2$.].

(6) If $A \subseteq B \subseteq \mathbb{R}^n$ show that $\text{co}(A) \subseteq \text{co}(B)$.

*(7) (i) For $A \subseteq \mathbb{R}^n$ show that $co(A)$ is the set consisting of all vectors of the form

$$\sum_{i=1}^{m} \lambda_i \mathbf{a}_i$$

where $m \in \mathbb{N}$; $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \in A$, and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are positive numbers with $\sum_{i=1}^{m} \lambda_i = 1$.

(ii) Use (i), the last theorem of section 3.1 and Jensen's inequality to give an alternative proof for the final theorem of this chapter.

(8) Show that every closed convex subset of $\mathbb{R}^n$ is the intersection of the closed $\frac{1}{2}$ - spaces which contain it.

(9) Let D be a convex set.

(i) If $\mathbf{x}$ is an interior point of D and $\mathbf{y} \in D$, show that for $\lambda \in (0, 1]$ the point $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$ is an interior point of D.

(ii) Using (i) deduce that the set of interior points of D, $intD$, is a convex set.

and

(iii) $D \subseteq \overline{intD}$ [in fact, $\overline{intD} = \bar{D}$].

*(10) Show that the closure of a convex set is itself convex. Hence deduce that for any set S we have $\overline{co}\ S$ is the smallest closed convex set containing S.

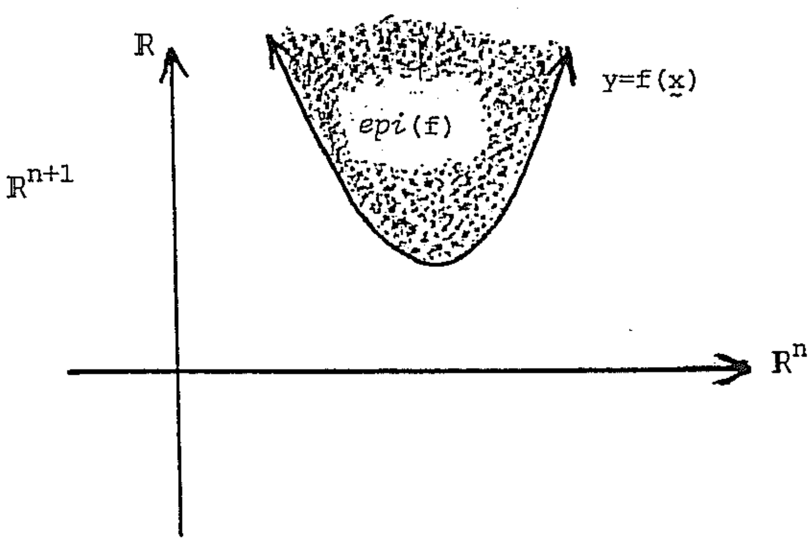
(11) Show $\mathbf{x}$ is an extreme point of $C \subseteq \mathbb{R}^n$ if and only if whenever $\frac{1}{2}(\mathbf{y}+\mathbf{z}) = \mathbf{x}$ with $\mathbf{y}, \mathbf{z} \in C$ we must have $\mathbf{y} = \mathbf{z}$ $(= \mathbf{x})$.

*(12) Show that every convex polyhedron has only a finite number of extreme points.

(13) Let C be a convex subset of $\mathbb{R}^n$. Show that $f : C \to \mathbb{R}$ is a convex function if and only if its epigraph;

$$epi(f) := \{(\mathbf{x}, y) \in C \times \mathbb{R} : f(\mathbf{x}) \leq y\},$$

is a convex subset of $\mathbb{R}^{n+1}$



(14) Let C be a convex subset of $\mathbb{R}^n$ and let f and g be convex functions from C into $\mathbb{R}$. Show that:

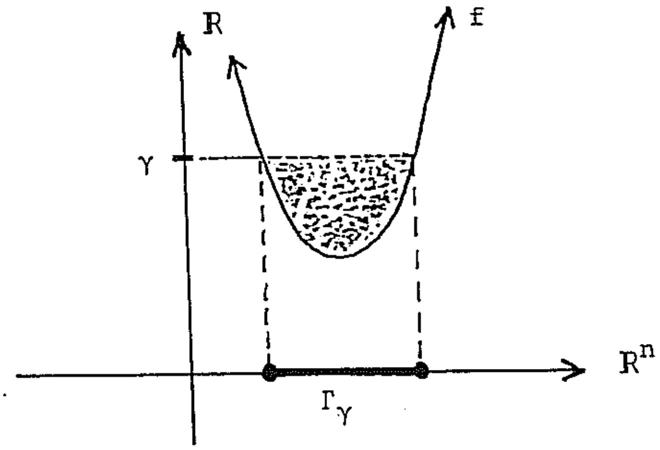
(i) αf is a convex function whenever $\alpha \geq 0$;

(ii) $f + g$ is a convex function;

(iii) the maximum of f and g; $(f \vee g)(\mathbf{x}) :=$ maximum $\{f(\mathbf{x}), g(\mathbf{x})\}$ is a convex function.

(15) Let C be a convex subset of $\mathbb{R}^n$ and let $f : C \to \mathbb{R}$ be a convex function. Show that

(i) For any $\gamma \in \mathbb{R}$ the (possibly empty) set $\Gamma_\gamma := \{\mathrm{x} \in C : f(\mathrm{x}) \leq \gamma\}$ is a convex subset of C.



(ii) The set of points in C at which f attains its minimum is a (possibly empty) convex subset of C.

(iii) If f is strictly convex, the set described in (ii) contains at most one point.

(16) An **affine** function $A : \mathbb{R}^n \to \mathbb{R}$ has the form $A(\mathrm{x}) = T\mathrm{x} + b$ where T is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$ and b is a fixed number.

(i) Show that $A(\mathrm{x}) = \mathrm{a} \cdot \mathrm{x} + b$ for some fixed vector $\mathrm{a} \in \mathbb{R}^n$. [Hence the affine functions from $\mathbb{R}$ to $\mathbb{R}$ are precisely those functions whose graph is a straight line.]

(ii) Show that A is both convex and concave.

(17) Let C be the set of *probability density functions* on $\{1, 2, \ldots, n\}$; that is

$$C := \{\mathrm{p} = (p_1, p_2, \ldots, p_n) : \sum_{j=1}^{n} p_j = 1 \text{ and } p_j \geq 0 \text{ for } j = 1, 2, \ldots, n\}.$$

Show that x is an extreme point of C if and only if it equals one of the n basis vectors

$$\mathrm{e}_1 = (1, 0, \ldots, 0),\ \mathrm{e}_2 = (0, 1, 0, \ldots, 0),\ \ldots,\ \mathrm{e}_n = (0, \ldots, 0, 1).$$

(18) Find the optimal mixed strategy $\mathbf{b} = (b_1, b_2)$ for player B in our game with pay-off matrix for A's winnings

A \ B	5	10
5	-5	10
10	5	-10

*(19) Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function. Prove that f is (strictly) convex if and only if the second derivative of f is a (strictly) positive function.

(20) (i) Prove Jensen's inequality.

(ii) [*General Arithmetic - Geometric Mean inequality*]. For $i = 1, 2, \ldots, n$ let x_i and α_i be strictly positive real numbers with $\sum_{i=1}^{n} \alpha_i = 1$. Prove that

$$x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \leq \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n.$$

[Hint: Write $x_i^{\alpha_i}$ as $\exp(\alpha_i \ell n x_i)$ and note that $\exp(x)$ is a convex function on $\mathbb{R}$.]

CHAPTER 4 - Linear Programming

In this chapter we explore linear programming problems and the simplex algorithm for their solution. Linear programming problems are undoubtedly the most frequently encountered type of optimisation problem, and the simplex algorithm is one of the best known and most commonly applied algorithms in the field.

We have seen several examples of a linear programming problem; Chapter 1 examples 1(a) and (b), and also example 4 (see the end of Chapter 3).

The general form of such a problem is:

$$\begin{aligned}
\textit{minimize}: \quad & f(\mathbf{x}) = \mathbf{c}\cdot\mathbf{x} = c_1x_1 + c_2x_2 + \ldots + c_nx_n, \\
\textit{subject to}: \quad & \mathbf{a}_1\cdot\mathbf{x} = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\
& \ldots \\
& \mathbf{a}_k\cdot\mathbf{x} = a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n \le b_k \\
& \ldots \\
& \mathbf{a}_{k+l}\cdot\mathbf{x} = a_{k+l\ 1}x_1 + \cdots + a_{k+l\ n}x_n \ge b_{k+l} \\
& \ldots \\
& \mathbf{a}_m\cdot\mathbf{x} = a_{m1}x_1 + \cdots + a_{mn}x_n \ge b_m \\
& x_i \ge 0 \text{ for } \ i \in I \subseteq \{\ 1, 2, \cdots, n\}.
\end{aligned}$$

The first m constraints are a mixture of equality and inequality constraints. The last set of constraints require that certain variables be positive and are consequently known as *positivity constraints.* While these are just a special type of inequality constraint it will prove convenient to distinguish them from the others.

Basic observation: The set of $\mathbf{x}$'s satisfying an equality constraint $\mathbf{a}_i\cdot\mathbf{x} = b_i$ is a hyperplane in $\mathbb{R}^n$. The $\mathbf{x}$'s satisfying an inequality constraint $\mathbf{a}_j\cdot\mathbf{x} \underset{(\ge)}{\le} b_j$, including a positivity constraint (of the form $\mathbf{e}_k\cdot\mathbf{x} \ge 0$), constitute a closed $\frac{1}{2}$ - space.

Hence the constraint set C (the set of feasible solutions, or those $\mathbf{x}$'s satisfying all the constraints) is an intersection of hyperplanes and closed $\frac{1}{2}$-spaces and so forms a closed convex set, indeed a polytope.

Thus if f achieves a finite minimum on C it is achieved at an extreme point of C and there are only finitely many of these.

Example Our diet problem, chapter 1 example 1:

$$\begin{aligned} \textit{minimize:} \quad & 50x_1 + 10x_2 \\ \textit{subject to:} \quad & x_1 + x_2 \leq 3 \\ & x_1 + 5x_2 \leq 10 \\ & 2x_1 + x_2 \geq 1 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

has the constraint set whose extreme points are tabulated below (see Figure on Page 4).

extreme points	value of f	
$(0,2)$	20	
$(0,1)$	10	$\leftarrow$ minimum value
$(3,0)$	150	
$(5/4,7/4)$	80	
$(1/2,0)$	25	

So the minimum of the objective function is 10 which is achieved at $x_1 = 0, x_2 = 1$. That is, make the chilli mix 100 % beans, no meat.

- taste familiar!

From this point our development must necessarily be somewhat recipe-like, however all the details are presented so that the rationale for what we do can be discerned. I hope that this will indeed prove to be the case.

Our general **plan of attack** is:

1. Develop effective methods for locating the extreme points of C.

2. Develop an efficient algorithm for searching the extreme points to determine one at which f is minimal.

In order to embrace all varieties of linear programming problems in a single approach it is first necessary to replace each problem by an equivalent one (i.e. a problem with the same constraint set and optimal solution as the original problem) which has a cannonical form that can be assumed throughout our subsequent treatment.

Reduction to Standard Form

Given a linear programming problem, if it is a maximisation problem replace the objective function $f(\mathbf{x})$ by $-f(\mathbf{x})$ to obtain a minimisation problem. Then, starting with the given linear programming problem in the form

$$\begin{aligned} \textit{minimize}: \quad & f(\mathbf{x}) = \mathbf{c} \cdot \mathbf{x}, \\ \textit{subject to}: \; & \mathbf{a}_1 \cdot \mathbf{x} = b_1 \\ & \dots \\ & \mathbf{a}_k \cdot \mathbf{x} \leq b_k \\ & \dots \\ & \mathbf{a}_{k+l} \cdot \mathbf{x} \geq b_{k+l} \\ & \dots \\ & x_i \geq 0 \text{ for } \; i \in I \end{aligned}$$

we perform the following series of steps. After describing each step in a general way it will be illustrated by performing it on the following example.

Example:

$$\begin{aligned} \textit{minimize:} \quad & x_1 + x_2 + x_3 \\ \textit{subject to:} \quad & x_1 + 2x_2 = 1 \\ & x_1 - x_2 + x_3 \geq 2 \\ & x_1 + x_3 \leq 5 \\ & x_2 \leq 0 \\ & \quad \text{and the positivity constraint :} \\ & x_1 \geq 0 \end{aligned}$$

Step 1 Replace each equality constraint $\mathbf{a}_j \cdot \mathbf{x} = b_j$ by the equivalent pair of inequalities

$$\mathbf{a}_j \cdot \mathbf{x} \leq b_j$$

$$\mathbf{a}_j \cdot \mathbf{x} \geq b_j$$

Applying Step 1 to the example we have

$$\begin{aligned}
\textit{minimize:} \quad & x_1 + x_2 + x_3 \\
\textit{subject to:} \quad & x_1 + 2x_2 \leq 1 \\
& x_1 + 2x_2 \geq 1 \\
& x_1 - x_2 + x_3 \geq 2 \\
& x_1 + x_3 \leq 5 \\
& x_2 \leq 0 \\
& x_1 \geq 0
\end{aligned}$$

Step 2 Rewrite each inequality constraint of the form $\mathbf{a}_j \cdot \mathbf{x} \geq b_j$ (including those which arose from Step 1, but excluding the positivity constraints) by

$$(-\mathbf{a}_j) \cdot \mathbf{x} \leq (-b_j)$$

At this stage, apart from positivity constraints all constraints, m in number say, will be of the form $\mathbf{a}_j \cdot \mathbf{x} \leq b_j$.

Applying Step 2 to our example yields

$$\begin{aligned}
\textit{minimize:} \quad & x_1 + x_2 + x_3 \\
\textit{subject to:} \quad & x_1 + 2x_2 \leq 1 \\
& -x_1 + -2x_2 \leq -1 \\
& -x_1 + x_2 - x_3 \leq -2 \\
& x_1 + x_3 \leq 5 \\
& x_2 \leq 0 \\
& x_1 \geq 0
\end{aligned}$$

Step 3 Introduce m new variables, $x_{n+1}, \ x_{n+2}, \ \ldots, \ x_{n+m}$, one for each constraint, referred to as **slack variables**, using them to replace each of the constraints $\mathbf{a}_j \cdot \mathbf{x} \leq b_j$ $(j = 1, 2, \cdots, m)$ by the equivalent requirement

$$\begin{cases} \mathbf{a}_j \cdot \mathbf{x} + x_{n+j} = b_j, \\ x_{n+j} \geq 0 & \text{(that is, a positivity constraint for } x_{n+j}\text{).} \end{cases}$$

REMARK: Slack variables for inequalities which were originally $\geq$ before Step 2 are sometimes called **surplus variables**. The distinction between surplus and slack variables is important for certain economic applications, but will not concern us here.

After Step 3 our example becomes

$$\begin{aligned} \textit{minimize:} \quad & x_1 + x_2 + x_3 \\ \textit{subject to:} \quad & x_1 + 2x_2 + \underline{x_4} = 1 \\ & -x_1 - 2x_2 + \underline{x_5} = -1 \\ & -x_1 - x_2 - x_3 + \underline{x_6} = -2 \quad \text{(underlined variables are slack variables)} \\ & x_1 + x_3 + \underline{x_7} = 5 \\ & x_2 + \underline{x_8} = 0 \\ & x_1,\ \underline{x_4},\ \underline{x_5},\ \underline{x_6},\ \underline{x_7},\ \underline{x_8} \geq 0. \end{aligned}$$

At this stage, excluding the positivity constraints, each constraint is an equality constraint involving a slack variable. Also, with the possible exception of some of the original variables, all variables satisfy a positivity constraint.

Step 4 To ensure every variable satisfies a positivity constraint we may adopt either of the following methods.

(i) Since every real number can be written as the difference of two positive ones, if x_j $(j \in \{1, 2, \cdots, n\})$ fails to satisfy a positivity constraint express it as the difference of two new variables

$$x_j = u_j - v_j$$

and require each of the new variables u_j and v_j to be positive.

(ii) If x_j $(j \in \{1, 2, \cdots,\ n,\})$ fails to satisfy a positivity constraint select one of the equality constraints in which x_j has a non-zero coefficient, solve this for x_j and use it to eliminate x_j from the problem.

Relabelling variables, our problem is now reduced to **standard form:**

Minimize: $\mathbf{c} \cdot \mathbf{x}$

Subject to: $A\mathbf{x} = \mathbf{b}$

$\mathbf{x} \geq 0$

where $\mathbf{x} = (x_1, x_2 \ldots, x_n, \underbrace{x_{n+1}, \ldots, x_{n+m}}_{\text{slack variables}})$ and A is an $m \times (n+m)$ matrix of the form.

$$A = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n\ \underbrace{\mathbf{a}_{n+1}\ \cdots\ \mathbf{a}_{n+m}}_{\text{identity matrix}}]$$

with the (column) vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{n+m}$ denoting the $m+n$ columns of A.

Using the method of Step 4 (i) for our example we arrive at the following standard form for the problem.

$$\begin{aligned}
\textit{Minimize:}\quad & x_1 + u_2 - v_2 + u_3 - v_3 \\
\textit{subject to:}\quad & x_1 + 2u_2 - 2v_2 + x_4 = 1 \\
& -x_1 - 2u_2 + 2v_2 + x_5 = -1 \\
& -x_1 + u_2 - v_2 - u_3 + v_3 + x_6 = -2 \\
& x_1 + u_3 - v_3 + x_7 = 5 \\
& u_2 - v_2, +x_8 = 0 \\
& x_1,\ u_2,\ v_2\, u_3,\ v_3,\ x_4,\ x_5,\ x_6,\ x_7,\ x_8 \geq 0.
\end{aligned}$$

Alternatively if we use Step 4 (ii) we obtain as a standard form the following problem.

$$\begin{aligned}
\textit{Minimize:}\quad & 5 - x_7 - x_8 \\
\textit{subject to:}\quad & x_1 - 2x_8 + x_4 = 1 \quad (\textit{ using } x_2 = -x_8 \text{ and } x_3 = 5 - x_1 - x_7) \\
& -x_1 + 2x_8 + x_5 = -1 \\
& -x_8 + x_7 + x_6 = 3 \\
& x_1,\ x_4,\ x_5,\ x_6,\ x_7,\ x_8 \geq 0.
\end{aligned}$$

Note x_7 and x_8 should no longer be regarded as slack variables for this problem.

As noted in Step 4 it is usual to relabel the variables and express the standard form of the problem in terms of the generic variables $\mathbf{x} = (x_1, x_2, \cdots, x_{n+m})$. Thus if we used Step 4(ii) and relabelled the variables as follows

$$\begin{aligned}
& x_1 \longrightarrow x_1 \\
& x_2 \quad \text{deleted} \\
& x_3 \quad \text{deleted} \\
& x_4 \longrightarrow x_4 \\
& x_5 \longrightarrow x_5 \\
& x_6 \longrightarrow x_6 \\
& x_7 \longrightarrow x_3 \\
& x_8 \longrightarrow x_2.
\end{aligned}$$

we would obtain as a standard form the equivalent problem:

$$\begin{aligned} \textit{Minimize:} \quad & -x_2 - x_3 \\ \textit{subject to:} \quad & x_1 - 2x_2 + x_4 = 1 \\ & -x_1 + 2x_2 + x_5 = -1 \\ & x_2 + x_{3x_6} = 3 \\ & x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6 \geq 0. \end{aligned}$$

That is,

$$\begin{aligned} \textit{Minimize:} \quad & (0, -1, -1, 0, 0, 0) \cdot \mathbf{x} \\ \textit{subject to:} \quad & \begin{bmatrix} 1 & -2 & 0 & 1 & 0 & 0 \\ -1 & 2 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \\ & \mathbf{x} \geq 0 \,. \end{aligned}$$

Remarks

1. Our reason for putting all our Linear Programming problems into this particular standard form is, as we shall see, that from this form it is relatively easy to determine extreme points of the constraint set.
2. The standard form of a Linear Programming problem involves a considerable amount of redundant information. The coefficients of initial equality constraints are duplicated, the number of variables increased, particularly when positivity constraints are absent and method (i) is adopted. Since Linear Programming problems are often large this redundance should be avoided when "storing" the problem in a computer. Although we will not pursue such matters, many algorithms have been devised which, while "working" on the standard form, avoid redundant storage. These methods usually result from a careful analysis of how the data is used by the basic algorithms we will develop.

As a further example consider our **diet problem** from Chapter 1:

$$\begin{aligned} \textit{minimize}: \quad & 50x_1 + 10x_2 \\ \textit{subject to:} \quad & x_1 + x_2 \leq 3 \\ & x_1 + 5x_2 \leq 10 \\ & 2x_1 + x_2 \geq 1 \\ & x_1 \geq 0, x_2 \geq 0. \end{aligned}$$

The standard form for this problem is:

$$\textit{minimize:} \quad 50x_1 + 10x_2 = (50, 10, 0, 0, 0) \cdot \mathbf{x}$$

$$\textit{subject to:} \quad \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 5 & 0 & 1 & 0 \\ -2 & -1 & 0 & 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 10 \\ -1 \end{bmatrix}$$

$$\mathbf{x} \geq 0$$

Basic Solutions

Consider a linear programming problem reduced to a standard form:

Minimize: $\mathbf{c} \cdot \mathbf{x}$

subject to: $A\mathbf{x} = \mathbf{b}$

$\mathbf{x} \geq 0$

where, A is the $m \times (n+m)$ matrix

$$A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n \ \underbrace{\mathbf{a}_{n+1} \ \cdots \ \mathbf{a}_{n+m}}_{\text{identity matrix}}]$$

In the arguments which follow it is important to recall that

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_{n+m}\mathbf{a}_{n+m}.$$

Let $\mathbf{a}_{j_1}, \mathbf{a}_{j_2}, \ldots, \mathbf{a}_{j_m}$ be any m linearly independent columns of A. At least one such selection exists, as the last m columns are the unit vector basis

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ \cdot \\ \cdot \\ cr \end{bmatrix}, \cdots \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

and there are at most $\binom{n+m}{m} = \dfrac{(n+m)(n+m-1)\ldots(n+1)}{m(m-1)\ldots 2.1}$ such selections.

The (unique) vector $\mathbf{x}$ with $\mathrm{x}_j = 0, j \neq j_1, j_2, \ldots, j_m$ and

$$[\mathbf{a}_{j_1} \quad \mathbf{a}_{j_2} \quad \cdots \quad \mathbf{a}_{j_m}] \begin{bmatrix} x_{j_1} \\ x_{j_2} \\ \cdot \\ \cdot \\ \cdot \\ x_{j_m} \end{bmatrix} = \mathbf{b}$$

is clearly a solution of $Ax = \mathbf{b}$. We call it a **basic solution.** "Basic" because the non-zero entries $x_{j_1}, x_{j_2}, \ldots, x_{j_m}$ are the (unique) "coordinates" for the expansion of $\mathbf{b}$ in terms of the *basis* $\mathbf{a}_{j_1}, \mathbf{a}_{j_2}, \ldots, \mathbf{a}_{j_m}$ of $\mathbb{R}^m$.

If in addition $x_{j_1}, x_{j_2} \cdots x_{j_n} \geq 0$ (that is, $\mathbf{x} \geq 0$) we call $\mathbf{x}$ a **basic feasible solution,** it is basic and feasible as it satisfies all the constraints, including the positivity constraints.

EXAMPLE: For the standard form of our diet problem:

Minimize: $50x_1 + 10x_2$

$$\text{subject to:} \quad \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 5 & 0 & 1 & 0 \\ -2 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 \\ 10 \\ -1 \end{bmatrix}$$

$$x_1, x_2, x_3, x_4, x_5 \geq 0$$

we have the following basic solutions.

Selection of Columns	*Basic solution*	
1 2 3	(-5/9, 19/9, 13/9, 0,0)	
1 2 4	(-2,5, 0, -13, 0)	
1 2 5	(5/4, 7/4, 0, 0, 2)	*a basic feasible solution*
1 3 4	(1/2, 0, 5/2, 19/2, 0)	*a basic feasible solution*
1 3 5	(10, 0, -7, 0, 19)	
1 4 5	(3, 0, 0, 7, 5)	*a basic feasible solution*
2 3 4	(0, 1, 2, 5, 0)	*a basic feasible solution*
2 3 5	(0, 2, 1, 0, 1)	*a basic feasible solution*
2 4 5	(0, 3, 0, -5, 2)	
3 4 5	(0, 0, 3, 10, -1)	

Here, for example the basic solution corresponding to the selection 1 2 3 is obtained by setting $x_4 = x_5 = 0$ and solving

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 5 & 0 \\ -2 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 10 \\ -1 \end{bmatrix}.$$

The other basic solutions are obtained similarly.

Note that if $(x_1, x_2, x_3, x_4, x_5)$ satisfies the constraints of our standard form then (x_1, x_2) satisfies the original constraints (x_3, x_4, x_5 were the additional slack variables introduced), so from the basic feasible solutions obtained above we have the following feasible solutions for our original problem

$$(5/4,\ 7/4)$$
$$(1/2,\ 0)$$
$$(3,\ 0)$$
$$(0,\ 1)$$
$$(0,\ 2)$$

A comparison with the results of our previous graphical analysis shows that these are precisely the extreme points of the constraint set! This is no coincidence, as the following theorem shows.

Fundamental Theorem of Linear Programming: *For the linear programming problem in standard form*

$$\begin{aligned} \textit{minimize:} \quad & \mathbf{c} \cdot \mathbf{x} \\ \textit{subject to:} \quad & \underbrace{[\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n\ I_{m\times m}]}_{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq 0 \end{aligned}$$

the vector $\mathbf{x}$ *is an extreme point of the constraint set* C *if and only if* $\mathbf{x}$ *is a basic feasible solution for some selection of* m *linearly independent columns of* A.

Proof: ($\Leftarrow$) Let $\mathbf{x} = (0, \ldots, x_{j_1}, \ldots, x_{j_m}, 0, \ldots, 0) \geq 0$ be a basic feasible solution (corresponding to the choice of linearly independent columns $\mathbf{a}_{j_1}, \mathbf{a}_{j_2}, \ldots \mathbf{a}_{j_m}$). Assume $\mathbf{x} = \frac{1}{2}(\mathbf{y} + \mathbf{z})$ for some $\mathbf{y}, \mathbf{z} \in C$. Since $\mathbf{y}$, $\mathbf{z} \geq 0$, both must have zero components in the same places as $\mathbf{x}$ does.

Thus

$$\begin{aligned} \mathbf{b} = A\mathbf{y} &= y_{j_1}\mathbf{a}_{j_1} + y_{j_2}\mathbf{a}_{j_2} + \ldots + y_{j_m}\mathbf{a}_{j_m} \\ &= A\mathbf{z} = z_{j_1}\mathbf{a}_{j_1} + z_{j_2}\mathbf{a}_{j_2} + \ldots + z_{j_m}\mathbf{a}_{j_m}, \end{aligned}$$

and so, since $\mathbf{a}_{j_1}, \ldots \mathbf{a}_{j_m}$ are linearly independent, $y_{j_1} = z_{j_1}, y_{j_2} = z_{j_2}, \ldots, y_{j_m} = z_{j_m}$.

That is, $\mathbf{y} = \mathbf{z}$ and $\mathbf{x}$ is an extreme point of C.

($\Rightarrow$) Let $\mathbf{x} = (0, \ldots, 0, x_{j_1}, 0 \ldots x_{j_k}, 0 \ldots)$, be an extreme point of C where $x_{j_1}, x_{j_2}, \ldots, x_{j_k}$ are the only non-zero (necessarily positive) components of $\mathbf{x}$. Since $A\mathbf{x} = \mathbf{b}$ we have

$$x_{j_1}\mathbf{a}_{j_1} + x_{j_2}\mathbf{a}_{j_2} + \ldots + x_{j_k}\mathbf{a}_{j_k} = \mathbf{b}$$

Claim: $\mathbf{a}_{j_1}, \mathbf{a}_{j_2}, \ldots, \mathbf{a}_{j_k}$ are linearly independent.

Once this claim is established we have, $k \leq$ rank $A =$ row rank$A \leq m$, so $\mathbf{x}$ is the basic feasible solution corresponding to any choice of m linearly independent columns of A which includes the k columns $\mathbf{a}_{j_1}, \ldots \mathbf{a}_{j_k}$, thereby establishing the required result.

To verify the claim suppose $y_{j_1}\mathbf{a}_{j_1} + \ldots + y_{j_k}\mathbf{a}_{j_k} = 0$ (and put $y_j = 0$ when $j \neq j_k$ for any k).

Choose $\alpha > 0$ so that

$$\min_k \; x_{j_k} \geq \alpha \; \max_k \; | y_{j_k} |$$

(Here we have assumed $\mathbf{x} \neq \mathbf{0}$. If $\mathbf{x} = \mathbf{0}$ is an (extreme) point of C then $\mathbf{b} = A\mathbf{x} = \mathbf{0}$ and so $\mathbf{0}$ is clearly a basic feasible solution.)

Let $\mathbf{u} = \mathbf{x} + \alpha\mathbf{y}$ and $\mathbf{v} = \mathbf{x} - \alpha\mathbf{y}$, then $\mathbf{u}, \mathbf{v} \geq 0$ (by the choice of α) and $A\mathbf{u} = A\mathbf{x} + \alpha A\mathbf{y} = \mathbf{b} + \mathbf{0} = \mathbf{b}$ similarly $A\mathbf{v} = \mathbf{b}$.

Thus $\mathbf{u}, \mathbf{v} \in \mathbf{C}$. But $\mathbf{x} = \frac{1}{2}(\mathbf{u} + \mathbf{v})$, so $\mathbf{u} = \mathbf{v}$ and hence $\mathbf{y} = \mathbf{0}$. In particular

$$y_{j_1} = y_{j_2} = \ldots = y_{j_k} = 0$$

and so $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_k}$ are linearly independent.

□

We are now ready to develop the simplex method for solving linear programming problems. This method devised by G.B. Dantzig in 1947 while working on military logistic problems was first published by him in 1951.

The essential idea is to move from one extreme point (basic feasible solution) to another in a systematic, efficient and computationally expedient way until we arrive at the solution.

For computational convenience we first introduce a convenient representation of the problem which we will work with from now on.

Tableau Form

The linear programming problem in standard form

minimize: $\mathbf{c} \cdot \mathbf{x} - d$

(it is convenient for the algorithm we will develop to allow an affine cost function.)

subject to: $A\mathbf{x} = \mathbf{b}$

$\mathbf{x} \geq \mathbf{0}$,

$$\text{where} \quad A = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & \\ \cdots & & & I_m \\ a_{m1} & \cdots & a_{mn} & \end{array} \right]$$

is conveniently represented in the following *tableau* form.

x_1	x_2	$\cdots$	x_n	-1	
a_{11}	a_{12}	$\cdots$	a_{1n}	b_1	$-x_{n+1}$
a_{21}	a_{22}	$\cdots$	a_{2n}	b_2	$-x_{n+2}$
$\cdots$		$\cdots$		$\cdots$	
a_{m1}	a_{m2}	$\cdots$	a_{mn}	b_m	$-x_{n+m}$
c_1	c_2	$\cdots$	c_n	d	f

Note: (1) The body of the tableau corresponds to rewriting the constraints represented by $A\mathbf{x} = \mathbf{b}$, that is

$$a_{11}x_1 + \cdots + a_{1n}x_n + x_{n+1} = b_1$$

$$a_{21}x_1 + \cdots + a_{2n}x_n + x_{n+2} = b_2$$

$$\cdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n + x_{m+n} = b_m,$$

in the form

$$a_{11}x_1 + \cdots + a_{1n}x_n - b_1 = -x_{n+1}$$

$$a_{21}x_1 + \cdots + a_{2n}x_n - b_2 = -x_{n+2}$$

$$\cdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n - b_m = -x_{n+m}.$$

(2) This tableau, together with the assumed positivity constrains $\mathbf{x} \geq 0$, completely summarizes the problem.

(3) The variables along the top border of a tableau are referred to as the **nonbasic variables** for the tableau while those appearing down the right hand edge of the tableau are its **basic variables.**

(4) The form of the tableau is such that the basic variables can be easily determined from a knowledge of the non-basic ones. In particular, setting all the non-basic variables to zero gives $x_{n+1} = b_1, x_{n+2} = b_2, \cdots, x_{n+m} = b_m$ as a basic solution, though not necessarily a basic feasible solution unless all the b_i are positive. We refer to this particular basic solution as the **basic solution corresponding to the tableau.**

EXAMPLE: The (initial) Tableau form of our Diet Problem is:

x_1	x_2	-1	
1	1	3	$-x_3$
1	5	10	$-x_4$
−2	−1	−1	$-x_5$
50	10	0	f

For which the corresponding (nonfeasible) basic solution is $\mathbf{x} = (0, 0, 3, 10, -1)$.

The Pivoting Operation

This operation transforms the problem (tableau) into an equivalent one with the rôle of one of the basic variables and one of the non-basic variables interchanged. The transformation being effected by the usual elementary operations of linear algebra.

Before describing the operation generally consider the example of *"pivoting"* about the 5 in the second row and second column of the above tableau. We aim to arrive at an equivalent tableau with the rôle of x_2 and x_4 interchanged. This may be achieved via the following steps. Divide the x_4-row (pivot row) by the coefficient of x_2 to obtain an equation in which x_2 has unit coefficient.

x_1	x_2	-1	
1	1	3	$-x_3$
1/5	1	2	$-x_{4/5}$
−2	−1	−1	$-x_5$
50	10	0	f

Now, subtract appropriate multiples of the new x_4-row to eliminate x_2 from each of the other rows including the bottom row reprenting the objective function.

x_1	x_2	-1	
$4/5$	0	1	$-x_3 + x_{4/5}$
$1/5$	1	2	$-x_{4/5}$
$-9/5$	0	1	$-x_5 - x_{4/5}$
48	0	-20	$f + 2x_4$

Now rearrange each of the equations to obtain the sought for tableau

x_1	x_4	-1	
$4/5$	$-1/5$	1	$-x_3$
$1/5$	$1/5$	2	$-x_2$
$-9/5$	$1/5$	1	$-x_5$
48	-2	-20	f

Note:

(1) By carrying the objective function along as part of the tableau we have as a result of the pivoting expressed it in terms of the new non-basic variables, x_1 and x_4.

(2) As a result of the above pivot we have obtained an equivalent problem (tableau) from which we can read off $x_1 = 0, x_2 = 2, x_3 = 1, x_4 = 0, x_5 = 1$ as a basic **feasible** (since all the b_i's of the new tableau are positive) solution (c.f. our earlier calculation of all basic feasible solutions).

In general, **pivoting about the I, J'th entry**, a_{IJ}, of a tableau:

x_1	$\cdots$	x_J	$\cdots$	x_j	$\cdots$	x_n	-1	
a_{11}		a_{1J}		a_{1j}		a_{1n}	b_1	$-x_{n+1}$
	$\cdots$		$\cdots$		$\cdots$			$\cdots$
a_{i1}		a_{iJ}		a_{ij}		a_{in}	b_i	$-x_{n+i}$
	$\cdots$		$\cdots$		$\cdots$			$\cdots$
a_{I1}		$\mathbf{a_{IJ}}$		a_{Ij}		a_{In}	b_I	$-x_{n+I}$
	$\cdots$		$\cdots$		$\cdots$			$\cdots$
a_{m1}		a_{mJ}		a_{mj}		a_{mn}	b_m	$-x_{n+m}$
c_1	$\cdots$	c_J	$\cdots$	c_j	$\cdots$	c_n	d	f

the steps illustrated above will produce the new tableau:

x_{n+I}	x_j	-1	
$-\frac{a_{iJ}}{a_{IJ}}$	$a_{ij} - \frac{a_{iJ}a_{Ij}}{a_{IJ}}$	$b_i - \frac{a_{iJ}b_I}{a_{IJ}}$	$-x_{n+i}$
$\frac{1}{a_{IJ}}$	$\frac{a_{Ij}}{a_{IJ}}$	$\frac{b_I}{a_{IJ}}$	$-x_J$
$-\frac{c_J}{a_{IJ}}$	$c_j - \frac{c_J a_{Ij}}{a_{IJ}}$	$d - \frac{c_J B_I}{a_{IJ}}$	$f(\text{x})$

That is;

(1) The entry in the pivot position is replaced by its reciprocal.

(2) Each entry in the pivot row (except the pivot itself) is divided by the pivot value a_{IJ}

(3) Each entry in the pivot column (except the pivot itself) is divided by the pivot value and multiplied by -1.

(4) From each entry off the pivot row and column subtract the product of the entry of the same column in the pivot row a_{Ij} and the entry of the same row in the pivot column a_{iJ} divided by the pivot value.

Schematically, for $i \neq I$ and $j \neq J$;

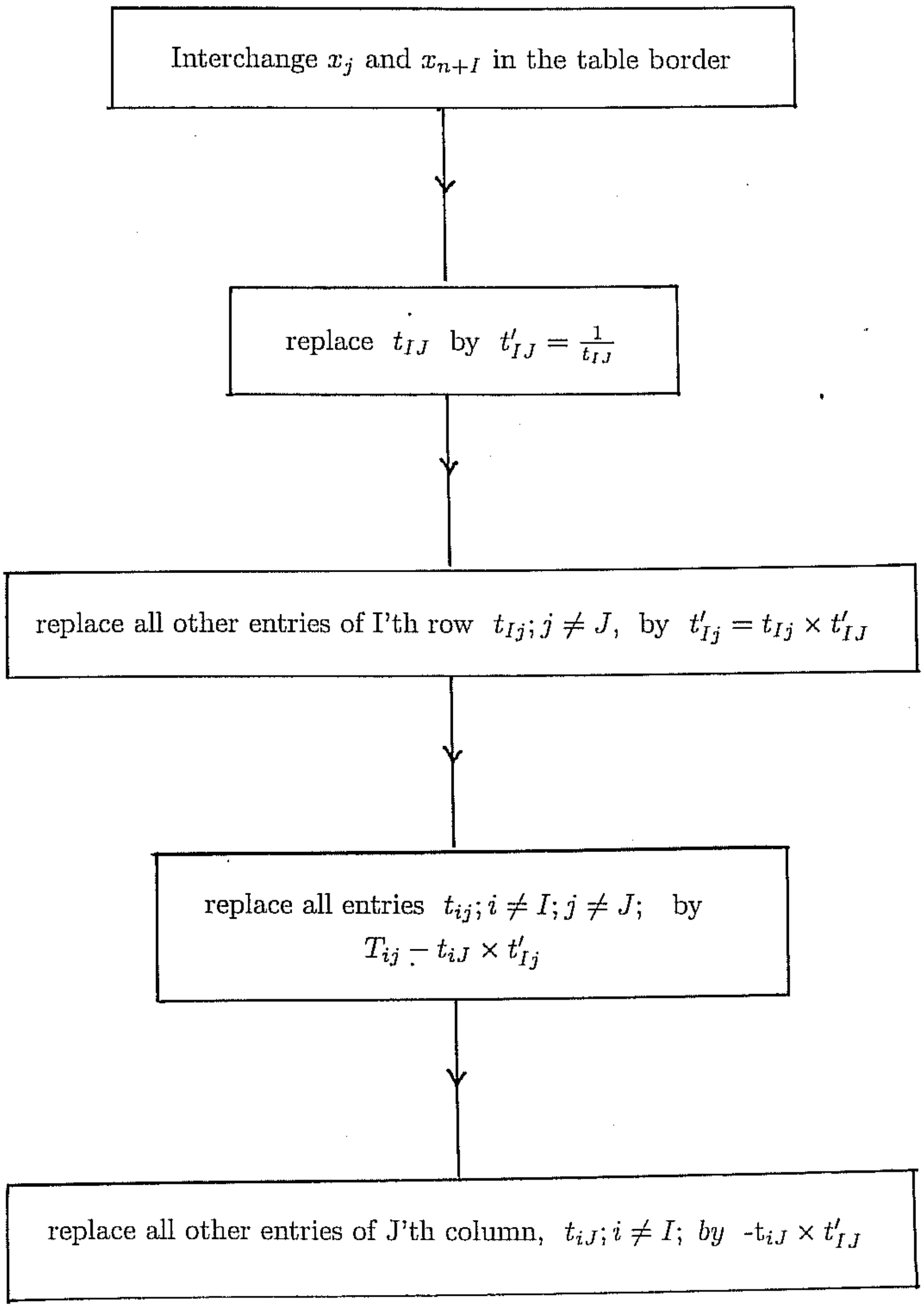
$$\begin{array}{ccc} a_{iJ} & \leftarrow & a_{ij} \\ & & \downarrow \\ a_{IJ} & & a_{Ij} \end{array}$$

the ij'th entry becomes

$$a_{ij} - \frac{a_{iJ}a_{Ij}}{a_{IJ}}.$$

In many applications the tableaus are large (sometimes both m and n are of the order of 1,000). When implementing the pivot operation by a computer, particularly in the case of such large tableaus, it is desirable to overwite the existing tableau. A convenient and efficient algorithm for effecting this is given by the following flow chart.

Labelling entries in the body of the table as t_{ij} $(i = 1, \ldots, m+1)$
we have the following algorithm for pivoting about the I, J'th position:


Interchange x_j and x_{n+I} in the table border
replace t_{IJ} by $t'_{IJ} = \frac{1}{t_{IJ}}$
replace all other entries of I'th row $t_{Ij}; j \neq J$, by $t'_{Ij} = t_{Ij} \times t'_{IJ}$
replace all entries $t_{ij}; i \neq I; j \neq J$; by
$T_{ij} - t_{iJ} \times t'_{Ij}$
replace all other entries of J'th column, $t_{iJ}; i \neq I$; *by* -$t_{iJ} \times t'_{IJ}$

FUNDAMENTAL OBSERVATIONS:

1. As already noted, if

x_1	$\cdots$	x_n	-1	
a_{11}	$\cdots$	a_{1n}	b_1	$-x_{n+1}$
$\cdots$				
a_{m1}	$\cdots$	a_{mn}	b_m	$-x_{n+m}$
c_1	$\cdots$	c_n	d	f

is a tableau for a given linear programming problem, then

$$\mathbf{x}_B = (\underbrace{0, \ \ldots, \ 0}_{n}, \ b_1, \ldots, b_m)$$

is a basic feasible solution if and only if $b_1, b_2, \ldots, b_m \geq 0$.

2. For any $\mathbf{x}$ in the constraint set we have $x_1, \ldots, x_n \geq 0$, and so $f(\mathbf{x}) = c_1x_1 + \ldots + c_nx_n - d \geq -d = f(\mathbf{x}_B)$ if $c_1, c_2, \ldots, c_n \geq 0$. It follows that

$$\mathbf{x}_B = (\underbrace{0, \ \ldots, \ 0}_{n}, \ b_1, \ldots, b_m)$$

is an optimal solution if and only if $b_1, b_2, \ldots, b_m \geq 0$ **and** $c_1, c_2, \ldots, c_n \geq 0$.

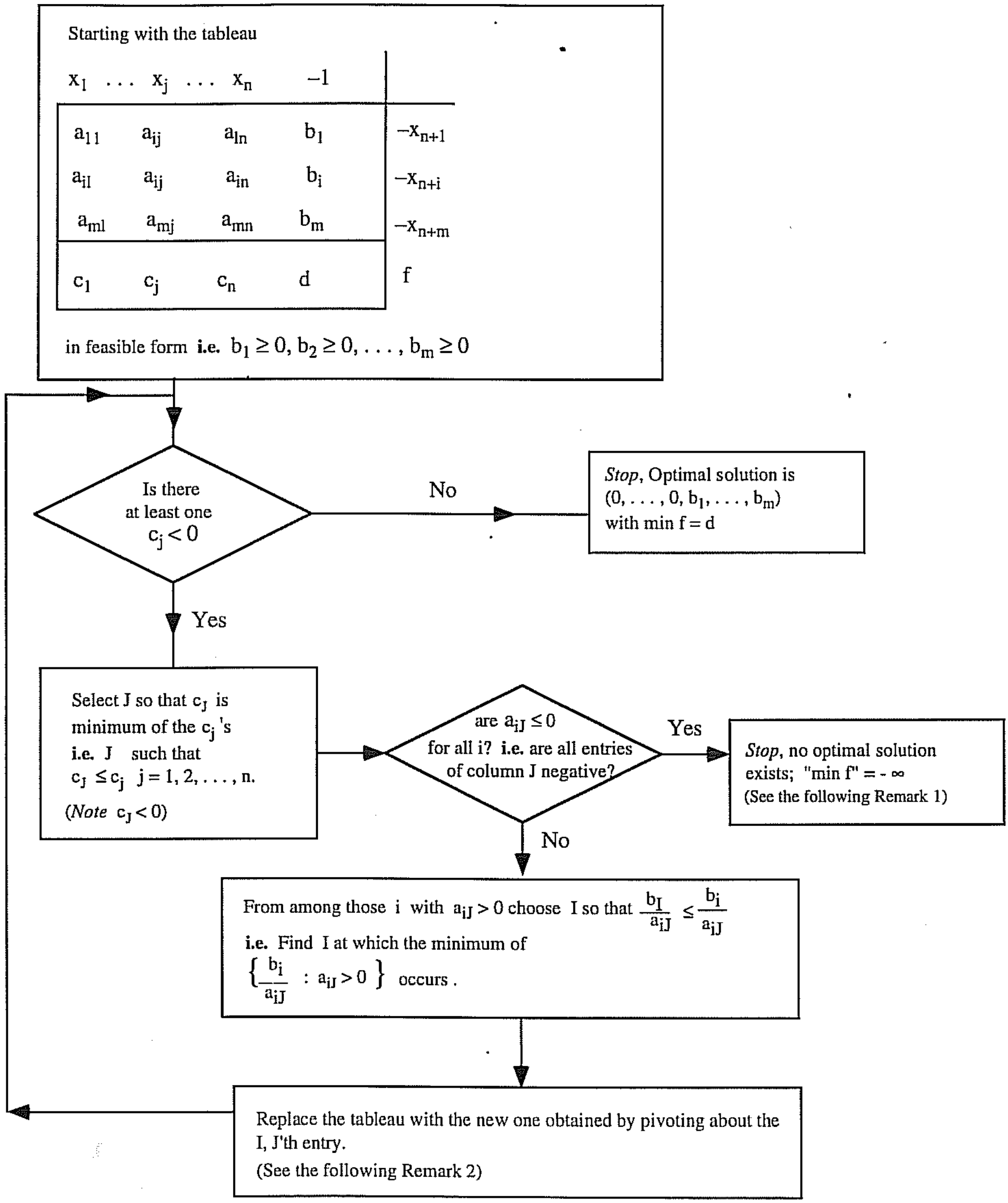
3. Further, **if** $c_j \underset{(\neq)}{>} 0$ **for** $j = 1, 2, \ldots n$ **then the vector** $\mathbf{x}_B$ **given above is the unique optimal solution.**
   This leads directly to the **strategy for the simplex method,** namely;

Stage I: By means of a suitable sequence of pivot operations transform to an equivalent tableau in which all the $b_1 (i = 1, 2, \ldots, m) \geq 0$.
(For example, the effect of pivoting about 5 in the tableau of our diet problem.) That is, locate a basic feasible solution, or equivalently an extreme point of the constraint set. A tableu with all $b_i \geq 0$ is sometimes said to be in *feasible form.*

Stage II: (if possible getting all $c_j \geq 0$). Here we attempt to pivot to a new tableau with all b_i still positive but for which $d_{new} \leq d_{old}$; that is, at the basic feasible solution corresponding to the new tableau the value of the cost function is less than or equal to its value at the previous basic feasible solution. "Geometrically" this means we move from one extreme point to another (adjacent one) at which the cost is no greater. Since there are only a finite number of extreme points, at most $\binom{n+m}{m}$, provided we continue to move to untried extreme points (that is, we don't return to a previously encountered extreme point *cycling* does not occur) we must eventually arrive at an extreme point where the cost is minimal or else find that there is no solution (the "minimum" cost is $-\infty$). As we shall see the simplex method allows us to easily recognise this last possibility should it occur.

Since many naturally occuring problems are initially in feasible form (ie all b_i's are positive) and since stage II is easier to describe (and indeed needed in Stage I) we begin by presenting the algorithm for it.

STAGE II ALGORITHM



EXAMPLE we saw that a feasible tableau, corresponding to the extreme point (0, 2, 1, 0, 1), for the *Diet problem* was:

x_1	x_4	-1	
$\frac{4}{5}$	$-\frac{1}{5}$	1	$-x_3$
$\frac{1}{5}$	$\frac{1}{5}$	2	$-x_2$
$-\frac{9}{5}$	$\frac{1}{5}$	1	$-x_5$
48	-2	-20	f

applying the algorithm leads to pivoting about the 3, 2 entry to produce:

x_1	x_5	-1	
(-1)	1	2	$-x_3$
(2)	-1	1	$-x_2$
-9	5	5	$-x_4$
30	10	-10	f

So optimal solution is $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 5, x_5 = 0$ with $\min f = 10$, see our earlier graphical solution.

Note: The tableau entries enclosed in brackets (that is, the new a_{ij} which are not in the pivot row or column) are not needed to decide whether or not the tableau is optimal, nor to find the optimal solution in the case when it is an optimal tableau. Thus calculation of these entries should be left until last and need only be made if the tableau is not optimal so a further pass through the algorithm is required.

The following remarks analyse in detail key steps in the Stage II algorith.

REMARK 1: In the case when the pivot colum (J) has been selected and the entries a_{iJ} are negative for all i, we have for any $\lambda > 0$ that the vector $\mathbf{x}_\lambda$ obtained by setting the J'th non–basic variable equal to λ and all the other non–basic variables equal to 0 and then using the tableau to solve for the remaining (basic) variables [the i'th basic variable will be $b_i - a_{iJ}\lambda$], is a feasible solution.

Further $f(\mathbf{x}_\lambda) = -d + c_J\lambda$, where $c_J < 0$, so letting $\lambda \to \infty$ we see that

$$\inf_{\mathbf{x}\in C} f(\mathbf{x}) = -\infty.$$

REMARK 2: Our choice of I ensures that the new tableau resulting from a pivot about the I, J'th entry remains feasible. To see this note that if $a_{iJ} > 0$, then by choice $\dfrac{b_I}{a_{IJ}} \leq \dfrac{b_i}{a_{iJ}}$

and so new $b_i = b_i - \dfrac{a_{iJ}b_I}{a_{IJ}} \geq 0$. On the other hand if $a_{iJ} \leq 0$ the inequality $b_i - \dfrac{a_{iJ}b_I}{a_{IJ}} \geq 0$ is ensured (as $b_I \geq 0$ and by choice $a_{IJ} \geq 0$). Hence the new tableau is feasible.

Further, since $c_J < 0$ we have that

$$d_{\text{new}} = d_{\text{old}} - \frac{c_J b_I}{a_{IJ}} \geq d_{\text{old}}.$$

That is, the value of the objective function is decreased (never increased).

REMARK 3: As each step of the algorithm, when encountered, can always be carried out we are sure that either the algorithm must terminate at one of the two *stops* (and so an optimal solution found), or since there are only a finite number of feasible tableaus (extreme points) available it must cycle back to a previously encountered tableau. We now analyse this possibility of *cycling*. We show that by a slight modification of the algorithm (which in no way affects its working) we can prevent cycling from occuring. If cycling is precluded, the algorithm must then terminate at an optimal solution. This shows that *any linear programming problem represented by a feasible tableau has a solution.*

Cycling: As already noted the possibility of the above algorithm cycling cannot be excluded, indeed this phenomenon has been shown to occur in specially constructed examples. None-the-less the situation does not seem common in applications of the simplex method to problems which arise in natural ways. Further, it can be shown that any linear programming problem may be "perturbed" in such a way that the degeneracies which lead to cycling are excluded. In practice this often means that round off errors which build up from iteration to iteration perturb the problem sufficiently to prevent indefinite cycling.

Relatively recently R.G. Bland [1977] proposed a remarkably simple modification to the Stage II algorithm presented above which prevents cycling. **Bland's modification** is now included in many implementations of the simplex algorithm. It consists of the following. **When selecting the pivot column, choose J so that $c_j < 0$ and so that the associated variable, x_k, has the smallest possible valued subscript.** (note, that for the simplex algorithm to function it was not necessary that c_J be minimal, only that c_J be strictly negative.)

In case of a tie when selecting the pivot row; that is, when there is more than one value of I for which $\frac{b_I}{a_{IJ}} \leq \frac{b_1}{a_{iJ}}$ for all i with $a_{iJ} > 0$, break the tie by choosing I so that the associated variable has the smallest possible subscript.

To see that Bland's Modification precludes cycling, we suppose that we have a tableau in feasible form for which the modified algorithm cycles and derive a contradiction.

Delete from the tableau those rows and columns which do not contain pivots occurring in the cycle to obtain a new tableau for which the algorithm still cycles and such that during the course of the cycle each variable is swapped from the top of the tableaux to the right hand side and subsequently back again (on completion of the cycle, the tableau is returned to its initial form).

For the algorithm to have cycled the cost $-d$ must have remained constant throughout the cycle (it can only decrease and must return to its initial value). Since the new cost after pivoting about the I'th row is $-d + c_J b_I / a_{IJ}$ (where $c_J < 0$), and since each row is involved in a pivot during the cycle we must have $b_I = o$ for all I.

Let N be the largest valued subscript of any variable associated with the problem, then the tableau from which c_N is swapped from the top border to the right hand side has the form:

x_N			-1	
			0	
			.	
			.	
			.	
			0	
			.	
			.	
			.	
			0	
$c_j \geq 0$	$c_J < 0$	≥ 0	d	f

For J to have been selected, all the c_j with $j \neq J$ must be positive or else a variable with smaller subscript could be swapped); while the tableau from which x_N is swapped from the right hand side to its original position in the top border has the form:

$0 \cdots$	1	0			
$x_{k1} \cdots$	x_{kJ}	$\cdots x_{k_n}$	-1		
			0	$x_{k_{n+1}}$	$-a_{ij}$
			.	.	.
			.	.	.
	$a_{iJ} \geq 0$		.	.	.
	$(i \neq I)$				
	$a_{IJ} > 0$		0	$-x_N$	$-a_{IJ}$
			.		
	≤ 0		0	$-x_{k_{n+M}}$	$-a_{mJ}$
≥ 0	$C\prime_J < 0$	≥ 0	d	f	

(Since all b_i's are zero, all rows with $a_{iJ} > 0$ are tied and so for the row associated with x_N to be chosen all other rows, being associated with variables of smaller index, must have $a_{iJ} < 0$).

It is easily seen that a solution (which is neither basic or feasible, positivity is violated) to this second tableau; that is, a solution to $A\mathbf{x} = \mathbf{b}$, is:

$$x_{k_j} = 0 (j = 1, 2, \ldots, n;\ j \neq J),\ x_{k_j} = 1$$

and $x_{k_{n+i}} = -a_{iJ}(i = 1, 2, \ldots, m)$ all of which are positive except $x_{k_{n+i}} = x_N = -a_{IJ}$ which is negative. *See outer border of the second tableau).*

From the second tableau we see that for this solution

$$f = -d + c'_J < -d \quad (\text{as } c'_J < 0)$$

On the other hand, substituting this solution into the expression for f obtained from the first tableau we have

$$f = -d - c_J a_{IJ} + \sum_{j=1}^{n} c_j z_j ,$$

where $z_j = 0, 1$ or $-a_{iJ}(i \neq I)$ and so $z_j \geq 0$, yielding $f \geq -d$.

A contradiction, since the effect of pivoting is to use the relationship $A\mathbf{x} = \mathbf{b}$ to express f in terms of a different set of arguments and so the expression for f from any tableau may be used to evaluate $f(\mathbf{x})$ for any solution $\mathbf{x}$ *of* $A\mathbf{x} = \mathbf{b}$.

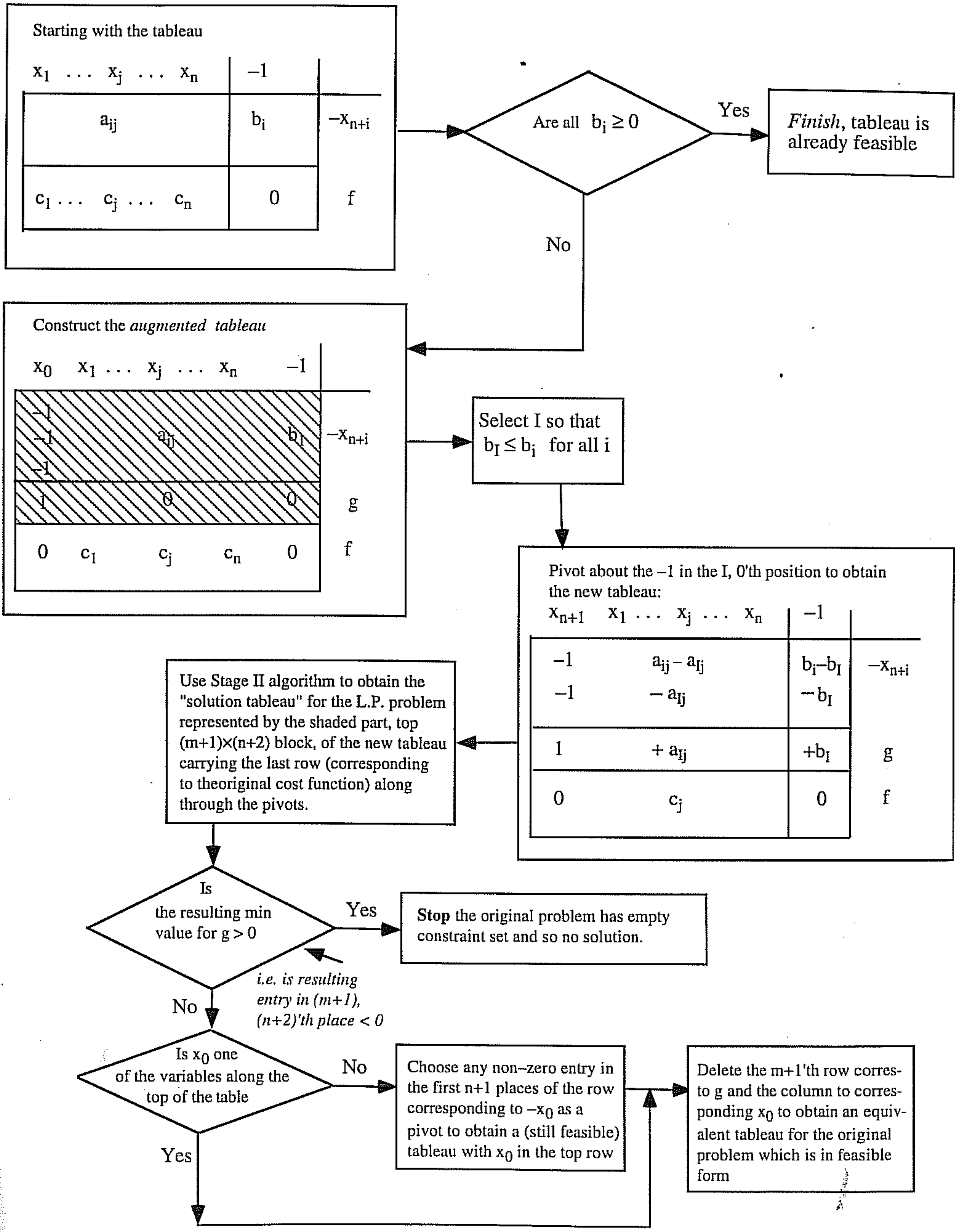
REMARK The following is an example, due to Beale (1955), of a linear programming problem which may cycle when the simplex algorithm as described on pages 64, 65 is employed.

$$\begin{aligned}
\text{minimize:} \quad & -0.75x_1 + 150x_2 - 0.02x_3 + 6x_4 \\
\text{subject to:} \quad & 0.25x_1 - 60x_2 - 0.04x_3 + 9x_4 \leq 0 \\
& 0.50x_1 - 90x_2 - 0.02x_3 + 3x_4 \leq 0 \\
& x_3 \leq 1 \\
& x_1, \quad x_2, \quad x_3, \quad x_4 \geq 0.
\end{aligned}$$

If in the standard algorithm ties are resolved by always choosing the lowest number row or column to pivot in, the example exhibits a cycle of length 6.

STAGE 1 ALGORITHM, obtaining, if possible, an initial feasible tableau - one with all $b_i \geq 0$. Geometrically: locating an extreme point of the constraint set. Our approach is that due to Rockafellar [1964].

STAGE I ALGORITHM



EXAMPLE *Diet Problem*

x_1	x_2	-1	
1	1	3	$-x_3$
1	5	10	$-x_4$
−2	−1	−1	$-x_5$
50	10	0	$f(\mathbf{x})$

augmented tableau

x_0	x_1	x_2	-1	
−1	1	1	3	$-x_3$
−1	1	5	10	$-x_4$
−1	−2	−1	−1	$-x_5$
1	0	0	0	$g(\mathbf{x})$
0	50	10	0	$f(\mathbf{x})$

x_5	x_1	x_2	-1	
−1	3	2	4	$-x_3$
−1	3	6	11	$-x_4$
−1	2	1	1	$-x_0$
1	−2	−1	−1	$g(\mathbf{x})$
0	50	10	0	$f(\mathbf{x})$

applying Stage II:

x_5	x_0	x_2	-1	
$\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	$\frac{5}{2}$	$-x_3$
$\frac{1}{2}$	$-\frac{3}{2}$	$\frac{9}{2}$	$\frac{19}{2}$	$-x_4$
$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-x_1$
0	1	0	0	$g(\mathbf{x})$
25	−25	−15	−25	$f(\mathbf{x})$

deleting row and column gives as a tableau in feasible form:

x_5	x_2	x_1	
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{5}{2}^{\bullet}$	$-x_3$
$\frac{1}{2}$	$\frac{9}{2}$	$\frac{19}{2}$	$-x_4$
$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-x_1$
25	-15	-25	$f(\mathrm{x})$

solving by Stage II.

x_5	x_1	-1	
	-1	2	$-x_3$
	-9	5	$-x_4$
-1	2	1	$-x_2$
10	30	-10	$f(\mathrm{x})$

i.e. min cost of 10 at $(0, 1, 2, 5, 0)$. See the previous solution.

DUALITY

Given a linear programming problem in **standard form** prior to introduction of slack variables, here referred to as the **Primal Problem P**;

Minimize: $f(\mathrm{x}) = \mathrm{c} \cdot \mathrm{x}$

Subject to: $A\mathrm{x} \leq \mathrm{b}$

$\mathrm{x} \geq 0$

That is;

Minimize: $c_1 x_1 + \cdots + c_n x_n$

$$\underset{(m \times n)}{\begin{bmatrix} a_{11} & a_{\ell n} \\ \cdots & \\ a_{ml} & a_{mn} \end{bmatrix}} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \leq \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix}$$

$$x_1, x_2, \ldots, x_n \geq 0$$

We may associate the **Dual Problem $\mathbf{P}^*$**;

Minimize: $f^8(\mathbf{y}) = \mathbf{b} \cdot \mathbf{y}$

Subject to: $(-A^T)\mathbf{y} \leq \mathbf{c}$

$\mathbf{y} \geq 0$

That is;

Minimize: $b_1 y_1 + \cdots + b_m y_m$

$$\underset{(n \times m)}{\begin{bmatrix} -a_{11} & \cdots & a_{ml} \\ -a_{ln} & & -a_{mn} \end{bmatrix}} \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{bmatrix} \leq \begin{bmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix}$$

$$y_1, \ldots, y_n \geq 0$$

Observation: The dual of the dual problem is

Minimize: $\mathbf{c} \cdot \mathbf{x}$

Subject to: $-(-A^T)^T \mathbf{x} \leq \mathbf{b}$ ie. $A\mathbf{x} \leq \mathbf{b}$

$\mathbf{x} \geq 0$

which we recognise as **the primal problem.**

This observation, that $(P^*)^* = P$, is one reason for calling the problems P and P^* "dual".

DUALITY THEOREM of linear programming [Gale, Klee and Rockafellar, 1951].
Let C be the constraint set of the primal problem P and let C^* be the constraint set for the dual problem P^* defined above. Then

(i) If $\mathbf{x} \in C$ and $\mathbf{y} \in C^*$ we have $f^*(\mathbf{y}) \leq -f(\mathbf{x})$

(ii) If $\mathbf{x}_0 \in C$ and $\mathbf{y}_0 \in C^*$ are such that $f^*(\mathbf{y}_0) = -f(\mathbf{x}_0)$ then $\mathbf{x}_0$ is an optimal solution for P and $\mathbf{y}_0$ is an optimal solution for P^*.

(iii) If one of the problems P or P^* has an optimal solution $\mathbf{x}_0$ or $\mathbf{y}_0$, respectively, then so does the other and $F^*(\mathbf{y}_0) = -f(\mathbf{x}_0)$.

Proof:

(i)

$$\begin{aligned} f^*(\mathbf{y}) &= \mathbf{b} \cdot \mathbf{y} \\ &\geq A\mathbf{x} \cdot \mathbf{y} \quad \text{as} \quad \mathbf{y} \geq 0 (\mathbf{y} \in C^*) \quad \text{and} \quad A\mathbf{x} \leq \mathbf{b} (\mathbf{x} \in C) \\ &= \mathbf{x} \cdot A^T \mathbf{y} \\ &\geq -\mathbf{x} \cdot \mathbf{c} \quad \text{as} \quad \mathbf{x} \geq 0 \quad \text{and} \quad \mathbf{y} \in C^* \quad \text{so} \quad -A^T\mathbf{y} \leq \mathbf{c} \quad \text{or} \quad A^T\mathbf{y} \geq -\mathbf{c} \\ &= f(x)\ . \end{aligned}$$

(ii) If $f^*(\mathrm{y}_0) = -f(\mathrm{x}_0)$; $\mathrm{y}_0)$; $\mathrm{y}_0 \in C^*, \mathrm{x}_0 \in C$, then since by (i) $f^*(\mathrm{y}) \geq -f(x_0)$ for all y we see that $\min_{\mathrm{y} \in C^*} f^*(\mathrm{y} = -f(\mathrm{x}_0) = f^*(\mathrm{y}_0)$.
Similarly for f.

(iii) Our proof of this is obtained by forging the **Link between the Simplex Method of Solution for the Primal and Dual Problems.**
Introducing slack variables $x_{n+1}, \ldots, x_{n+m}$ our primal problem is expressed in tableau form by:

	x_1	$\cdots$	x_j	$\cdots$	x_n	-1	
$-y_1$	a_{11}				a_{ln}	b_1	$-x_{n+1}$
$-y_1$			a_{ij}			b_i	$-x_{n+i}$
$-y_m$	a_{ml}				a_{mn}	b_m	$-x_{n+m}$
-1	c_1	$\cdots$	c_j	$\cdots$	c_n	0_d	$f(\mathrm{x})$
	$-y_{m+1}$		$-y_{m+j}$		$-y_{m+n}$	$-f*(\mathrm{y})$	

Introducing slack variables $y_{m+1}, \ldots, y_{m+n}$ for the dual problem we see that it may be represented on the same tableau as indicated. Applying our fundamental observations to this tableau for the dual problem we see that for a basic feasible solution we require all $c_j \geq 0$ and for that solution to be optimal $b_i \geq 0\ \forall_i$. **But**, this is precisely what the simplex algorithm achieves in solving the primal problem. Thus, whenever the primal problem has a solution (obtainable by the Simplex Method) so does the dual problem and, by the duality observation, vice versa.

EXAMPLE

For our diet problem:

Primal problem:

Minimize: $50x_1 + 10x_2 =: f(x)$

Subject to: $x_1 + x_2 \leq 3$

$x_1 + 5x_2 \leq 10$

$-2x_1 - x_2 \leq -1$

$x_1,\ x_2 \geq 0$

Dual problem:

Minimize: $3y_1 + 10y_2 - y_3 =: f*(x)$

Subject to: $-y_1 - y_2 + 2y_3 \leq 50$

$-y_1 = 5y_2 + y_3 \leq 10$

$y_1, y_2 \geq 0$

Combined tableau:

	x_1	x_2	-1	
$-y_1$	1	1	3	$-x_3$
$-y_2$	1	5	10	$-x_4$
$-y_3$	-2	-1	-1	$-x_5$
-1	50	10	0	$f(\mathbf{x})$
	$-y_4$	$-y_5$	$-f*(\mathbf{x})$	

Stage I

Augmented tableau

	x_0	x_1	x_2	-1	
$-y_1$	-1	1	1	3	$-x_3$
$-y_2$	-1	-1	1	5	$-x_4$
$-y_3$	-1	-2	-1	-1	$-x_5$
	1	0	0	0	$g(\mathbf{x})$
-1	0	50	10	0	$f(\mathbf{x})$
	*	$-y_4$	$-y_5$	$-f*(\mathbf{x})$	

	x_5	x_1	x_2	-1	
$-y_1$	-1	3	2	4	$-x_3$
$-y_2$	-1	3	6	11	$-x_4$
*	-1	2	1	1	$-x_0$
	1	-2	-1	-1	g
-1	0	50	10	0	f
	$-y_3$	$-y_4$	$-y_5$	$-f*$	

Stage II solution for min g

	x_5	x_0	x_2	-1	
$-y_1$	$\frac{1}{2}$	$\frac{-3}{2}$	$\frac{1}{2}$	$\frac{5}{2}$	$-x_3$
$-y_2$	$\frac{1}{2}$	$-\frac{3}{2}$	$\frac{9}{2}$	$\frac{19}{2}$	$-x_4$
$-y_4$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-x_1$
	0	1	0	0	g
-1	25	-25	-15	-25	f
	$-y_3$	$*$	$-y_5$	$-f*$	

Note: $*$ and x_0 remain aligned and so are deleted together.
feasible tableau

	x_5	x_2	x_1	
$-y_1$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{5}{2}$	$-x_3$
$-y-2$	$\frac{1}{2}$	$\frac{9}{2}$	$\frac{19}{2}$	$-x_4$
$-y_4$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-x_1$
-1	25	-15	-25	f
	$-y_3$	$-y_5$	$-f*$	

Stage II

	x_5	x_1	-1	
$-y_1$		-1	2	$-x_3$
$-y_2$		-9	5	$-x_4$
$-y_5$	-1	2	1	$-x_2$
-1	10	30	-10	f
	$-y_3$	$-y_4$	$-f*$	

whence $\min f = 10$ at $(0,1,2,5,0)$
$\min f^* = -10$ at $(0,0,30,0)$.

Note: (1) The correspondences:

$y_1 \leftrightarrow x_{n+1} \quad i = 1,2,\ldots,m$

$y_{m+1} \leftrightarrow x_1 \quad i = 1,2,\ldots,n$

are preserved throughout, and may be used to fill in the y-borders without the need of carrying them through the pivots.

(2) Dual problems are of theoretical importance and arise naturally in Games theory (see Exercise 13). Further, if the coefficients of the objective function are all positive (often the case for a cost function) then the dual problem is in feasible form and so may be solved using only stage II of the algorithm with its dual solution giving a solution to the original problem.

Exercises

If you have access to and familiarity with a computer, you are urged to solve the problems indicated with a # by developing programmes to

(i) input a tableau

(ii) output a tableau

(iii) perform a pivot about a specified tableau entry

(iv) effect stage II of the simplex algorithm

(v) effect stage I of the simplex algorithm

(vi) determine the solution to the dual problem.

1) Convert the following optimization problem into a linear programming problem.

Minimize: $|x| + |y|$

Subject to: $x + y \leq 1$.

2) Convert each of the following linear programming problems into *standard form.*

(a) *Maximise:* $x_1 + 4x_2 + x_3$

Subject to:

$$\begin{aligned} 2x_1 - 2x_2 + x_3 &= 4 \\ x_1 \qquad\quad - x_3 &\geq 1 \\ x_2 \geq 0; \quad x_3 &\geq 0 \end{aligned}$$

(b) The linear programming problem which results from question 1.

(c) *Minimize:* $3x_1 - 2x_2$

Subject to:

$$\begin{aligned} 2x_1 - x_2 &\geq -2 \\ x_1 - 2x_2 &\leq 3 \\ x_1 + 2x_2 &\leq 11 \\ x_1 \geq 0, x_2 &\geq 0 \end{aligned}$$

3) Graphically determine the extreme points of the constraint set of problem 2(c). Hence solve the problem.

4) Determine all basic solutions for the standard form of the problem in 2(c). Identify the basic feasible solutions. Verify the solution obtained in 3).

*5. Given a linear programming problems in the form

$$\begin{aligned} &\textit{Minimize:} && \mathbf{c}\cdot\mathbf{x} \\ &\textit{Subject to:} && A\mathbf{x}\le\mathbf{b} \\ &&& \mathbf{x}\ge 0 \end{aligned} \qquad \ldots(1)$$

where A is an $m\times n$ matrix, let the problem

$$\begin{aligned} &\textit{Minimize:} && \mathbf{c}_s\cdot\mathbf{x}_s \\ &\textit{Subject to:} && [A\vdots I_m]\mathbf{x}_s=\mathbf{b} \\ &&& \mathbf{x}_s\ge \mathbf{0} \end{aligned} \qquad \ldots(2)$$

correspond to a *standard form* of the problem (1).

Show that there is a one-to-one correspondence between the constraint set C for problem (1) and the constraint set $C_s \subseteq \mathbb{R}^{n+m}$, of (2). Show that under this correspondence extreme points of C correspond to extreme points of C_s.

6) Express the linear programming problems of question 2) in *Tableau form*.

#7) For the tableau

x_1	x_2	x_3	x_4	-1	
2	−2	−2	1	4	$-x_5$
−2	2	2	−1	−4	$-x_6$
−1	1	0	1	−1	$-x_7$
−1	1	−4	−1	0	

Perform a pivot about: (i) the (2, 3) position
(ii) the (2, 1) position

#8) Solve using the simplex method the linear programming problem in 2(c).

*9) Estimate approximately the *worst case complexity* (maximum number of operations $+, -, \times$ and $\div$ necessary) for the simplex algorithm when applied to a tableau of size $n\times n$ which is in feasible form.

#10) Solve, using the simplex method;

$$\begin{aligned} &\textit{Minimize:} && -3x_1+2x_2 \\ &\textit{Subject to:} && 4x_1+x_2\le 16 \\ &&& x_1+x_2\ge 7 \\ &&& 2x_1-x_2\le -4 \\ &&& x_1\ge 0, x_2\ge 0. \end{aligned}$$

*11) Estimate the (worst case) complexity for stage I of the simplex algorithm. Combine this with (9) to obtain an estimate for the over-all complexity of the simplex method. [*Note:* Extensive experience has shown that the algorithm can be expected to reach an optimum solution after the order of n privot operations!]

#12) For the problem;

$$\begin{aligned} \textit{Minimize:} \quad & 5x_1 - 3x_2 \\ \textit{Subject to:} \quad & 2x_1 - x_2 + 4x_3 \leq 4 \\ & x_1 + x_2 + 2x_3 \leq 5 \\ & 2x_1 - x_2 + x_3 \leq 1 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

(a) Find a solution using the simplex method.
(b) What is the dual problem?
(c) What is the solution to the dual problem?

13) For our game theory example (see end of Chapter 3), show that the Linear programming problems for A and $B's$ optimal strategy are dual problems. Use the simplex algorithm to solve for both strategies.

#14) For the linear programming problems of 2(c) and (10), find the respective dual problems and their solutions.

*15) For a linear programming problem in the form;

$$\begin{aligned} \textit{Minimize:} \quad & \mathbf{c} \cdot \mathbf{x} \\ \textit{Subject to:} \quad & A\mathbf{x} = \mathbf{b} \quad (A \text{ an } m \times (n+m) \text{ matrix}) \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

show that the dual problem may be expressed as

$$\begin{aligned} \textit{Minimize:} \quad & -\mathbf{b} \cdot \mathbf{z} \\ \textit{Subject to:} \quad & A^T\mathbf{z} \leq \mathbf{c} \end{aligned}$$

(note the absence of positivity constraints)

[Hint: take $\mathbf{z} = (y_{m+1} - y_1,\ y_{m+2} - y_2,\ \ldots,\ y_{2m} - y_m)$]

CHAPTER 5 - Non-Linear Optimization

For $\mathbf{x} \in \mathbb{R}^n$ we consider the problem;

Minimize: $\quad f_0(\mathbf{x})$

Subject to:
$$\left.\begin{array}{l} f_1(\mathbf{x}) \leq 0 \\ \cdots \\ f_i(\mathbf{x}) \leq 0 \\ \cdots \\ f_m(\mathbf{x}) \leq 0 \end{array}\right\} \quad \mathbf{f}(\mathbf{x}) \leq 0 (\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m)$$

where the $f_i (i = 0, \ldots, m)$ are given (not necessarily linear) functions of the n variables $x_1, x_2, \ldots, x_n$.

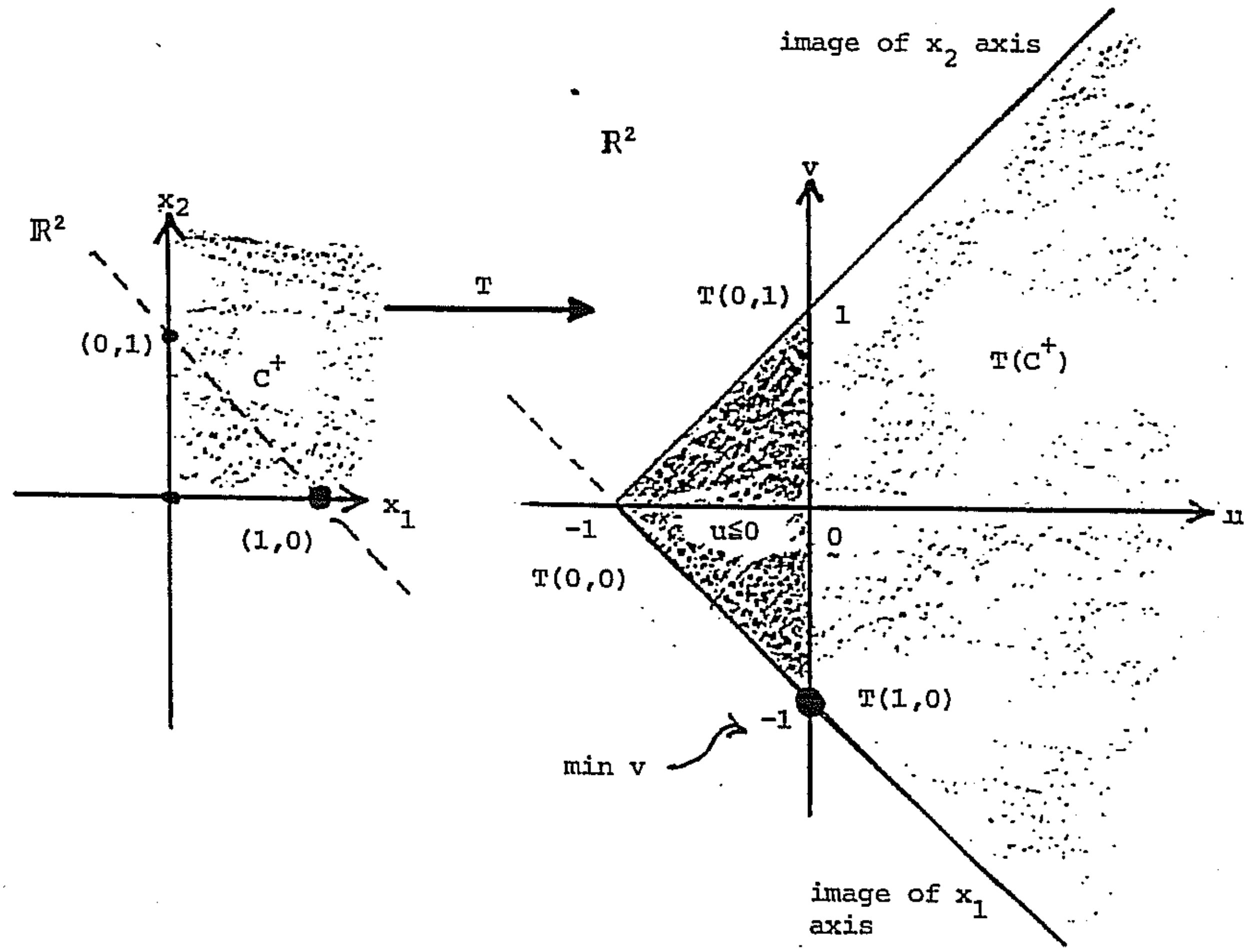
To gain perspective on our method of approach we examine a few *examples* of the case $n = 2, m = 1$.

(1) To get us thinking in the right direction we begin by exploring the **linear programming problem**;

Minimize: $\quad f_0(x_1,\ x_2) = x_2 - x_1$
Subject to: $\quad f_1(x_1,\ x_2) = x_1 + x_2 - 1 \leq 0$
$\quad x_1 \geq 0,\ x_2 \geq 0,$

but from a different geometric view point.
Define the (affine) mapping

$$\begin{aligned} T : \mathbb{R}^2 \to \mathbb{R}^2 : (x_1,\ x_2) \mapsto & (f_1(x_1, x_2), f_0(x_1, x_2)) \\ := & (x_1 + x_2 - 1,\ x_2 - x_1) \\ = & (x_1 + x_2,\ x_2 - x_1) - (1, 0) \\ = & \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned}$$



We see that the image under T of the positive quadrant C^+ (consisting of those $\mathbf{x}'s$ satisfying the positivity constraint) is the cone illustrated. Translated our problem becomes, Find;

Minimum:	v
Subject to:	$u \leq 0$
	$(u, v) \in T(C^+)$

Clearly this is -1, at $(0, -1) = T(1, 0)$. So the solution is $\min f_0 = -1$ at $x_1 = 1, x_2 = 0$, the correctness of which is easily verified.

A similar transformation and reinterpretation of the non-linear case is possible.

(2) *Minimize:* $2x + y$

Subject to: $x^2 + y^2 \leq 1$

Define:

$$T : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (u, v) := (x^2 + y^2 - 1, 2x + y)$$

Since there are no positivity constraints we seek the image of all of $\mathbb{R}^2$ under T ($T(\mathbb{R}^2)$, the range of T).

To do this, consider the image under T of the line $y = mx$ (since $\mathbb{R}^2$ is the union of such lines over all m, $T(\mathbb{R}^2)$ will be the union of the images).

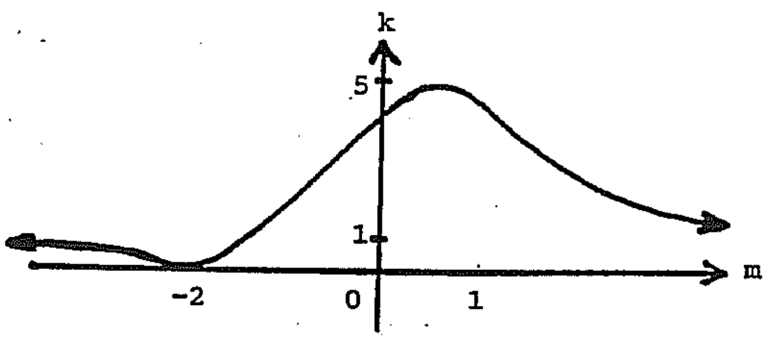
Now

$$v = (m + 2)x \text{ and } u = (m^2 + 1)x^2 - 1 \Rightarrow v^2 = \frac{(m+2)^2}{m^2+1}(u + 1)$$

Thus $T(\mathbb{R}^2)$ is the union of the family of parabolas $v^2 = k(u + 1)$ where k ranges over all possible values of $\frac{(m+2)^2}{m^2+1}$

ie. $0 \leq k \leq \max_m \frac{(m+2)^2}{m^2+1}$

$= 5$



$$\frac{(2m+4)(m^2+1) - 2(m+2)^2 m}{(m^2+1)^2} = 0$$

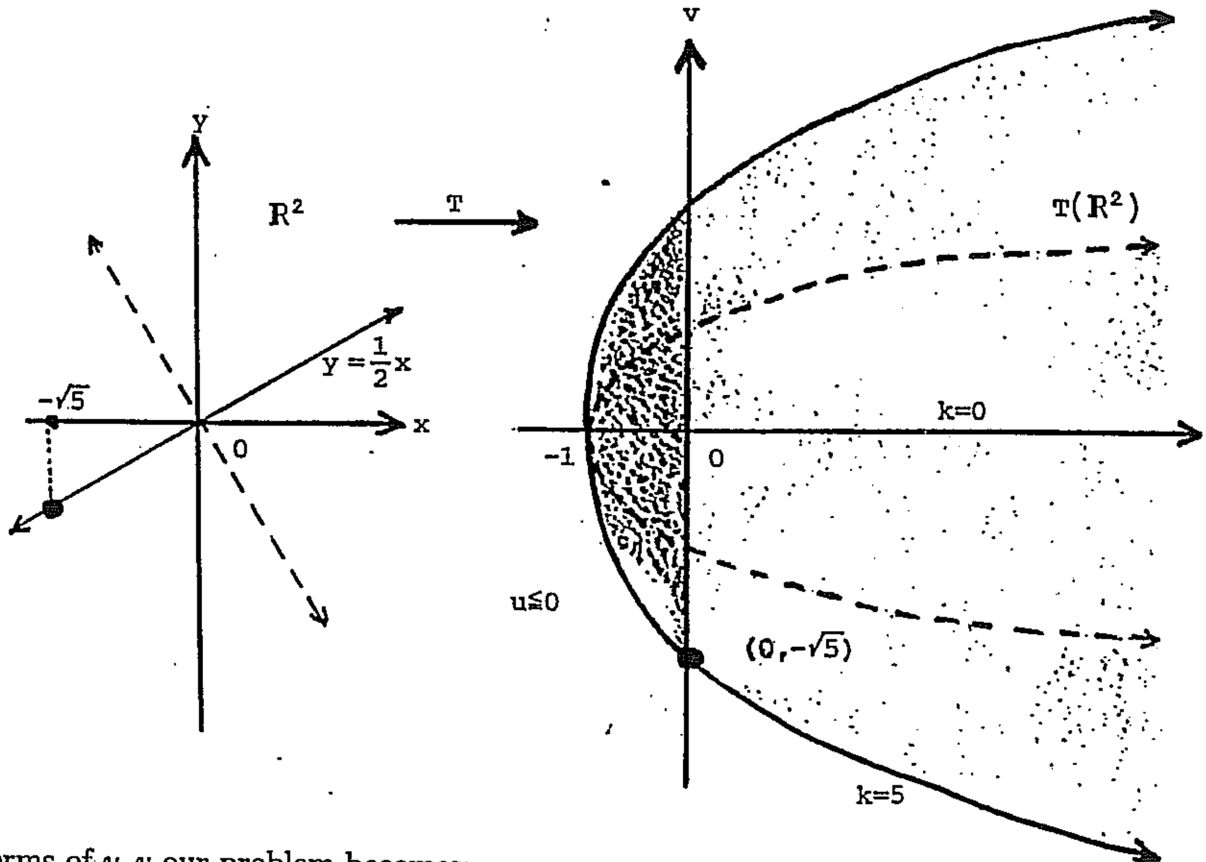
$$\Rightarrow 2m^2 + 3m - 2 = 0$$

$$(2m - 1)(m + 2) = 0$$

So

$$m = \frac{1}{2}, \quad -2$$

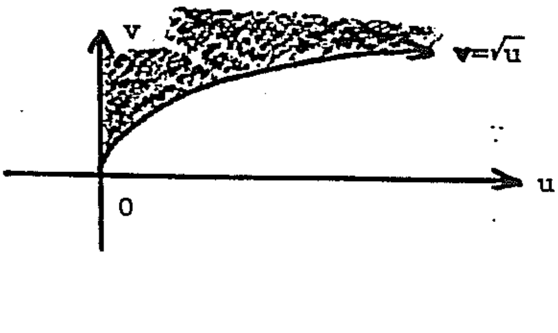
$$\therefore\ k = 5, \quad k = 0$$



In terms of u, v our problem becomes;

Minimize: v
Subject to: $u \leq 0$
$(u, v) \in T(\mathbb{R}^2)$

Clearly the solution is $-\sqrt{5}$ at $(0, -\sqrt{5}) = T(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$.
That is, the minimum is $-\sqrt{5}$ at $x = -\frac{2}{\sqrt{5}}$, $y = -\frac{1}{\sqrt{5}}$.
Note: Here $T(\mathbb{R}^2)$ is convex, in general however, even when f_0 and f_1 are convex functions, this need not be the case.
$f_0 = x^2 + y^2$ and $f_1 = x^4 \Rightarrow v = \sqrt{u} + y^2 \geq \sqrt{u}$ and $u \geq 0$.



(3) **A nonconvex example**

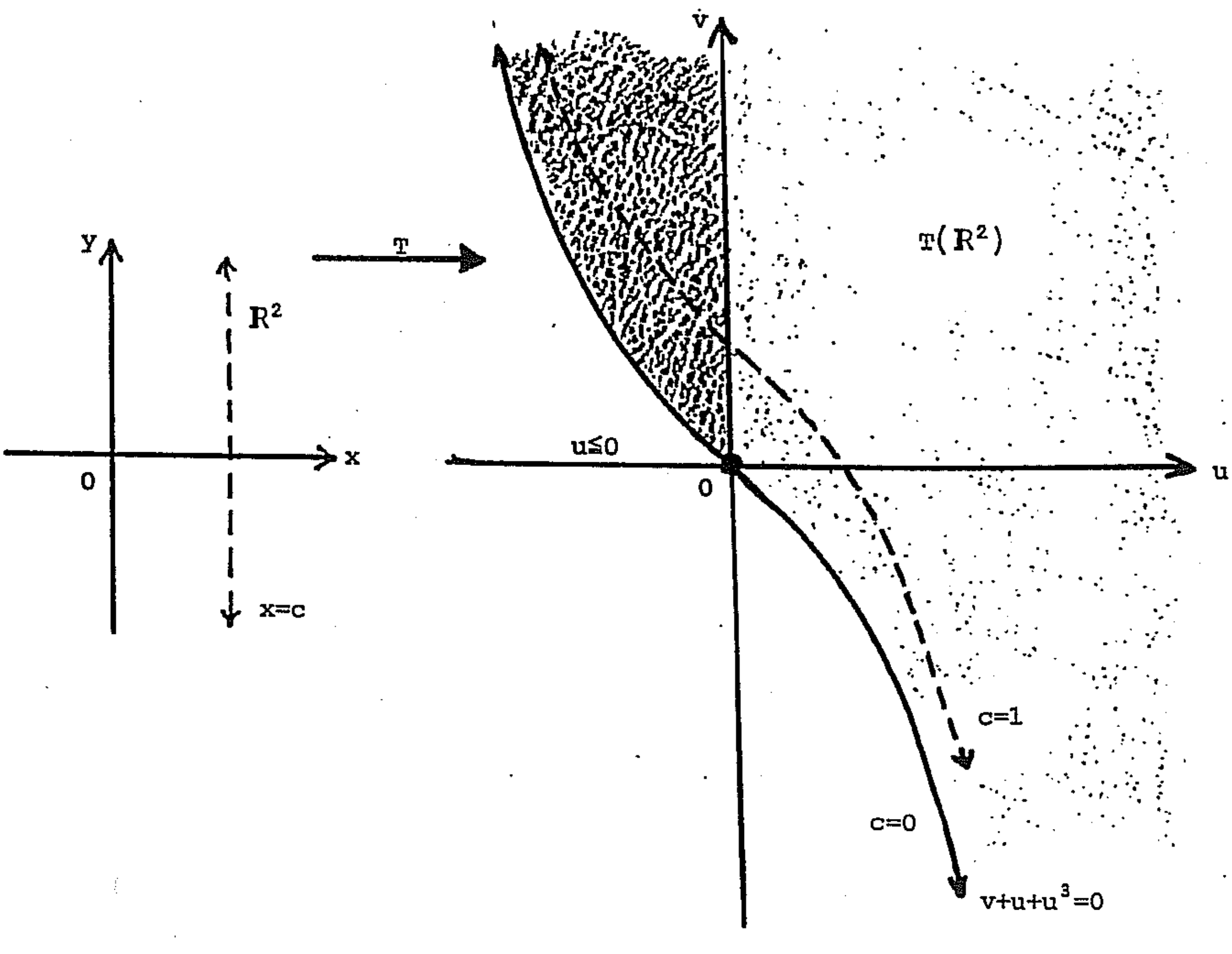
Minimize: $f_0(x, y) := x^2 + y + y^3$
Subject to: $f_1(x, y) := -y \leq 0$
Define:

$$T : \mathbb{R}^2 \mapsto \mathbb{R}^2 : (x, y) \to (u, v) := (-y, x^2 + y + y^3)$$

To determine $T(\mathbb{R}^2)$ observe: All points (c, y) on the line $x = c$ transform to points (u, v) satisfying: $v = c^2 - u - u^3$ or points on the curve $u + u^3 + v = c^2 \geq 0$.

Thus $T(\mathbb{R}^2)$ is the union, over all values of c, of such curves. That is,

$$T(\mathbb{R}^2) = \{(u, v) : u + u^3 + v \geq 0\}$$



In terms of u, v our problem is;

$$\begin{aligned} &\textit{Minimize:} && v \\ &\textit{Subject to:} && u \leq 0 \\ & && (u,v) \in T(\mathbb{R}^2) \end{aligned}$$

Clearly the solution is 0 at $(0, 0)$.

So min f_0 is 0 at $x = 0, y = 0$.

Remark Clearly any problem of the form;

$$\begin{aligned} &\textit{Minimize:} && f_0(x,y) \\ &\textit{Subject to:} && f_1(x,y) \leq 0, \end{aligned}$$

can be analysed in this way. The possibility of extending into higher dimensions (more constraint equations and unknowns) is also obvious. However, as the last two examples serve to illustrate, explicitly determining the range of T is in general a forbidding task. Instead of attempting this we shall seek distinguishing properties of the point in the image space where the minimum occurs (such properties then give necessary conditions for a solution).

MULTIPLIER RULES - THE BASIC STRATEGY

Given the following problem: For $\mathrm{x} \in \mathbb{R}^n$

$$
\begin{array}{ll}
\textit{Minimize:} & f_0(\mathrm{x}) \\
\textit{Subject to:} & f_1(\mathrm{x}) \leq 0 \\
 & \dots \\
 & f_m(\mathrm{x}) \leq 0,
\end{array}
$$

define $T : \mathbb{R}^n \to \mathbb{R}^{m+1} : \mathrm{x} \mapsto (f_1(\mathrm{x}), \dots, f_m(\mathrm{x}), f_0(\mathrm{x}))$.

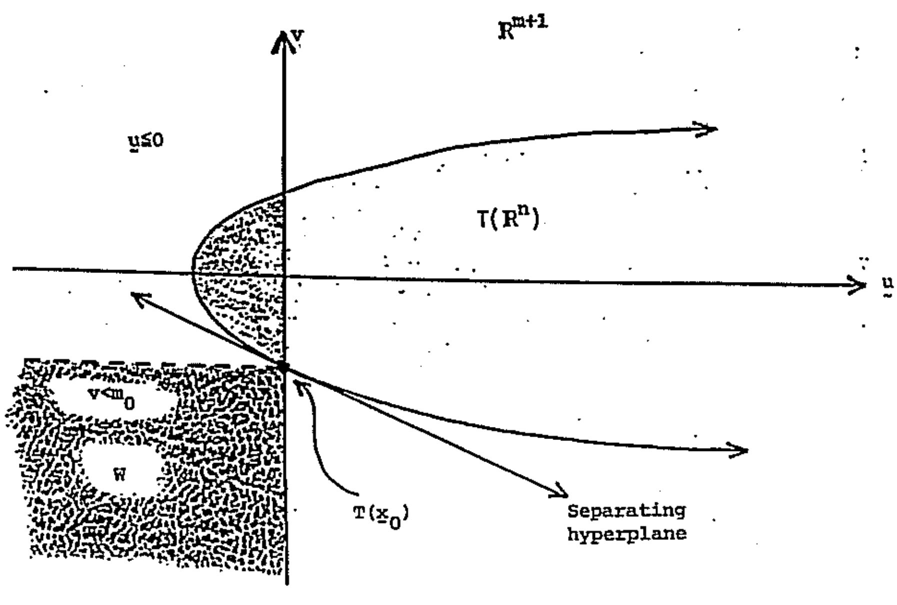
Now suppose the minimum value of f_0 subject to our constraints is m_0, then

$$W := \{(u_1, \dots, u_m, v) : u_1 \leq 0, \dots, u_m \leq 0, v < m_0\}$$

is clearly a convex subset of $\mathbb{R}^{m+1}$ with interior points, and we have

$$W \cap T(\mathbb{R}^n) = \phi$$

while, $\overline{W} \cap T(\mathbb{R}^n)$ consists precisely of the image under T of those points x_0 satisfying the constraints at which the minimum occurs. The converse is also true.



For motivation, we begin by considering the case when $T(\mathbb{R}^n)$ is a convex subset of $\mathbb{R}^{m+1}$ *(illustrated by Example (2) above).* In this case we can separate W and $T(\mathbb{R}^n)$ by a hyperplane $(\lambda_1, \ldots, \lambda_m, \lambda_0).(u_1, \ldots, u_m, v) = c$, where $\|\underline{\lambda}\| = 1$. That is, for $u_1, \ldots, u_m \leq 0$ and $v \leq m_0$ we have

$$\lambda_1 u_1 + \ldots + \lambda_m u_m + \lambda_0 v \leq \underline{\lambda} \cdot T(\mathrm{x}_0) \leq \underline{\lambda} \cdot T(\mathrm{x}).$$

Since $u_1, \ldots, u_m$ and v can assume arbitrarily large negative values it follows that $\lambda_0, \lambda_1, \ldots, \lambda_m$ must all be positive (otherwise the lower inequality would be violated).

Further, observing that for any $j \in \{1, \ldots, m\}$

$$(f_1(\mathrm{x}_0), \ldots, \frac{1}{2} f_j(\mathrm{x}_0), \ldots, f_0(\mathrm{x}_0)) \in \overline{W}$$

we have

$$\lambda_1 f_1(\mathrm{x}_0) + \ldots + \frac{1}{2}\lambda_j f_j(\mathrm{x}_0) + \ldots + \lambda_0 f(\mathrm{x}_0) \leq \lambda \cdot T(\mathrm{x}_0) = \lambda_1 f_1(\mathrm{x}_0) + \ldots \lambda_j f_j(\mathrm{x}_0) + \ldots$$

or

$$-\frac{1}{2}\lambda_j f_j(\mathrm{x}_0) \leq 0.$$

On the other hand, since $\lambda_j \geq 0$ and $f_j(\mathrm{x}_0) \leq 0$ we have $-\frac{1}{2}\lambda_j f_j(\mathrm{x}_0) \geq 0$.

That is, for $j = 1, 2, \ldots, m$ we have $\lambda_j f_j(\mathrm{x}_0) = 0$.

Thus we have proved (in the case $T(\mathbb{R}^n)$ is convex) the **convex multiplier rule.**

Convex Multiplier Rule

There exists positive $\lambda_0, \lambda_1, \ldots, \lambda_m$, not all zero, such that if $\phi(\mathrm{x}) := \sum_{j=0}^{m} \lambda_j f_j(\mathrm{x})$ then

$$\lambda_0 f_0(\mathrm{x}_0) = \phi(\mathrm{x}_0) \leq \phi(\mathrm{x}) \quad \text{for all } \mathrm{x} \in \mathbb{R}^n.$$

That is, ϕ **has an unconstrained (global) minimum at** x_0 (the point where f_0 is a min subject to the constraints) and

$$\lambda_j f_j(\mathrm{x}_0) = 0 \quad j = 1, 2, \ldots, m$$

(constraints where $\lambda_j \neq 0$ are said to be *active*, these are the constraints where $f_j(\mathrm{x}_0) = 0$ so small changes in x can violate them.)

Example

For the problem of Example 2, we see that the separating hyperplane is $v + \frac{\sqrt{5}}{2}u = -\sqrt{5}$, that is,

$$\lambda_0 = 1, \lambda_1 = \frac{\sqrt{5}}{2}$$

$$\phi(x, y) = 2x + y + \frac{\sqrt{5}}{2}(x^2 + y^2 - 1)$$

$$= \frac{\sqrt{5}}{2}(x + \frac{2}{\sqrt{5}})^2 + \frac{\sqrt{5}}{2}(y + \frac{1}{\sqrt{5}})^2 - \sqrt{5}$$

which has a minimum of $-\sqrt{5}$ at $x =$ at $x = -\frac{2}{\sqrt{5}}, y = -\frac{1}{\sqrt{5}}$.

Also note $f_1(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}) = (-\frac{2}{\sqrt{5}})^2 + (-\frac{1}{\sqrt{5}})^2 - 1 = 0$, that is, f_1 is an active constraint.

We now observe how the above rule could have been used to solve the problem.

We seek $\mu, \lambda \geq 0$ (not both zero) and x_0, y_0 so that $\phi(x, y) := \mu(2x+y)+\lambda(x^2+y^2-1)$ has a global minimum at (x_0, y_0) and $\lambda(x_0^2 + y_0^2 - 1) = 0$.

Now, if $\mu = 0$, then $\lambda \neq 0$ so $x_0^2 + y_0^2 = 1$ and $\phi(x_0, y_0) = 0$ which is not a minimum as $\phi(0, 0) = -\lambda < 0$.

So we must have $\mu \neq 0$. Dividing ϕ by μ and replacing $\frac{\lambda}{\mu}$ by λ we obtain the equivalent problem:

$$2x + y + \lambda(x^2 + y^2 - 1) \text{ is a minimum at } (x_0, y_0)$$

and

$$\lambda(x_0^2 + y_0^2 - 1) = 0.$$

In case $\lambda = 0, x_0$ and y_0 are unconstrained and $\phi(x, y) = 2x + y$ has no minimum. Thus the only possibility for a solution is to have $\lambda \neq 0$.

Differentiating we obtain:

$$\frac{\partial \phi}{\partial x} = 2 + 2\lambda x = 0 \text{ at } x_0 = -\frac{1}{\lambda}$$
$$\frac{\partial \phi}{\partial y} = 1 + 2\lambda y = 0 \text{ at } y_0 = -\frac{1}{2\lambda}$$

and from $x_0^2 + y_0^2 - 1 = 0$ we have

$$\frac{1}{\lambda^2} + \frac{1}{4\lambda^2} - 1 = 0$$

or

$$\lambda = \pm\frac{\sqrt{5}}{2}.$$

Since we require $\lambda > 0$ the only possibility is $\lambda = \frac{\sqrt{5}}{2}$ in which case $x_0 = -\frac{\sqrt{2}}{5}$ and $y_0 = -\frac{1}{\sqrt{5}}$. Thus the point $(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$ is the only possible location of a minimum for f_0 satisfying the constraint. Since a minimum exists (continuous function on a compact set) this **must** be its location. (cf. previous graphical solution).

Even when its is true, establishing the convexity of $T(\mathbb{R}^n)$ is usually not practicable. Accordingly we now turn to the general case where $T(\mathbb{R}^n)$ **need not be convex** (cf. Example 3 and the problem in the Note at the end of Example 2). Here the strategy is to "approximate" $T(\mathbb{R}^n)$ (at least locally near $T(\mathbf{x}_0)$) by a convex set K and then use an argument similar to that above, separating W and K, to obtain a multiplier rule. We distinguish two cases.

(1) **The convex case;** $f_0, f_1, \ldots, f_m$ are convex functions.

(2) **The smooth case;** $f_0, f_1, \ldots, f_m$ are continuously differentiable (at $\mathbf{x}_0$).

1. **The convex case:** In this case we construct a convex set K so that $W \cap K = \phi$ and $T(\mathbb{R}^n) \subseteq K$. Since separating K from W also separates $T(\mathbb{R}^n)$ we can use precisely the same argument as before to obtain
   convex multiplier Rule:
   If $f_0, f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ are convex functions and $\mathbf{x}_0$ minimizes f_0 subject to the constraints $f_j(\mathbf{x}) \leq 0 (j = 1, 2, \ldots, m; \mathbf{x} \in \mathbb{R}^n)$ then there exists positive $\lambda_0, \lambda_1, \ldots, \lambda_m$

(not all zero) so that for

$$\phi(\mathbf{x}) := \sum_{j=0}^{m} \lambda_j f_j(\mathbf{x}) \text{ we have}$$

$$\lambda_0 f_0(\mathbf{x}_0) = \phi(\mathbf{x}_0) \le \phi(\mathbf{x}) \text{ for all } \mathbf{x} \in R^n$$

and

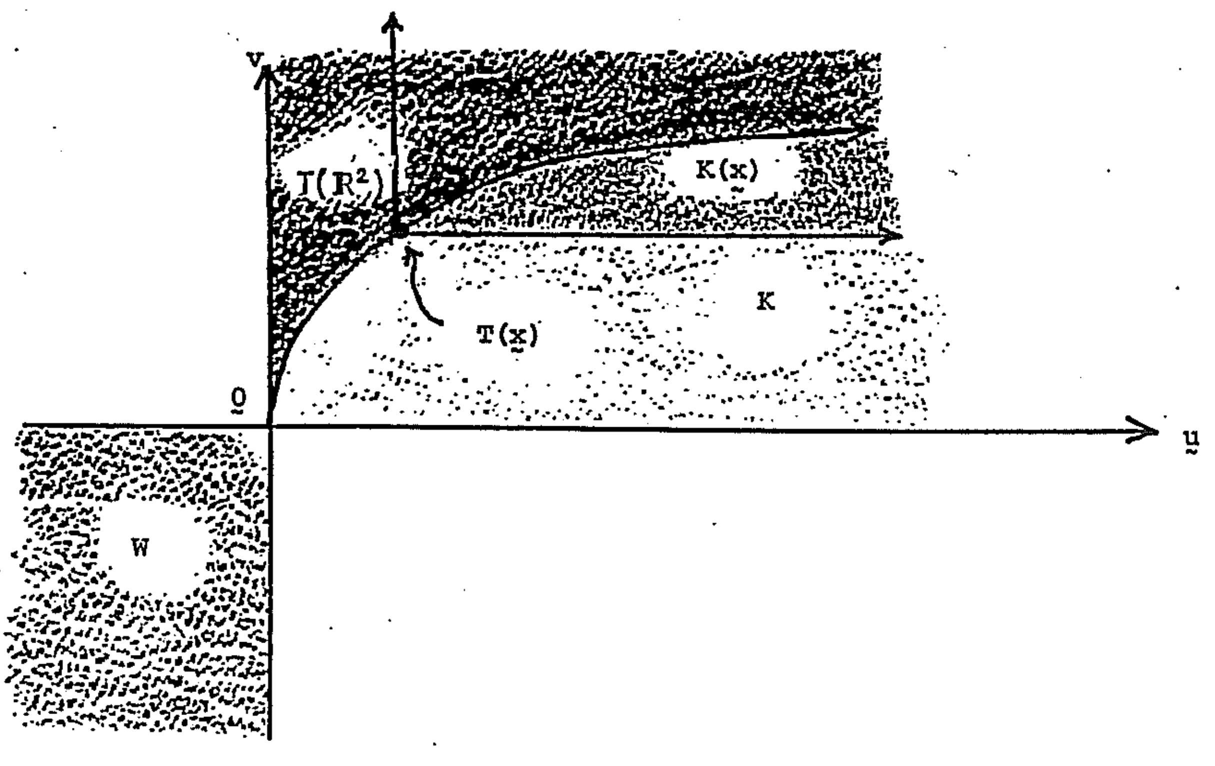
$$\lambda_j f_j(\mathbf{x}_0) = 0 \text{ for } j = 1, 2, \ldots, m.$$

Construction of K:
For $\mathbf{x} \in \mathbb{R}^n$ let

$$K(\mathbf{x}) := \{(u_1, u_2, \ldots, u_m, v) : f_1(\mathbf{x}) \le u_1, \ldots, f_m(\mathbf{x}) \le u_m, f_0(\mathbf{x}) \le v\}$$

and set

$$K := \bigcup_{\mathbf{x} \in \mathbb{R}^n} K(\mathbf{x})$$



We verify that K has the desired properties.
That $K \supseteq T(\mathbb{R}^n)$ is obvious, since $T(\mathbf{x}) \in K(\mathbf{x})$.

To show $W \cap K = \phi$: suppose not, then there exist $(u_1, \ldots, u_m, v) \in W \cap K$. But this implies $v < m_0$ and $f_0(\mathbf{x}) \leq v$ for some $\mathbf{x} \in \mathbb{R}^n$, with $f_j(\mathbf{x}) \leq u_j \leq 0$ $(j = 1, 2, \ldots, m)$ by the definition of W. That is, $f_0(\mathbf{x}) < m_0$ and $\mathbf{x}$ satisfies the constraints, contradicting the definition of m_0.

To show K is convex, let $\mathbf{u} = (u_1, u_2, \ldots, u_m, v)$ and $\mathbf{r} = (r_1, r_2, \ldots, r_m, s)$ be points of K. That is for some $\mathbf{x}$ and $\mathbf{y}$ we have

$$\mathbf{u} \in K(\mathbf{x}) \text{ and } \mathbf{r} \in K(\mathbf{y}).$$

Then for $0 \leq \lambda \leq 1$ we have

$$\begin{aligned} \mathbf{w} &= \lambda(u_1, \ldots, u_m, v) + (1-\lambda)(r_1, \ldots, r_m, s) \\ &= (\lambda u_1 + (1-\lambda)r_1, \ldots, \lambda v + (1 - \lambda s)) \\ &\geq (\lambda f_1(\mathbf{x}) + (1-\lambda) f_1(\mathbf{y}), \ldots, \lambda f_0(\mathbf{x}) + (1-\lambda) f_0(\mathbf{y})) \\ &\geq (f_1(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}), \ldots, f_0(\lambda \mathbf{x} + (1-\lambda)\mathbf{y})) \end{aligned}$$

so $\mathbf{w} \in K(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \subset K$.
Applying this to the problem in the Note at the end of Example 2, we seek

$\lambda \geq 0, \mu \geq 0$ (not both zero) and x_0, y_0 so that

$\mu(x^2 + y^2) + \lambda x^4$ is a global minimum at (x_0, y_0)

and

$$\lambda x_0^4 = 0.$$

If $\mu = 0$ than $\lambda \neq 0$ and so $x = 0$ in which case the minimum is at any point $(0, y)$. If $\mu \neq 0$ then in case $\lambda = 0$ we have a minimum at $(0, 0)$, while if $\lambda \neq 0$, taking partial derivatives we require

$$2x + 4x^3\lambda = 0$$

and

$$2y = 0$$

together with

$$x^4 = 0.$$

Leading again to a minimum at $(0, 0)$. So our search for a solution to the original problem is narrowed down to points of the form $(0, y)$. From which we readily see that the minimum is 0 at $(0, 0)$.

Corollary: **Karush, Kuhn-Tucker Condition for Optimality:**

(1) Let C be the feasible set $\{\mathrm{x} \in \mathbb{R}^n : f_j(\mathrm{x}) \leq 0, j = 1, 2, \ldots, m\}$ for the problem

$$\begin{array}{ll} \textit{Minimize:} & f_0(\mathrm{x}) \\ \textit{Subject to:} & f_1(\mathrm{x}) \leq 0 \\ & \ldots \\ & f_m(\mathrm{x}) \leq 0 \end{array}$$

If no $f_j (j = 1, 2, \ldots, m)$ is identically zero on C, then the conclusions of the Convex Multiplier Rule hold with a $\lambda_0 \geq 0$.

(2) Conversely: **If the conclusions of the convex Multiplier Rule hold with a** $\lambda_0 > 0$ **then** x_0 is an optimal solution.

Note: (2) gives a sufficient condition for optimality. For example, we see that (0, 0) is the optimal solution for the problem in the Note to Example 2.

Proof:

(1) Suppose not, ie $\lambda_0 = 0$ is the only possible value for λ_0. Then

$$0 = \lambda_0 f_0(\mathrm{x}_0) = \phi(\mathrm{x}_0) \leq \phi(\mathrm{x})$$

on the other hand if $\mathrm{x} \in C$

$$\begin{aligned} \phi(\mathrm{x}) :&= \lambda_0 f_0(\mathrm{x}) + \lambda_1 f_1(\mathrm{x}) + \ldots + \lambda_m f_m(\mathrm{x}) \\ &= 0 + (+ve)(-ve) + \ldots + (+ve)(-ve) \\ &\leq 0 \end{aligned}$$

That is, $\phi(\mathrm{x}) \equiv 0$ on C.
Since each term $\lambda_j f_j(\mathrm{x}) \leq 0$, they too must all be zero, but not all the λ_j are zero so we must have at least one $f_j(\mathrm{x}) \equiv 0$ on C contrary to our assumption.

(2) Suppose $\lambda_0 > 0$ then for $\mathrm{x} \in C$ we have

$$\phi(\mathrm{x}_0) \leq \phi(\mathrm{x})$$

or

$$\lambda_0 f_0(\mathrm{x}_0) \leq \lambda_0 f_0(\mathrm{x}) + \underbrace{\lambda_1 f_1(\mathrm{x}) + \ldots \lambda_m f_m(\mathrm{x})}_{\leq 0}$$

Thus, $\lambda_0 f_0(\mathrm{x}_0) \leq \lambda_0 f_0(\mathrm{x})$ and, since $\lambda_0 > 0$, we have

$$f_0(\mathrm{x}_0) \leq f_0(\mathrm{x}).$$

Remark: A useful result in this context, which we state without proof, is the following sufficient condition for convexity of $f : \mathbb{R}^n \to R$, which generalises; $f : \mathbb{R} \to \mathbb{R}$ is convex if $f'' \geq 0$.

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ have continuous second order partial derivatives $f_{ij}^2 := \frac{\partial^2 f}{\partial x_j \partial x_i}$ throughout a convex open set $U \subseteq \mathbb{R}^n$. Then f is convex on U if and only if the *Hessian matrix*

$$H(\mathrm{x}) := \begin{bmatrix} f_{11}^2 & \cdots & f_{1n}^2 \\ \cdots\cdots\cdots & & \\ f_{n1}^2 & \cdots & f_{nn}^2 \end{bmatrix} \text{(partial derivatives calculated at } \mathbf{x}\text{)}$$

is *positive semi-definite* for each $\mathrm{x} \in U$; that is,

$$\mathrm{y}^T H(\mathrm{x})\mathrm{y} \geq 0 \text{ for all } \mathrm{y} \in \mathbb{R}^n.$$

Moreover, f is strictly convex if $H(\mathbf{x})$ is *positive definite* on U; that is,

$$\mathrm{y}^T H(\mathrm{x})\mathrm{y} > 0 \text{ for all } \mathrm{y} \in \mathbb{R}^n, \mathrm{y} \neq \mathbf{0}.$$

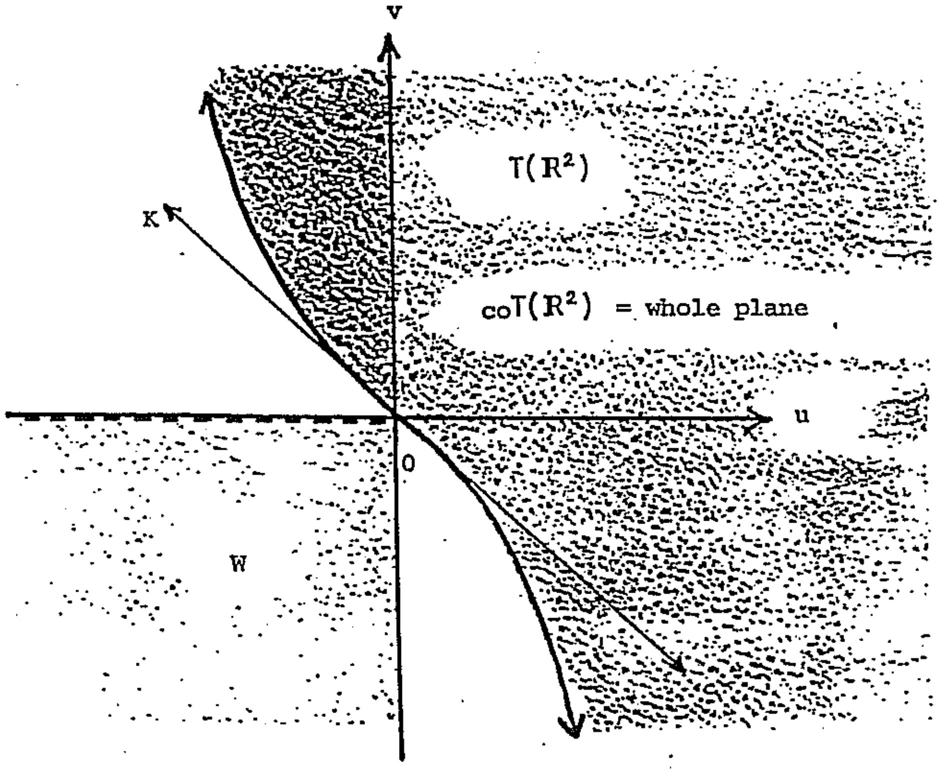
A convenient criteria for this is that each of the matrices

$$[f_{11}^2]\,, \begin{bmatrix} f_{11}^2 & f_{12}^2 \\ f_{21}^2 & f_{22}^2 \end{bmatrix}, \begin{bmatrix} f_{11}^2 & f_{12}^2 & f_{13}^2 \\ f_{21}^2 & f_{22}^2 & f_{23}^2 \\ f_{31}^2 & f_{32}^2 & f_{33}^2 \end{bmatrix}, \ldots, \mathrm{H}$$

have strictly positive determinant.

2. The Smooth Case

Here, as example 3 typifies, $T(\mathbb{R}^n)$ is neither convex, nor can it be included in any convex set (the smallest would be its convex hull) which remains disjoint from W, even locally. Thus we must find a different construction for K.



Preliminaries:

Recall: $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0 \in \mathbb{R}^n$ if there exists an affine function $A : \mathbb{R}^n \to \mathbb{R}^m : \mathbf{x} \mapsto A(\mathbf{x}) := L(\mathbf{x}) + \mathbf{b}$, where L is linear, such that

$$\mathbf{f}(\mathbf{x}) = A(\mathbf{x}) + \mathbf{r}(\mathbf{x})\|\mathbf{x} - \mathbf{x}_0\|$$

where

$$\mathbf{r}(\mathbf{x}) \to \mathbf{0} \text{ as } \mathbf{x} \to \mathbf{x}_0 .$$

Clearly $\mathbf{f} = (f_1, \dots, f_m)$ is differentiable if and only if each of the $f_i : \mathbb{R}^n \to \mathbb{R} (i = 1, \dots, m)$ is differentiable.

If f is differentiable at $\mathbf{x}_0$ we have $\mathbf{f}(\mathbf{x}_0) = L(\mathbf{x}_0) + \mathbf{b}$, that is $\mathbf{b} = f(\mathbf{x}_0) - L(\mathbf{x}_0)$, and if L corresponds to the matrix $[\ell_{ij}]$ with respect to the standard bases, letting $\mathbf{x} = \mathbf{x}_0 + h\mathbf{e}_j$

$(j = 1, 2 \ldots, n)$ we have

$$\mathbf{f}(\mathbf{x}_0 + h\mathbf{e}_j) - \mathbf{f}(\mathbf{x}_0) = hL(\mathbf{e}_j) + \mathbf{r}(\mathbf{x}_0 + h\mathbf{e}_j)h$$

so,

$$\frac{f_i(\mathbf{x}_0 + h\mathbf{e}_j) - f_i(\mathbf{x}_0)}{h} = \ell_{ij} + r_i(\mathbf{x}_0 + h\mathbf{e}_j).$$

Since $r_i(\mathbf{x}_0 + h\mathbf{e}_j) \to 0$ as $h \to 0$ we have

$$\ell_{ij} = \left.\frac{\partial f_i}{\partial x_j}\right|_{\mathbf{x}_0}.$$

ie. L is the *Jacobian* of f at x_0, $Df\big|_{\mathbf{x}_0}$ (or $\mathrm{Df}(\mathbf{x}_0)$) represented by the $m \times n$ matrix

$$D\mathbf{f}(\mathbf{x}_0) = \left[\left.\frac{\partial f_i}{\partial x_j}\right|_{\mathbf{x}_0}\right]_{(m\times n)}.$$

Theorem: f: $\mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0$ if and only if all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are *continuous* at $\mathbf{x}_0$.

Examples

(1) $f : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto x^2 + y^2$.
has

$$\begin{aligned} Df(x_0, y_0) &= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \\ &= [2x_0, 2y_0]. \end{aligned}$$

So, $\mathbf{b} = \mathrm{x}_0^2 + y_0^2 - 2x_0 - 2y_0$
and we have

$$x^2 + y^2 = 2x_0x + 2y_0 - x_0^2 - y_0^2 + r(\mathbf{x})\sqrt{(x - x_0)^2 + (y - y_0)^2}.$$

From which we see

$$r(\mathbf{x}) = \sqrt{(x - x_0)^2 + (y - y_0)^2}.$$

Note, $r(\mathbf{x}) \to 0$ as $\mathbf{x} \to \mathbf{x}_0$.

(2) $T : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (-y, x^2 + y + y^3)$
has

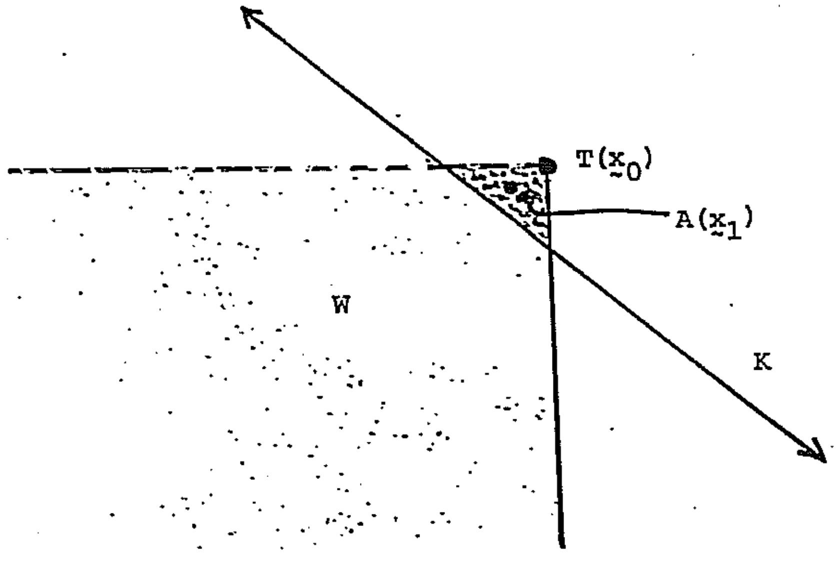
$$DT(x_0) = \begin{bmatrix} 0 & -1 \\ 2x_0 & 1 + 3y_0^2 \end{bmatrix}$$

which is continuous and so T is differentiable.

Our idea, in the case of a differentiable $T : \mathbb{R}^n \to \mathbb{R}^{m+1}$ is to replace ("approximate" near and Tx_0) the set $T(\mathbb{R}^n)$ by

$$K = A(\mathbb{R}^n) = DT\big|_{x_0}(\mathbb{R}^n) + T(x_0) - DT\big|_{x_0}(x_0) .$$

We begin by showing that K can be separated from W. Assume this were not true then $K \cap \text{ int } W \neq \phi$.



That is, there exists $x_1 \in \mathbb{R}^n$ s.t. $(u_{11}, \ldots u_{1m}, v_1) := A(x_1) \in \text{int } W$ so $u_{11}, \ldots, u_{1m} < 0$ and $v_1 < m_0$. That is, there exists $p > 0$ such that $u_{11}, \ldots, u_{1m} < -p$ and $v_1 < m_0 - p$.

For any $h > 0$ let

$$(u_{h1}, \ldots, u_{hm}, v_h) := A(hx_1 + (1-h)x_0) = hA(x_1) + (1-h)Tx_0.$$

Then $v_h \leq m_0 - hp$ and $u_{hi} \leq -hp$.

Now, from the differentiability of T we have

$$\|T(h\mathrm{x}_1 + (1-h)\mathrm{x}_0) - A(h\mathrm{x}_1 + (1-h)\mathrm{x}_0)\|$$
$$\leq \|\mathrm{r}(h\mathrm{x}_1 + (1-h)\mathrm{x}_0)\| h \|\mathrm{x}_1 - \mathrm{x}_0\|$$
$$\leq \tfrac{hp}{2}, \text{ for } h \text{ sufficiently small (since } h\mathrm{x}_1 + (1-h)\mathrm{x}_0 \to \mathrm{x}_0 \text{ and so } \mathbf{r} \to \mathbf{0}, \text{ as } h \to 0).$$

In particular then, taking h sufficiently small we will have that $T(h\mathrm{x}_1 + (1-h)\mathrm{x}_0)$ has its first m components negative indeed less than $hp/2$ and its last component less than $m_0 - hp/2 < m_0$ contradicting the choice of x_0 and m_0.

Now, since K and W can be separated and $T(\mathrm{x}_0) \in \bar{W}$ we have there exist $\lambda_1, \ldots, \lambda_m, \lambda_0$ not all zero, so that

$$\begin{aligned}\lambda_1 u_1 + \ldots + \lambda_m u_m + \lambda_0 v &\leq \underset{\sim}{\lambda} \cdot T(\mathrm{x}_0) \\ &\leq \underset{\sim}{\lambda} \cdot A(\mathrm{x})\end{aligned}$$

whenever $(u_1, \ldots, u_m, v) \in W$ and $\mathrm{x} \in \mathbb{R}^n$. Here $\underset{\sim}{\lambda}, := (\lambda_1, \lambda_2, \ldots, \lambda_m, \lambda_0)$.

So

$$\underset{\sim}{\lambda} \cdot T(\mathrm{x}_0) \leq \underset{\sim}{\lambda} \cdot [DT(\mathrm{x}_0)\mathrm{x} + T\mathrm{x}_0 - DT(\mathrm{x}_0)\mathrm{x}_0]$$

or

$$\underset{\sim}{\lambda} \cdot DT(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0) \geq 0 \text{ for all } \mathrm{x} \in \mathbb{R}^n .$$

That is, $\underset{\sim}{\lambda} \cdot DT(\mathrm{x}_0)\mathrm{y} \geq 0$ for all $\mathrm{y} \in \mathbb{R}^n$. Now suppose $\underset{\sim}{\lambda} \cdot DT(\mathrm{x}_0)\mathrm{y}_0 > 0$ for some $\mathrm{y}_0 \in \mathbb{R}^n$, then $\underset{\sim}{\lambda} \cdot DT(\mathrm{x}_0)(-\mathrm{y}_0) \lneq 0$ a contradiction. Thus $\underset{\sim}{\lambda} \cdot DT(\mathrm{x}_0)\mathrm{y} = 0$ for all $\mathrm{y} \in \mathbb{R}^n$.

In particular then,

$$(\lambda_1, \ldots, \lambda_m, \lambda_0) \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ & & \\ \frac{\partial f_0}{\partial x_1} & \cdots & \frac{\partial f_0}{\partial x_n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = 0 \text{ for all } y_1, y_2, \ldots, y_n.$$

That is, if $\phi := \lambda_1 f_1 + \ldots + \lambda_m f_m + \lambda_0 f_0$, we have

$$\begin{bmatrix} \frac{\partial \phi}{\partial x_1}, & \cdots & \frac{\partial \phi}{\partial x_{n-1}}, & \frac{\partial \phi}{\partial x_n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = 0 \text{ for all } y_1, \ldots, y_n.$$

so, $\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2}, \ldots, \frac{\partial \phi}{\partial x_n} = 0.$

That $\lambda_0, \lambda_1, \ldots, \lambda_n \geq 0$ follows as in the proof of the convex Multiplier Rule, as does $\lambda_i f_i(\mathrm{x}_0) = 0$ for $i = 1, 2, \ldots, m$.

Thus we have proved, **John Multiplier Rule:**

Given $f_0, f_1, \ldots, f_m : R^n \to \mathbb{R}$ let $C := \{\mathrm{x} \in \mathbb{R}^n : f_i(\mathrm{x}) \leq 0 \text{ for } i = 1, 2, \ldots, m\}$. If $f_0, f_1, \ldots, f_m$ are continuously differentiable at $\mathrm{x}_0 \in C$ and $f_0(\mathrm{x}_0) \leq f_0(\mathrm{x})$ for all $\mathrm{x} \in C$ then there exists $\lambda_0, \lambda_1, \ldots, \lambda_m \geq 0$ (not all zero) so that if

$$\phi(\mathrm{x}) = \sum_{j=0}^{m} \lambda_j f_j(\mathrm{x})$$

then

$\frac{\partial \phi}{\partial x_i}(\mathrm{x}_0) = 0 \quad (i = 1, \ldots, n)$ [That is, x_0 is a stationary point of ϕ.]

and

$$\lambda_j f_j(\mathrm{x}_0) = 0 \quad (j = 1, 2, \ldots, m)$$

Illustrative Application (to Example 3)

Minimize: $f_0(x, y) = x^2 + y + y^3$
Subject to: $f_1(x, y) = -y \leq 0$
Let $\phi(x, y) = \lambda_0 x^2 + \lambda_0 y + \lambda_0 y^3 - \lambda_1 y$, then we require

$$2\lambda_0 x_0 = 0$$
$$(\lambda_0 - \lambda_1) + 3\lambda_0 y_0^2 = 0$$
$$\lambda_1 y_0 = 0$$

If $\lambda_0 \neq 0$ then $x_0 = 0$ in which case $\lambda_1 \neq 0$ implies $y_0 = 0$ and $\lambda_1 = \lambda_0$ while $\lambda_1 = 0$ implies $3y_0^2 = -1$ (which is impossible). On the other hand $\lambda_0 = 0 \Rightarrow \lambda_1 = 0$ which is dissallowed. So the only solution is with $\lambda_0 = \lambda_1 > 0$ in which case $x_0 = y_0 = 0$ (cf. earlier graphical solution).

Since Multiplier methods have led us to consider unconstrained optimization problems it is appropriate that we conclude the course with a brief examination of such problems.

Unconstrained Optimization

The following result, known as **Ekeland's variational principal**, was established by I. Ekeland in 1973 for all normed linear spaces. We will only consider it for the finite dimensional space $\mathbb{R}^n$, in which case we can give a very simple proof, due to J.B. Hiriart-Urruty in 1983.

Theorem: Let V be a closed subset of $\mathbb{R}^n$ and $f : V \to \mathbb{R}$ a lower semi-continuous function which is bounded below. Given $\epsilon > 0$ let $\mathbf{u} \in V$ be such that

$$f(\mathbf{u}) \leq \inf_{\mathbf{x} \in V} f(\mathbf{x}) + \epsilon,$$

then for every $\lambda > 0$ there exists $\mathbf{v} \in V$ with

(1) $f(\mathbf{v}) \leq f(\mathbf{u})$,

(2) $\|\mathbf{u}-\mathbf{v}\| \leq \epsilon/\lambda$,

(3) for all $\mathbf{x} \in V, f(\mathbf{x}) \geq f(\mathbf{v}) - \lambda\|\mathbf{v}-\mathbf{x}\|$.

Proof: Let $g(\mathbf{x}) := f(\mathbf{x}) + \lambda\|\mathbf{x}-\mathbf{u}\|$, then g is also lower semi-continuous and bounded below. The set $V_0 := \{\mathbf{x} \in V : g(\mathbf{x}) \leq f(\mathbf{u})\}$ is closed, non-empty [$\mathbf{u} \in V_0$], and bounded [if there existed $\mathbf{x_n} \in \mathbf{V_0}$ with $\|\mathbf{x}_n\| \to \infty$ then $g(\mathbf{x}_n) \to \infty$, contradicting $g(\mathbf{x}_n) \leq f(\mathbf{u}]$, hence V_0 is compact, and so g restricted to V_0 achieves its minimum at some point $\mathbf{v}$. This is necessarily also a minimum for g on V, so

$$f(\mathbf{v}) + \lambda\|\mathbf{v}-\mathbf{u}\| \leq f(\mathbf{x}) + \lambda\|\mathbf{x}-\mathbf{u}\|, \text{ for all } \mathbf{x} \in V \qquad \ldots \qquad (*)$$

Substituting $\mathbf{x} = \mathbf{u}$ we get

$$f(\mathbf{v}) + \lambda\|\mathbf{v}-\mathbf{u}\| \dot{<} f(\mathbf{u}) ,$$

from which (1) follows.

Further, since $f(\mathbf{u}) \leq \inf_{\mathbf{x}\in V} f(\mathbf{x}) + \epsilon \leq f(\mathbf{v}) + \epsilon$, we also have

$$f(\mathbf{u}) - \epsilon + \lambda\|\mathbf{v}+\mathbf{u}\| \leq f(\mathbf{v}) + \lambda\|\mathbf{v}+\mathbf{u}\| \leq f(\mathbf{u})$$

from which we readily deduce (2).

To verify (3) we need only note that from (*) we have

$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{v}) + \lambda\big[\|\mathbf{v}-\mathbf{u}\| - \|\mathbf{x}-\mathbf{u}\|\big] \\ &\geq f(\mathbf{v}) - \lambda\|\mathbf{v}-\mathbf{x}\|, \end{aligned}$$

as $\quad \|\mathbf{v}-\mathbf{u}\| - \|\mathbf{x}-\mathbf{u}\| \geq -\|\mathbf{v}-\mathbf{x}\|$.

Corollary 1 If V is a closed subset of $\mathbb{R}^n, f : V \to \mathbb{R}$ is lower semi-continuous and bounded below. Then for any $\epsilon > 0$ and $\mathbf{u} \in V$ with

$$f(\mathbf{u}) \leq \inf_{\mathbf{x}\in V} f(\mathbf{x}) + \epsilon,$$

there exists $\mathbf{v} \in V$ with

(i) $\|\mathbf{v} - \mathbf{u}\| < \epsilon$, and

(ii) $f(\mathbf{x}) \geq f(\mathbf{v}) - \epsilon\|\mathbf{v} - \mathbf{x}\|$ for all $\mathbf{x} \in V$.

That is: There exists a 'small' perturbation of f; $g(\mathbf{x}) := f(\mathbf{x}) + \epsilon\|\mathbf{v} - \mathbf{x}\|$, which achieves its minimum at the point $\mathbf{v}$ near to $\mathbf{u}$ where f is almost a minimum - hence the name 'variational principle'.

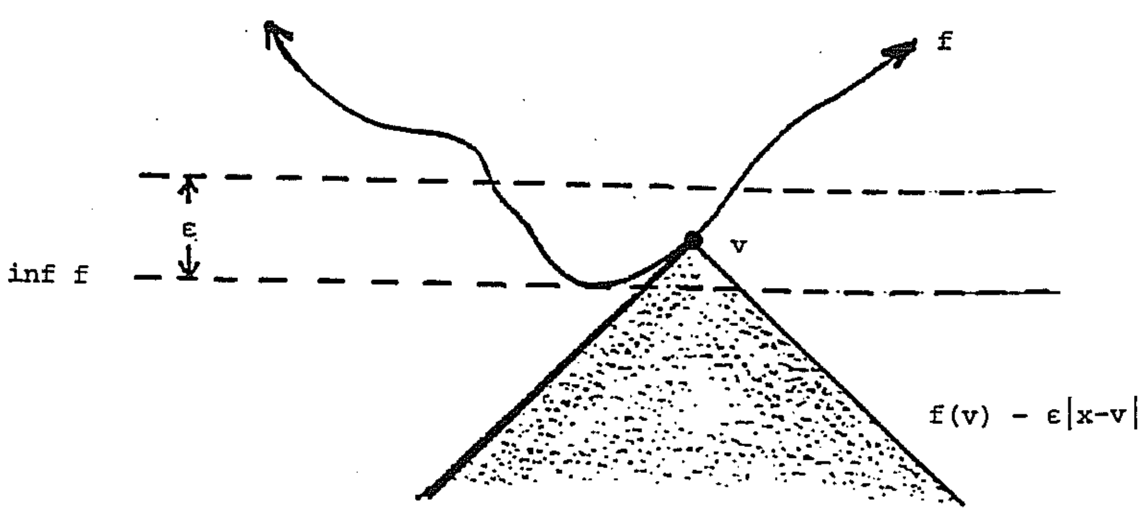
Proof: Replace ϵ by ϵ^2 and λ by ϵ in the theorem.

Corollary 2 If V is a closed subset of $\mathbb{R}^n$, $f : v \to \mathbb{R}$ is lower semi-continuous and bounded below, then for any $\epsilon > 0$ there exists $v \in V$ such that

$$f(\mathbf{v}) \leq \inf_{\mathbf{x} \in V} f(\mathbf{x}) + \epsilon,$$

and for all $\mathbf{x} \in V$ we have $f(\mathbf{x}) \geq f(\mathbf{v}) - \epsilon\|\mathbf{x} - \mathbf{v}\|$.

Proof: Take $\lambda = \epsilon$ in Theorem. Note: We also have $\|\mathbf{v} - \mathbf{u}\| < 1$.



Corollary 3 If f is differentiable on V, then given $\epsilon \geq 0$ there exists $\mathbf{v} \in V$ such that

$$f(\mathbf{v}) \leq \inf_{\mathbf{x} \in V} f(\mathbf{x}) + \epsilon,$$

and

$$|Df(\mathbf{v})\mathbf{x}| \leq \epsilon \|\mathbf{x}\|, \text{ for all } \mathbf{x} \in \mathbb{R}^n .$$

Proof: Choose $\mathbf{v}$ as in Corollary (2), then for $t > 0$ and $\mathbf{x} \in \mathbb{R}^n$ we have

$$f(\mathbf{v} + t\mathbf{x}) - f(\mathbf{v}) = tDf(\mathbf{v})(\mathbf{x}) + \mathbf{r}(\mathbf{v} + t\mathbf{x})t\|\mathbf{x}\| \geq -\epsilon t\|x\|.$$

That is, $Df(\mathbf{v})(\mathbf{x}) + \mathbf{r}(\mathbf{v} + t\mathbf{x})\|\mathbf{x}\| \geq -\epsilon\|\mathbf{x}\|$, let $t \to 0$, so $\mathbf{r}(\mathbf{v} + t\mathbf{x}) \to 0$, to obtain

$$Df(\mathbf{v})(\mathbf{x}) \geq -\epsilon\|\mathbf{x}\| \text{ for all } \mathbf{x} .$$

Substituting $-\mathbf{x}$ for $\mathbf{x}$ leads to

$$-\epsilon\|\mathbf{x}\| \leq Df(\mathbf{v})(\mathbf{x}) \leq \epsilon\|\mathbf{x}\| .$$

Note: Expressing Df as the Jacobian, we have

$$\left| \left[\frac{\partial f}{\partial x_1}(\mathbf{v}), \frac{\partial f}{\partial x_2}(\mathbf{v}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{v}) \right] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \right| \leq \epsilon \; \|\mathbf{x}\| .$$

Taking $\mathbf{x} = \mathbf{e}_i$ gives $\left| \frac{\partial f}{\partial x_i}(\mathbf{v}) \right| < \epsilon$ for all i.

If f achieves a minimum at $\mathbf{v}$ we may use $\epsilon = 0$ in the above argument to obtain the well known result:

Proposition 4 If f is differentiable on V and achieves a minimum at $\mathbf{v} \in V$ then $Df(\mathbf{v} = 0$.

In particular $\frac{\partial f}{\partial x_i}(\mathbf{v}) = 0$ for all i.

Proof: Immediate.

Remark: Proposition 4 gives us an often used necessary condition for a minimum of a differentiable function. The numerical stability of this necessary condition is ensured by Corollary 3. That is, among the points where the derivative is almost zero there are points at which the minimum is almost achieved.

This leaves us with the practical problem of finding, at least approximately, the minimum of

$$f : \mathbb{R}^n \to \mathbb{R}$$

One approach to this is via what are known as *gradient methods*, we will consider the simplest of these. The **method of steepest descent** consists of searching for the minimum by moving on the 'surface'

$$z = f(\mathbf{x}) \ ,$$

initially 'downhill' as steeply as possible and then continuing in the same direction until we cease to go down, at which point a new initial direction is chosen.

The direction of *steepest descent* from $\mathbf{x}$ is the projection,

$$\mathbf{d} := -Df\big|_{\mathbf{x}} = \left(-\frac{\partial f}{\partial x_1}, \dots, -\frac{\partial f}{\partial x_n}\right),$$

of the 'downward' normal to the surface at $\mathbf{x}$, that is: perpendicular to the level curve (contour line) for the surface through $\mathbf{x}$, see diagram.

Thus from $\mathbf{x}$ we move to the point

$$\mathbf{x}^* := \mathbf{x} + t^*\mathbf{d} \ ,$$

where t^* is the smallest value of $t > 0$ at which the function

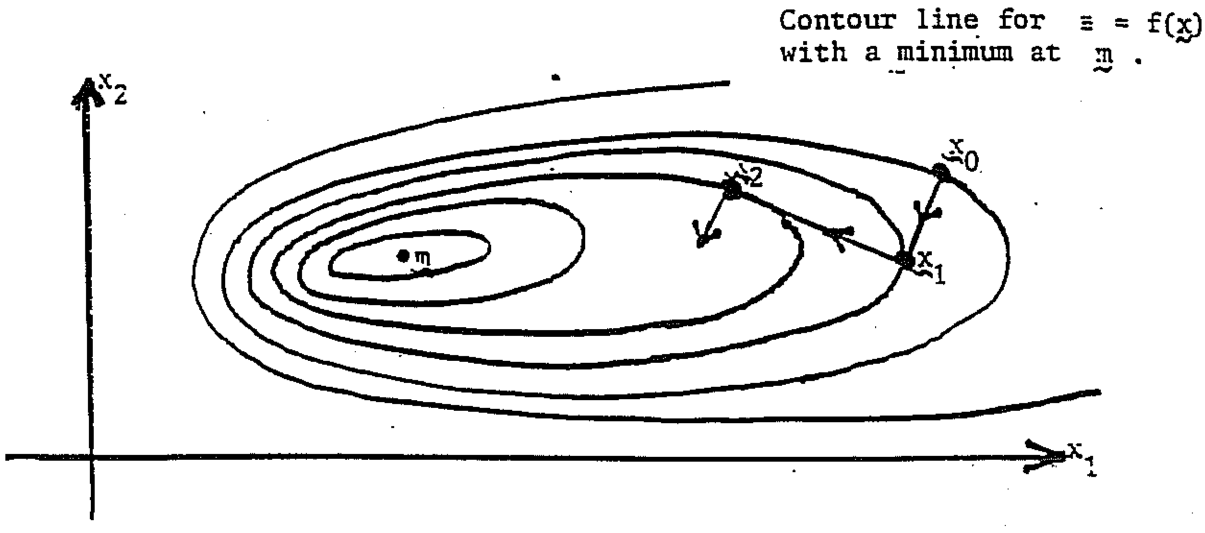
$$f^*(t) := f(\mathbf{x} + t\mathbf{d})$$

has a (local) minimum.

Starting with any point $\mathbf{x}_0$ this leads to a sequence of successive approximations for the minimum's location;

$$\mathbf{x}_0 \ , \ \mathbf{x}_1 \ , \mathbf{x}_2 \ , \ \dots \ , \ \mathbf{x}_n \ , \ \dots \ ,$$

where $\mathbf{x}_{n+1} = \mathbf{x}_n^*$.



Example: To illustrate, let us compute a couple of successive approximations to the minimum [−3 at (1, 1)] for the function

$$f(x,y) := 2x^2 + y^2 - 4x - 2y \ ,$$

starting at the origin, $\mathbf{x}_0 = (0,0)$.

$$\text{At} \quad \mathbf{x}_0; \quad \mathbf{d} = (-4x+4, -2y+2)\Big|_{(0,0)}$$
$$= (4,2),$$
$$f^*(t) = f(4t, 2t)$$
$$= 36t^2 - 20t,$$

which has a minimum at $t^* = \frac{5}{18}$.

Thus, $\mathbf{x}_1 = (\frac{10}{9}\ ,\ \frac{5}{9}) \doteq (1.11,\ 0.56)$.

At $\mathbf{x}_1$;

$$\mathbf{d} \doteq (-0.44,\ 0.88)$$
$$f^*(t) \doteq f(1.11 - 0.44t,\quad 0.56 + 0.88t)$$
$$\doteq 1.16t^2 - 0.97t - 2.8$$

which has a minimum at $t^* = 0.42$.

Thus, $x_2 = (0.93, 0.93)$ at which the function value is -2.984.

Thus we see that after two iterations we have located the minimum to within reasonable accuracy.

The method of steepest descent leads at each iteration to the problem of minimizing a function of one variable

$$f^* : \mathbb{R} \to \mathbb{R} .$$

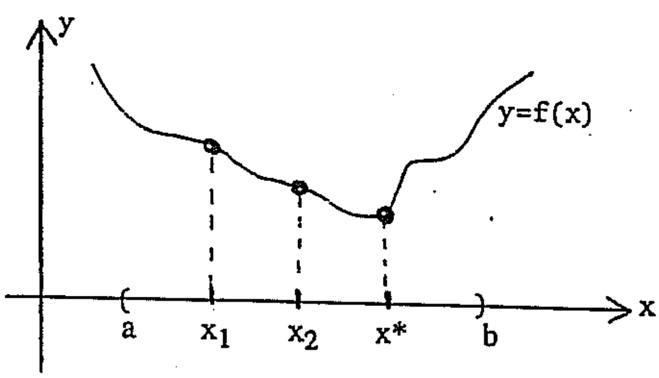
When the form of f^* is complicated we may need to seek its minimum numerically. We close with a consideration of this problem.

Suppose $f : \mathbb{R} \to \mathbb{R}$ has a unique minimum in the interval (a, b), at x^*, and is *unimodal*; that is,

$x_1 < x_2 < x^*$, or $x^* < x_2 < x_1$, implies

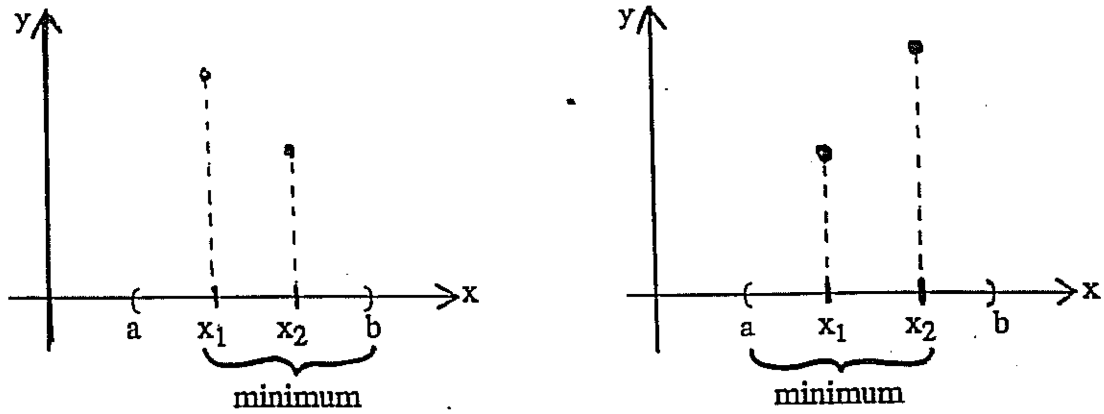
$$f(x_1) > f(x_2) > f(x^*) .$$

Every strictly convex function is unimodal, but not conversely - see sketch.



A unimodal function.

Our aim is to locate x^*, at least to within a given accuracy. Vital to accomplishing this is the observation that if $a < x_1 < x_2 < b$ then $x^* > x_1$ if $f(x_1) > f(x_2)$, and $x^* < x_2$ if $f(x_1) < f(x_2)$.

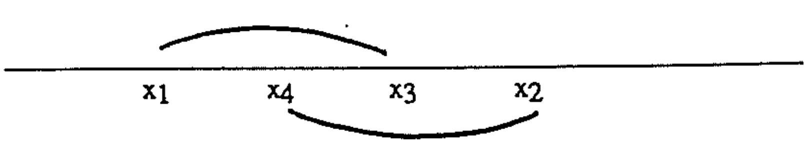


One approach is to evaluate f at $x_1 = a + \frac{1}{3}(b-a)$ and $x_2 = a + \frac{2}{3}(b-a)$ thereby locating x^* to within one of the intervals $(a, x_2]$ or $[x_1, b)$ both of which have length $\frac{2}{3}(b-a)$. We can now evaluate f one-third and two-thirds the way along the new interval and so locate x^* to within an interval of length $\left(\frac{2}{3}\right)^2(b-a)$.

Repeating this procedure we can locate x^* to within an interval of length $\left(\frac{2}{3}\right)^2(b-a)$ using $2n$ evaluations of f.

While this is an easily implemented (often adequate) algorithm it is possible to do better; that is, to achieve comparable accuracy with fewer evaluations. Each iteration of the above algorithm requires two new evaluations of the function. It would be 'nice' to use one of the last pair of evaluations as one of the values used at the next step, this is achieved with a **Fibonacci search.**

A key observation is that if the minimum of f is known to lie between x_1 and x_2, and the value of f is known at x_3, with $x_1 < x_3 < x_2$, then evaluating f at a fourth point x_4 located 'symmetrically' so that $x_2 - x_4 = x_3 - x_1$.



enables us to identify the minimum as lying between x_1 and x_3 or between x_4 and x_2, and so to within an accuracy of $x_3 - x_1 (= x_2 - x_4)$.

Further, as a moment's thought will show, any other location for x_4 may result in the minimum being located less precisely. Also note that the whole procedure is only possible provided x_3 is not the midpoint of x_1 and x_2.

We could now repeat the procedure starting with the three points $x_1 < x_4 < x_3$, or $x_4 < x_3 < x_2$, whichever is relevant, and locating a new point x_5.

Following this strategy with $x_1 = a$ and $x_2 = b$, and letting

$$\begin{aligned} \ell_1 &:= x_2 - x_1 \\ \ell_2 &:= x_3 - x_1 (= x_2 - x_4) \\ \ell_3 &:= x_4 - x_1 (= x_3 - x_5) \\ &\quad \text{or} \\ &\quad x_2 - x_3 (= x_5 - x_4), \end{aligned}$$

we seek a decreasing of lengths $\ell_1 > \ell_2 > \ell_3 > \ldots > \ell n > 0$ such that

$$\ell_k = \ell_{k+1} + \ell_{k+2} \quad (\text{k=1}, \ldots, n-2)$$

and $\ell_n = \frac{1}{2}\ell_{n-1}$, so that the process stops at the n'th step with the minimum of f located to within an interval of length ℓ_{n-1}.

Working backward, with $\ell_n := \delta$, we have

$$\begin{aligned} \ell_{n-1} &= 2\delta \\ \ell_{n-2} &= \ell_{n-1} + \ell_n = 2\delta + \delta = 3\delta \\ \ell_{n-3} &= \ell_{n-2} + \ell_{n-1} = 3\delta + 2\delta = 5\delta \\ \ell_{n-4} &= \ell_{n-3} + \ell_{n-2} = 5\delta + 3\delta = 8\delta \\ &\ldots\ldots \\ \ell_{n-k} &= \ell_{n-k+1} + \ell_{n-k+2} = F_k\delta + F_{k-1}\delta \\ &= F_{k+1}\delta \\ &\ldots\ldots \\ \ell_1 &= F_n\delta, \end{aligned}$$

where $F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8, F_6 = 13, F_7 = 21, F_8 = 34,$ $F_9 = 55, \ldots$ is the celebrated Fibonacci sequence (satisfying $F_{k+1} = F_k + F_{k-1}$).

From this we see that

$$\delta = \frac{b-a}{F_n} .$$

That is, to locate the minimum of f to within $2(b-a)/F_n$ using $n-1$ evaluations of f we should:

(1) Evaluate f at $a + \frac{F_{n-1}}{F_n}(b-a)$ and the symmetrically placed point in (a, b);

(2) Pick out the new interval within which the minimum must lie, and evaluate f at the point symmetrically placed in this interval with respect to the interior point already used;

(3) Repeat (2) until it terminates.

As an example, suppose we wish to locate the minimum of $f(x) = 2x^4 - x$ which is known to lie between 0 and 1.

To confine the minimum to an interval at length at most 0.05 we use $n = 9$, giving an uncertainty of $2/55 = 0.036$.
The resulting iterations are set out below. Numbers in brackets are the value of f at the relevant point.

(1)	x_1 0		x_3 34/55 (−0.326)	x_2 1
		x_4 21/55 (−0.339)		
(2)	x_1 0		x_4 21/55 (−0.339)	x_3 34/55
		x_5 13/55 (−0.230)		
(3)	x_5 13/55	x_4 21/55 (−0.339)		x_3 34/55
			x_6 26/55 (−0.373)	
(4)	x_4 21/55	x_6 26/55 (−0.3728)		x_3 34/55
			x_7 29/55 (−0.3727)	
(5)	x_4 21/55		x_6 26/55 (−0.373)	x_7 29/55
		x_8 24/55 (−0.364)		
(6)	x_8 24/55	x_6 26/55 (−0.373)		x_7 29/55
			x_9 27/55 (−0.375)	
(7)	x_6 26/55	x_9 27/55 (−0.374755)		x_7 29/55
			x_{10} 28/55 (−0.374749)	

From which we conclude that the minimum occurs between $x = 26/55 = 0.491$ and $x = 28/55 = 0.509$.
Of course, the actual minimum in this case is at $x = 1/2$.

Exercises

(1) Using graphical methods solve the following non linear optimization problems.

(a) Minimize: $x^2 + y^2$
Subject to: $1 - y \leq 0$

(b) Minimize: y
Subject to: $x^2 - y^3 \leq 0$

(c) Minimize: $x^2 - 3xy + y^2$
Subject to: $xy \leq 0$

(d) Minimize: x^2y
Subject to: $-y \leq 0$

(e) Minimize: xyz
Subject to: $x^2 + y^2 + z^2 - 3 \leq 0$, $x \geq 0$, $y \geq 0$, $z \geq 0$.

[Remark: this is the problem of finding the box of largest volume centred at the origin which can be fitted inside the sphere of radius $\sqrt{3}$ about the origin.]

(2) In (a) and (c) of (1) find values of μ and λ so that the functions

$$\mu(x^2 + y^2) + \lambda(1 - y)$$

and

$$\mu(x^2 - 3xy + y^2) + \lambda xy$$

respectively have a global minimum of the same value and at the same place as the minimum of the original constrained problems.

(3) Using the convex Multiplier Rule (and if applicable the Karush, Kuhn-Tucker condition) locate optimal solutions to 1(a) and (c).

(4) Show that the function $f(x, y) = x^2 + y^2 - xy$ is strictly convex.

(5) By considering the matrix $M = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$ show that the condition

$$[m_{11}], \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}, \ldots, M$$

have positive determinants is not sufficient to ensure that the quadratic form

$$\mathbf{x}^T M \mathbf{x}$$

is positive semi-definite.

(6) Solve the following problem.

Minimize: $x + y^2 - 2y$

Subject to: $-1 + x^2 \leq 0$

$y \leq 0$

(7) Starting at the point (1,1) use the method of steepest descent to find 3 sucessive approximations to the minimum of $f(x, y) = 9x^2 - 8xy + 3y^2$.

(8) In the method of steepest descent, show that at each change in the direction of descent the new direction is orthogonal to the old one.

(9) Starting at the initial point (1, 2, 2) use the method of steepest descent to find 2 further successive approximations to the maximum of

$$f(x, y, z) = 4x - x^2 + 9y - y^2 + 10z - 2z^2 - 1/2yz.$$

(10) The function

$$f(x) = (3 - 4x)/(1 + x^2)$$

is known to have its minimum between 1 and 6.5 Use a Finonacci search to locate the minimum to an accuracy of ± 0.1.

APPENDIX – Suprema and Infima

Let S be a non-empty subset of the real numbers $\mathbb{R}$. We say S is **bounded below** if there exists a real number α such that $\alpha \leq s$, for all $s \in S$, and refer to α as a **lower bound** for S.

The **infimum** of S, denoted by $\inf S$, is the 'greatest lower bound' for S. That is $M = \inf S$ if and only if

(i) $M \leq s$ for all $s \in S$ (M is a lower bound for S) and

(ii) if α is a lower bound for S, then $\alpha \leq M$ (M is the *greatest* lower bound for S).

It is a fundamental property of $\mathbb{R}$ that every non-empty subset S which is bounded below has an infimum.

Note: The infimum of S need not belong to S. For example $\inf(0,1) = 0$ but $0 \notin (0,1)$. If it hapens that $\inf S$ is a member of S we usually refer to it as the **minimum** of S and denote it by $\min S$.

It will be convenient to write $\inf S = -\infty$ in the case when S is not bounded below.

The following simple result is assumed frequently in these notes.

Proposition: For $\emptyset \neq S \subseteq \mathbb{R}$ *there exists a sequence* (s_n) *of points of* S *with* $s_n \to \inf S$ *(from above).*

Note: In case $\inf S = -\infty$, this must be interpreted as 's_n diverges to $-\infty$'; that is, given any real number r, there exists $N \in \mathbb{N}$ such that $n > N \Rightarrow s_n < r$.

Proof: In case $M = \inf S > -\infty$, for each $n \in \mathbb{N}$ there must exist a point of S, call it s_n, with $s_n \leq M + \frac{1}{n}$ (otherwise $M + \frac{1}{n}$ would be a lower bound for S, contradicting the fact that M is the greatest lower bound). But then, since M is a lower bound, $M \leq s_n \leq M + \frac{1}{n}$ and so as $n \to \infty$ we have $s_n \to M$. The proof in case $\inf S = -\infty$ is similar and is left as an exercise.

□

The **supremum** (or least upper bound) of S may be defined and analysed similarly. Alternatively questions concerning suprema may be converted into questions about infima by noting that

$$\sup S = -\inf\{-s : s \in S\}.$$

THE UNIVERSITY OF NEWCASTLE

NEW SOUTH WALES

EXAMINATION

November, 1990

MATHEMATICS 314

Optimization

Time allowed – 2 hours

Total number of questions – 5

All questions may be attempted

All questions carry the same number of marks

Hand calculators are allowed

1. (a) Define concisely what is meant by each of the following:

   (i) A *closed* set.

   (ii) A *compact* set.

   (iii) A *convex* set.

   (iv) An *affine* set.

   (v) A *hyperplane* in $\mathbb{R}^n$.

   (vi) An *interior point* in a subset of $\mathbb{R}^n$.

   (vii) An *extreme point* of a convex set.

   (viii) An *upper semi–continuous* function.

   (ix) A *convex* function.

   (x) A *feasible solution* for the general optimization problem:

   $$\begin{array}{ll} \textit{minimize:} & f(x) \\ \textit{subject to:} & x \in C. \end{array}$$

   (b) Let $f: C \to \mathbb{R}$ be a strictly convex function on the nonempty closed bounded convex subset C of $\mathbb{R}^n$. What can you say about the problems of maximizing and minimizing f on C?
   [State clearly any theorems you have appealed to in giving your answer.]

2. (a) Show that

   (i) the union, and

   (ii) the sum

   of two compact sets is compact.

   (b) Let K be a nonempty compact subset of $\mathbb{R}^n$. For $x \in K$ show that there exists a point $k_x \in K$ such that

   $$\| x - k_x \| \le \| x - k \|, \quad \forall\, k \in K.$$

SEE OVER

THE UNIVERSITY OF NEWCASTLE

NEW SOUTH WALES

EXAMINATION

June 1991

MATH314

OPTIMIZATION

Time allowed – 2 hours

Total number of questions – 5

All questions may be attempted

All questions carry the same number of marks

Hand calculators are allowed

1. (a) Briefly describe what is meant by each of the following.

   (i) *a bounded* subset of $\mathbb{R}^n$.

   (ii) $\bar{A}$, the *closure* of the set $A \subseteq \mathbb{R}^n$.

   (iii) The *boundary* of $A \subseteq \mathbb{R}^n$.

   (iv) a *lower semi-continuous* function $f : \mathbb{R}^n \to \mathbb{R}$.

   (v) co(A), the *convex hull* of $A \subseteq \mathbb{R}^n$.

   (vi) The *dimension* of a convex subset of $\mathbb{R}^n$.

   (vii) Given a closed bounded convex set A, a point $\underset{\sim}{x} \notin A$ can be *separated* from A by a hyperplane.

   (viii) The *epigraph* of a convex function.

   (ix) A *strictly convex* function $f : C \to \mathbb{R}$, where C is a convex subset of $\mathbb{R}^n$.

   (x) a *feasible solution* for the optimization problem.

   minimize : $f(\underset{\sim}{x})$

   subject to : $\underset{\sim}{x} \in C \subseteq \mathbb{R}^n$.

   (b) Let $\mathbb{P}$ be a linear programming problem with a nonempty bounded constraint set. Starting with a basic feasible solution each application of Bland's modified simplex algorithm yields another basic feasible solution, distinct from previous ones (no cycling), at which the value of the objective function is no greater than its value at the previous one.

   Assuming the above, give a concise explanation of why repeated application of Bland's modified simplex algorithm must lead to a solution of $\mathbb{P}$. (State clearly any results you are assuming.

SEE OVER............2

THE UNIVERSITY OF NEWCASTLE

NEW SOUTH WALES

EXAMINATION

June 1992

MATH314

OPTIMIZATION

Time allowed – 2 hours

Total number of questions – 5

All questions may be attempted

All questions carry the same number of marks

Hand calculators are allowed

1. (a) State concisely what is meant by each of the following.

   (i) a *compact* set.

   (ii) a *lower semi–continuous* function.

   (iii) A *boundary point* of a subset of $\mathbb{R}^n$.

   (iv) The *dimension* of a convex subset of $\mathbb{R}^n$.

   (v) The point $\underset{\sim}{a}$ is *separated* from the convex set B by the *hyperplane* H.

   (vi) The *epigraph* of a convex function f.

   (vii) A *linear inequality constraint.*

   (viii) The *set of feasible solutions* for the optimization problem:

   minimize: $f(\underset{\sim}{x})$

   subject to: $\underset{\sim}{x} \in C$.

   (b) (i) State the fundamental theorem of linear programming.

   (ii) Very briefly discuss the phenomenon of cycling and how it may be prevented.

   (iii) Suppose $f_0, f_1, \ldots, f_m$ are convex functions from $\mathbb{R}^n$ to $\mathbb{R}$. What can you say about optimum solutions of the problem:

   minimize: $f_0(\underset{\sim}{x})$

   subject to: $f_1(\underset{\sim}{x}) \le 0$

   $\ldots$

   $f_m(\underset{\sim}{x}) \le 0$?

4. (a) Let $\mathcal{P}$ be the linear programming problem:

$$\begin{aligned} \text{maximize:} \quad & x_1 + x_2 \\ \text{subject to:} \quad & 3x_1 + x_2 \le 3 \\ & 2x_1 + x_2 \le 2 \\ & x_1 + 2x_2 \ge 3 \\ & x_1, x_2 \ge 0 . \end{aligned}$$

(i) Express $\mathcal{P}$ in a standard form.

State the dual problem $\mathcal{P}^*$.

Write down a tableau representing both the problem $\mathcal{P}$ and its dual $\mathcal{P}^*$.

(ii) Illustrate the simplex algorithm by using it to solve both the problem $\mathcal{P}$ and its dual $\mathcal{P}^*$.

[Marks will **not** be awarded for solutions obtained other than by the simplex algorithm.]

(b) Suppose

(T)

$x_{k_1} \dots x_{k_n}$	-1	
A	$\underset{\sim}{b}$	$-x_{k_{n+1}}$ ⋮ $-x_{k_{n+m}}$
$\underset{\sim}{c}$	d	f

is a feasible tableau for some linear programming problem.

Let $\underset{\sim}{x}_i = (\underset{\sim}{0}, \underset{\sim}{b}) \in \mathbb{R}^{n+m}$ be the corresponding basic feasible solution.

Let $\underset{\sim}{x}_f$ be the basic feasible solution corresponding to the tableau obtained from (T) by pivoting about a point chosen according to the requirements of the stage II simplex algorithm.

Show that $f(\underset{\sim}{x}_f) \le f(\underset{\sim}{x}_i)$.

.... 5.

5. (a) State the John multiplier rule and illustrate its use by showing that *the minimum* of

$$f(x, y) := \det \begin{pmatrix} a & b \\ x & y \end{pmatrix}$$

$$\textit{subject to } x^2 + y^2 \le h$$

is $-\sqrt{h(a^2 + b^2)}$, where a and b are fixed real numbers not both zero.

[Remark: By symmetry this establishes

$$\left| \det \begin{pmatrix} a & b \\ x & y \end{pmatrix} \right| \le \sqrt{(x^2+y^2)(a^2+b^2)} .$$

This is *Hadamard's inequality* in the special case of 2×2 matrices.]

(b) For the unconstrained minimization problem:

$$\text{minimize:} \quad x^4 + y^4 + x^2y^2 - x,$$

show that starting from the point $\underset{\sim}{x}_0 = (0, 0)$ the initial iteration for the method of steapest descent leads to the problem of minimizing

$$f^*(t) = t^4 - t.$$

Noting that the minimum lies in the interval [0, 1], illustrate the method of Fibbonacci search by using it to locate the minimum of f^* to within an accuracy of 0.25.

THE UNIVERSITY OF NEWCASTLE

NEW SOUTH WALES

EXAMINATION

June 1994

MATH314

OPTIMIZATION

Time allowed – 2 hours

Total number of questions – 5

All questions may be attempted

All questions carry the same number of marks

Hand calculators are allowed

1. (a) Define concisely what is meant by each of the following.

   (i) a *compact* subset of $\mathbb{R}^n$.

   (ii) a *lower semi–continuous* function $f: \mathbb{R}^n \to \mathbb{R}$.

   (iii) A *feasible solutions* for the optimization problem:

   $$\text{minimize: } f(\underset{\sim}{x}) \qquad \text{subject to: } \underset{\sim}{x} \in C.$$

   (b) In the following statement explain what is meant by each of the boldface terms

   There is a **supporting hyperplane** at each **boundary point** of a closed bounded convex subset of $\mathbb{R}^n$.

   (c) If C is a non–empty closed bounded convex subset of $\mathbb{R}^n$ and $f: C \to \mathbb{R}$ is a continuous concave function what can be said about the optimization problem

   $$\text{minimize: } f(\underset{\sim}{x}) \qquad \text{subject to: } \underset{\sim}{x} \in C.$$

2. (a) Let C be a closed subset of $\mathbb{R}^n$ and let K be a compact subset of $\mathbb{R}^n$, show that $C \cap K$ is a compact set.

   (b) Prove that a finite subset of $\mathbb{R}^n$ is compact.

   (c) Let A be a closed bounded subset of $\mathbb{R}^n$ and let $\underset{\sim}{x}_0$ be a point of $\mathbb{R}^n$ which is not in A. Prove that there exists a closest point of A to $\underset{\sim}{x}_0$.

3. (a) Prove that the ball $B_r[\underset{\sim}{x}]$, where $r > 0$ and $\underset{\sim}{x} \in \mathbb{R}^n$, is a convex set.

   (b) Let C be a nonempty convex subset.

   (i) For $\underset{\sim}{x}_0 \in C$, $B \subseteq C$, and $\lambda \in [0, 1]$ show that

   $$\lambda \underset{\sim}{x}_0 + (1-\lambda)B \subseteq C.$$

   (ii) For $\underset{\sim}{x}_0 \in C$, $\underset{\sim}{x}_1$ an interior point of C, and $\lambda \in [0, 1]$ show that

   $$\lambda \underset{\sim}{x}_0 + (1-\lambda)\underset{\sim}{x}_1 \in \text{int } C.$$

   [If it helps; you may assume that $B_r(\underset{\sim}{x}_1) = \underset{\sim}{x}_1 + B_r(\underset{\sim}{0})$.]

   (c) Let C be a convex subset of $\mathbb{R}^n$ and let $f: C \to \mathbb{R}$ be a convex function. If f achieves its maximum on C at a unique point $\underset{\sim}{x}_0$ show that $\underset{\sim}{x}_0$ is an extreme point of C.

SEE OVER ....2.

4. (a) Illustrate the simplex algorithm by using it to solve the following linear programming problem and its dual.

maximize: $x_1 - 2x_2$

subject to: $2x_1 + x_2 \geq 4$

$x_1 + 2x_2 \leq 4$

$x_1 \geq 0,\ x_2 \geq 0$.

[**Note:** No marks will be awarded for a solution obtained by a method other than the simplex algorithm].

(b) Let $C = \{\underset{\sim}{x} = (x_1, x_2) : x_1 \geq 0,\ x_2 \geq 0 \text{ and } x_1 + x_2 = 1\}$.

Reduce the optimization problem;

find $\underset{\sim}{a} \in C$ at which

$$\max \ \min_{\underset{\sim}{b} \in C} \ E(\underset{\sim}{a}, \underset{\sim}{b})$$

occurs, where $E(\underset{\sim}{a}, \underset{\sim}{b}) = 5a_1b_1 - 10a_2b_2$,

to a linear programming problem.
State clearly any results you are assuming.

(a) Briefly discuss the phenomenon of *cycling* in the simplex algorithm and how it can be avoided.

(b) (i) Show that the function $f(x, y) = (x-1)^2 + y$ is convex.

(ii) Illustrate the convex multiplier rule, and if appropriate the Karush, Kuhn–Tucker condition to investigate the non–linear optimization problem

minimize : $-(x+y)$

subject to: $(x-1)^2 + y \leq 0$.

(c) Solve the following *Linear Programming problem* by using the John Multiplier rule.

minimize : $-y$

subject to: $x + y - 1 \leq 0$

$-x \leq 0$

(d) For the unconstrained minimization problem

minimize : $x - (x+y)e^{-(x+y)}$,

show that the initial iteration for the method of steepest descent starting from (0, 1) leads to the problem of minimizing

$$f^*(t) := (t-1)e^{(t-1)} - t.$$

Illustrate the method of Fibbonacci search by using it to locate the minimum of f^* in the interval [0, 2] to within an accuracy of 0.5.