CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

1

Approaching optimization by means of examples

1.1 Definition of optimization problem

The subjects we want to consider are centred around the problem of finding the *infimum*, or if it exists, the *minimum*, of a given real-valued function f on a given set G of a given space X. Hence if we let $f: x \in X \mapsto f(x) \in R$ and define α by

(1.1.1) $$\alpha = \inf\{f(x) : x \in G\}, \quad G \subset X,$$

then the underlying problem is that of finding the value of α, or, if it exists, an $x_o \in G$ such that $f(x_o) = \alpha$. We shall refer to this basic problem as an *optimization problem*. Usually X is called the *space of decisions* or *decision space*, and $x \in X$ is called the *decision variable*, although when $x = (\xi_1, \ldots, \xi_n) \in R_n$, the components of x are also termed decision variables, so that this term is slightly ambiguous.

Understandably, in practical situations we want an $x_o \in G$ such that $f(x_o) = \alpha$. Such an x_o is called an *optimal solution*, or simply a *solution*, whereas any x in G is called a *feasible solution*. We say that our optimization problem is *solvable* if at least one optimal solution exists, and that it is *feasible* if at least one feasible solution exists.

The function f is the *objective function*. As stated before, its values are real numbers. In a more general type of optimization problem, however, where $f(x)$ is not a real number but, say, an s-dimensional vector, it is not surprising that f is then termed the *multi-objective function*. A fair question now would be how to define an optimization problem with many objectives. In order to achieve this, one must generalize the idea of infimum. We shall come back to this more

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

general question in 7.5 but it will only be touched on. See also Example 1.2.9.

The set G is the *constraint set* or *feasible region*. A very special example is where $G = X$, in which case we speak of an *unconstrained* problem. This is a natural terminology, since the larger G is, the fewer restrictions there are. As a second example, let functions $g_i\colon X \to R$, $i = 1, \dots, p$ and $h_j\colon X \to R$, $j = 1, \dots, q$ be given and let G be defined by

(1.1.2) $$G = \{x\colon g_i(x) \leqslant 0,\ i = 1, \dots, p, \quad h_j(x) = 0,\ j = 1, \dots, q\}.$$

In particular, if the g_i are absent, we arrive at the abstract formulation of many classical optimization problems in mechanics and other areas, problems which are often solved by the method of Lagrange multipliers. The subject of this book is nothing more than the generalization of ideas and results involving these multipliers!

Each separate condition $g_i(x) \leqslant 0$ or $h_j(x) = 0$ is called a *constraint*, or more specifically an *inequality constraint* or an *equality constraint*. Defining $g\colon X \to R_p$ and $h\colon X \to R_q$ by $g(x) = (g_1(x), \dots, g_p(x))$ and $h(x) = (h_1(x), \dots, h_q(x))$, and writing $g_i(x) \leqslant 0$, $i = 1, \dots, p$, as $g(x) \leqslant 0$, which is now an inequality relation between vectors, we can further write

(1.1.3) $$G = \{x\colon g(x) \leqslant 0, \quad h(x) = 0\}.$$

Somewhat inconsistently, $g(x) \leqslant 0$ and $h(x) = 0$ are also often called *constraints*, although strictly speaking each of these consists of a number of constraints.

Instead of (1.1.3) we frequently come across the following form of G,

(1.1.4) $$G = \{x\colon g(x) \leqslant 0, \quad h(x) = 0, \quad x \in C\},$$

where C is some given set of X. The main reason for slipping in another set is not so much that it would be impossible to define G completely in terms of inequality and equality constraints alone, but that we want to treat the constraints $g(x) \leqslant 0$ and $h(x) = 0$ differently from the constraint $x \in C$. In fact we shall not be too concerned about the nature of C, although general conditions (such as being convex) may be imposed on it.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

Instead of an infimum, we may want to find a *supremum*. Obviously, all we have said so far can be carried over to supremum problems, if only because of the following trivial relation

(1.1.5) $$\sup\{f(x)\colon x\in G\} = -\inf\{-f(x)\colon x\in G\}.$$

In practical situations X is often a Euclidean space or more generally a Banach space. Many of the theorems that follow, however, hold if X is a locally convex topological vector space. These more general spaces can be used fruitfully in a number of cases where the norm topology of a Banach space is replaced by a weaker one. The latter may be necessary in order that the dual space is a convenient one.

The variety of optimization problems in areas such as mathematics, engineering, economics and statistics is enormous. In order that the reader gets at least some idea of this variety, and to illustrate the meaning of the definitions given so far, we list a few of them in the next section.

1.2 Examples of optimization problems

1.2.1 **Example.** Given a closed set in a Euclidean space, find the least distance from the origin to that set. For example, if $x\in R_n$, A is an $m\times n$ matrix, $b\in R_m$ and $|Ax-b|$ is the Euclidean norm of $Ax-b$, find

$$\inf\{|Ax-b|\colon x\in R_n\}.$$

Then the said Euclidean space is R_m, and the closed set is $\{Ax-b\colon x\in R_n\}$. This is just the simple least squares approximation. If we replace the Euclidean spaces by Banach spaces this example becomes what is known as the *minimum norm problem*. This is further discussed in Example 3.14.19.

1.2.2 **Example.** The *Chebyshev approximation problem* is concerned with approximating some function q by means of a given finite set of functions by minimizing the maximum of the absolute value of the difference of $q(t)$ and a linear combination of the given functions. As an example let $q\colon[-1,+1]\to R$ and let the given functions be $1, t, \ldots, t^n$, $t\in[-1,+1]$. Then the problem is

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

to find

$$\inf\{f(x) \colon x \in R_{n+1}\},$$

where $x = (\xi_0, \ldots, \xi_n)$ and

$$f(x) = \max\left\{\left|q(t) - \sum_{i=0}^{n} \xi_i t^i\right| \colon -1 \leqslant t \leqslant +1\right\}.$$

Both this example and the previous one are examples of unconstrained optimization problems. The Chebyshev approximation problem is worked out in 7.2.

1.2.3 **Example.** Let $a_1, \ldots, a_n$ be given positive numbers whose sum is 1. Find

$$\sup\{\xi_1^{a_1} \ldots \xi_n^{a_n} \colon a_1\xi_1 + \ldots + a_n\xi_n = 1 \xi_i \geqslant 0, \quad i = 1, \ldots, n\}.$$

As will be shown in 7.1, the optimal solution is given by $\xi_i = 1$ for all i, which implies that

$$\xi_1^{a_1} \ldots \xi_n^{a_n} \leqslant a_1 + \ldots + a_n \xi_n \quad \text{for all} \quad \xi_i \geqslant 0, \quad i = 1, \ldots, n.$$

1.2.4 **Example.** Suppose a horizontal, rigid, homogeneous steel plate, which has the shape of a polygon with four vertices, is supported at these vertices. At a given point T of the plate a vertical force P is applied to it. This force is directed downwards. Obviously, P results in forces ξ_i acting on the supports, $i = 1, \ldots, 4$, and $P = \Sigma\xi_i$. The ith support remains rigid as long as $\xi_i \leqslant F_i$, where F_i is known, but suddenly breaks down as soon as $\xi_i > F_i$. Find the maximum P such that no support will break down. If we introduce a coordinate system in the plane of the plate such that $T = (0, 0)$ and if we let (a_i, b_i) be the ith vertex, and then observe the conditions for all forces to be in equilibrium, the problem becomes that of finding

$$\sup\{\Sigma\xi_i \colon \Sigma a_i \xi_i = 0, \quad \Sigma b_i \xi_i = 0, \quad \xi_i \leqslant F_i, \quad i = 1, \ldots, 4\}.$$

Notice that this problem has a very special form: everything is linear, moreover there are only a finite number of constraints, and $x = (\xi_1, \ldots, \xi_4)$ is finite dimensional. Problems of this type are termed *linear programming problems*, and are treated in 3.13.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

1.2.5 **Example.** Linear programming problems, as defined in the previous example, arise very often in economics. Consider the following so-called *production planning problem*. Let $x = (\xi_1, \ldots, \xi_n)$ represent the amounts ξ_j of n goods that must be produced, let $b = (\beta_1, \ldots, \beta_m)$ represent the maximal amounts β_i of m raw materials that may be used for producing the goods, and let the production of one unit of the jth good require the amounts α_{ij} of raw material i, $i = 1, \ldots, m$. The price at which the jth good can be sold on the market is γ_j per unit. Find an optimal production plan, that is find x, such that $\Sigma\gamma_j\xi_j$ is as large as possible; hence find

$$\sup\{\Sigma\gamma_j\xi_j: \Sigma\alpha_{ij}\xi_j \leqslant \beta_i, \quad i = 1, \ldots, m, \quad \xi_j \geqslant 0, \quad j = 1, \ldots, n\}.$$

or, putting $c = (\gamma_1, \ldots, \gamma_n)$ and letting α_{ij} be the ijth element of a matrix A, find

$$\sup\{cx: Ax \leqslant, x \geqslant 0\},$$

where, just as in (1.1.3), we have used inequality relations between vectors. Notice the very special kind of constraint $x \geqslant 0$, or $\xi_j \leqslant 0$, $j = 1, \ldots, n$. These are called *nonnegativity constraints*.

1.2.6 **Example.** So far our examples have involved *finite*-dimensional decision spaces only, which would give the wrong impression that we could restrict ourselves to Euclidean spaces. Here is a much more general example. Let t represent time and let a certain physical system (a rocket to one of the earth's satellites) be described adequately if its so-called *state* $x(t)$ is known at all relevant times. The state is thought of as a finite-dimensional vector, and might consist of position and velocity of the system. The system can be controlled by the *control* u, which like x is a function of t, and $u(t)$ is also taken to be a finite-dimensional vector. The meaning of the control is that if u is specified the state $x(t)$ is also fixed, or, in other words, that $x: t \mapsto x(t)$ can be solved uniquely from a certain set of equations $g(x, u) = 0$. The control might include the position of steering mechanisms, the rate of fuel consumption, and the like. It is quite probable

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

that limitations are set to u, as obviously follows from the examples suggested. Let us summarize those by $u \in G$. Finally, $f(x, u) \in R$ represents some measure of performance of the system, say the time required to reach a target, or the total fuel consumption. At any rate we assume that the system performs best if $f(x, u)$ is minimal. Then the problem is to find

$$\inf\{f(x, u) : g(x, u) = 0, \quad u \in G\}.$$

To be more specific, assume that $f(x, u)$ is some integral over a fixed period of time $[0, T]$, ruling out the possibility that $f(x, u)$ is the time to reach a target, and assume that x can be solved from an integral equation; then given T, G, x_0, ϕ and ψ the problem is to find

$$\inf\left\{\int_0^T \phi(x(t), u(t), t)\,\mathrm{d}t : x(t) = x_0 + \int_0^T \psi(x(t), u(t), t)\,\mathrm{d}t, u \in G\right\}.$$

This is a typical *fixed time optimal control problem*. A simpler problem is obtained if the integral equation, which is equivalent to a differential equation with initial conditions, is replaced by difference equations. A more complex problem arises if this differential equation is replaced by a *partial* differential equation.

The question of what the decision space should be can be answered in two ways. Either it is the space of the controls u, in which case the equation by which x can be solved in terms of u should not be considered a constraint, but merely a *defining equation*; or it is the product of the space of the controls and that of the states, in which case $g(x, u) = 0$ becomes a *constraint*. If x can be easily solved from this equation then the former approach looks better, but if solving for x is a difficult matter then it may be wise to follow the latter approach, where (x, u) becomes the decision variable, *not u*. A simple control problem is solved in 7.3.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

1.2.7 **Example.** Consider two persons, playing the following, perhaps not very exciting, *game*. Player i must select an element x_i from a given set G_i of, say, a Euclidean space X_i, $i = 1, 2$. They must do so independently from each other, that is to say the one player does not know what the other is selecting. Further, a function $\phi: X_1 \times X_2 \to R$ is given. After x_1 and x_2 have been chosen, player 1 must pay the amount $\phi(x_1, x_2)$ to player 2. End of the game!

Now what is the optimization problem here? Obviously player 1 wants to minimize $\phi(x_1, x_2)$, but at the same time player 2 wants to maximize this amount. So there are two conflicting objectives. Even if we forget about player 2 and consider only player 1's desire to minimize $\phi(x_1, x_2)$, there is the problem that he does not know anything about x_2. To resolve the difficulty, let us assume that player 1 adopts the following reasoning: if I select x_1, then the maximum loss I might incur is $\sup_{G_2} \phi(x_1, x_2)$, so let me select x_1 such that I minimize this supremum. This leads to the problem of finding

$$\inf_{G_1} \sup_{G_2} \phi(x_1, x_2),$$

where $\sup_{G_2} \phi(x_1, x_2)$ is the objective function. A similar reasoning in player 2's mind leads to the problem of finding

$$\sup_{G_2} \inf_{G_1} \phi(x_1, x_2),$$

where now $\inf_{G_1} \phi(x_1, x_2)$ is the objective function.

Hence we have managed to create *two* optimization problems. An interesting question in the *theory of games* is under what conditions a pair $(x_1^\circ, x_2^\circ) \in G_1 \times G_2$ exists such that

$$\sup_{G_2} \inf_{G_1} \phi(x_1, x_2) = \phi(x_1^\circ, x_2^\circ) = \inf_{G_1} \sup_{G_2} \phi(x_1, x_2)$$

or equivalently

$$\phi(x_1^\circ, x_2) \leqslant \phi(x_1^\circ, x_2^\circ) \leqslant \phi(x_1, x_2^\circ)$$

$$\text{for all } x_1 \in G_1 \text{ and all } x_2 \in G_2.$$

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

Because of the last relation such a pair is called a *saddle-point of* ϕ (with respect to G_1 and G_2).

Quite apart from the theory of games (a subject we shall not pursue very deeply) the existence of saddle-points will be one of the vital facts we shall be interested in. Then the function ϕ will be replaced by what will be termed the *Lagrangian function*.

Finally we remark that the theory of games encompasses a little more than just this very simple game!

1.2.8 **Example.** Let us now turn to a *stochastic* optimization problem. It is an example of *inventory control*, well known in economics. During a given number of periods a certain item has to be produced so as to satisfy a demand for it. The demand presents itself at the *end* of each period and is stochastic. If the demand at the end of any period turns out to be less than what is available at that moment then inventory costs are incurred, and if the demand is more than what is available, then shortage costs are incurred. Paying these costs does not mean that the surplus or the shortage disappears; on the contrary they accumulate. Apart from inventory costs and shortage costs we are faced with production costs. Assume for simplicity that only two periods are considered, that all costs are linear and the same for both periods, that c is the cost of production per unit, h is the inventory cost per unit and u is the shortage cost per unit. Further assume that at the beginning of period 1 there is neither a positive inventory, nor a positive shortage. Finally, and this is an important assumption, the production in the second period may depend on the actual demand in the first period, so that the problem is one with '*recourse*'.

The problem is how to produce in order that the *expected* total cost is minimized.

Let S be the set of possible demands, e.g. $S = [0, 10]$ or $[0, +\infty)$, etc. and let μ be the probability measure of the demand distribution. Define the function $(\cdot)^+$ by

$$r^+ = r \quad \text{if } r \geq 0, \quad r^+ = 0 \quad \text{if } r < 0, \quad r \in R.$$

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

Then the problem is to find

$$\inf\Big\{cx_1+\int_S [cx_2(s)+h(x_1-s)^+ + u(s-x_1)^+]\,\mu(\mathrm{d}s)$$
$$+\int_S\int_S [h(x_1+x_2(s)-s-s')^+ + u(s+s'-x_1-x_2(s))^+]\,\mu(\mathrm{d}s')\,\mu(\mathrm{d}s):$$
$$x_1 \geqslant 0,\quad x_2(s)\geqslant 0 \text{ for (almost) all } s\Big\}.$$

The decision variable is $x=(x_1,x_2)$ where $x_1 \in R$ but x_2 is a function on S to R. Hence X is an infinite-dimensional space, unless the given probability distribution is discrete and finite.

Usually this sort of problem is solved by *dynamic programming*, but in 7.4 we shall view the present problem from the standpoint taken in this book, that is from the standpoint of *mathematical programming* (see 2.9).

1.2.9 **Example.** This example, too, is stochastic in nature and arises in statistical decision theory. Let the possible outcomes of a random experiment form the set S. S is known, but the underlying probability distribution is not, although it is known that it is one out of M completely known distributions, which we indicate by $D_1,\ldots,D_M$. For simplicity we let $S=\{s_1,\ldots,s_N\}$, for some N; hence S is finite. With D indicating the unknown distribution, we let p_{mn} be the probability that the outcome is s_n when $D=D_m$, $n=1,\ldots,N$, $m=1,\ldots,M$. Clearly, all p_{mn} are known. Further let Q, $1\leqslant Q<M$ be given and consider two hypotheses H and K. H is the *null-hypothesis*: $D=D_m$ for some m satisfying $1\leqslant m\leqslant Q$; whereas K is the *alternative hypothesis*: $D=D_m$ for some m satisfying $Q+1\leqslant m\leqslant M$. Statisticians speak of making *an error of the first kind* if H is rejected, when it is in fact true, and of making *an error of the second kind* if H is accepted, when it is in fact false. They introduce a *testfunction* x from S to, say, the set $\{0,1\}$, to serve the following purpose. Suppose a single sample is drawn from S. If $x(s_n)=0$ then H is accepted, if $x(s_n)=1$ then H is rejected. In case $D=D_m$ *and* $1\leqslant m\leqslant Q$ then $\Sigma_{n=1}^{N} p_{mn}\,x(s_n)$ is the probability of making an error of the first kind. This probability is kept low by simply putting an upper bound, say 0.05, to it. And in case $D=D_m$ but

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521604915 - Approaches to the Theory of Optimization
J. Ponstein
Excerpt
More information

$Q+1 \leqslant m \leqslant M$ then $1-\Sigma_{n=1}^{N} p_{mn} x(s_n)$ is the probability of making an error of the second kind. Instead of putting bounds on this probability as well, it will be minimized. This would lead, however, to *multi-objective optimization* unless we take $Q = M-1$. With this further restriction we are thus led to the problem of finding

$$\sup \left\{ \sum_{n=1}^{N} p_{Mn} x(s_n) \colon \sum_{n=1}^{N} p_{mn} x(s_n) \leqslant 0.05, m = 1, \ldots, M-1; \right.$$
$$\left. x(s_n) = 0 \text{ or } 1, \quad n = 1, \ldots, N \right\}.$$

Since all relations are linear, except for the constraint $x(s_n) = 0$ or 1, this problem is an example of what is known as *integer* (*linear*) *programming*; 'integer' because the decision variables are required to be integer valued. Such problems are out of the scope of this book. If however, we replace x by a function from S to [0, 1], then the result is linear programming, which is one of the simplest applications of what is going to follow. Other variations are obtained by taking S and/or D infinite. Then the decision space X becomes infinite-dimensional or the number of constraints becomes infinite, or both. Yet linearity remains. In these cases one speaks of (*one-sided*) *infinite linear programming problems*, at least if $x\colon S \to [0, 1]$.

© Cambridge University Press www.cambridge.org