Gradient Optimization and Nonlinear Control

LAWRENCE HASDORFF

University of Houston
Clear Lake City, Texas

Library of Congress Cataloging in Publication Data:

Hasdorff, Lawrence, 1929-
Gradient optimization and nonlinear control.
"A Wiley-Interscience publication."
Includes bibliographical references and index.
1. Control theory. 2. Mathematical optimization.
3. Nonlinear theories. I. Title.
QA402.3.H35 629.8'36 75-40187
ISBN 0-471-35870-3

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

This book is intended to serve as an introduction to computation in control by an iterative, gradient, numerical method. This general approach is thought to be important, since it permits one to attack problems in control without an assumption of linearity. In effect it permits one to do for nonlinear systems many of the things that have heretofore been possible only for linear systems. The importance of being able to deal with nonlinear systems is readily evident to anyone who has to deal with any real system.

Fundamental in taking an iterative, gradient, numerical approach to control problems is the assumption of an enormous computational capability, basically to integrate the system differential equations involved. The modern, large-scale, digital computer makes all this possible. There are of course many different approaches to solving control problems with the digital computer. The gradient approach is felt to be a good one since it is a simple, generally well understood method that permits the solution of a truly large class of optimization problems, a class that extends well outside the area of control.

The general language and approach used here are those of elementary functional analysis. This selection was made because it is very general and because it is receiving increasing acceptance in a wide variety of fields. Also, from a functional analysis standpoint the basic ideas in gradient methods stand out with clarity and simplicity.

The particular gradient method that is emphasized and used here is conjugate gradient descent; it is by now a well known method and it exhibits quadratic convergence while requiring very little more computation than simple steepest descent. So far as convergence is concerned, it generally does much better than steepest descent.

It may be noted by scanning the table of contents that there is very little said about constraints directly. This is done because it is felt that constraints are a thicket that it is best not to get into deeply at the introductory level intended here. The importance of constraints is recognized, however they do tend to obscure the generally beautiful simplicity that one has with a gradient approach. Also, in control problems constraints can and are often handled as part of the nonlinearity of the dynamics. For instance, a magnitude constraint on a control input may be treated as a constraint on the control input or it may be treated as a saturation nonlinearity in series with the control input. Another approach, and one used in the text, is to introduce the constraints as penalty terms in the criterion.

The text falls naturally into two parts. Chapters 1 to 3 treat the general method of the iterative gradient approach. Here the general mathematical tools are introduced and applied to the development of the underlying theorems on conjugate gradient descent. The second part discusses the application of the general methods, developed in the first three chapters, specifically to problems in control. Those individuals interested only in control applications may limit themselves to Chapters 4 to 6. On the other hand, those interested only in the theory of conjugate gradient descent may limit themselves to the first three chapters. The author, of course, feels that the two parts complement each other, and in general neither part can be fully appreciated without the other.

A few words concerning the level of rigor intended are also in order here. In the first three chapters the basic mathematical tools are developed with some care. In the second part, which deals with applications almost entirely, a great deal of this care has been set aside with more emphasis placed on obtaining and applying specific results.

The level of the text is that of a first-year graduate student in applied mathematics or engineering. No real background in control is assumed, although this will be helpful in understanding and appreciating the choice of criteria used and the initial guesses made in doing the examples. Every method introduced is illustrated by an example. There has been a real attempt to choose an approach and to use mathematical tools that make the material both appealing and accessible to an audience outside the control field.

LAWRENCE HASDORFF

Clear Lake City, Texas
October 1975

Contents

Symbols

Symbol	Description	Where First Found
$\mathscr{H}$	Hilbert space	Fig. 1.1
$\mathscr{R}^n$	n-Dimensional real space	Fig. 1.1
x	Element of a space	(1.1)
$F(x)$	Value of a functional at x	(1.1)
x^*	Minimizing element of a functional	(1.2)
t_0	Initial time of the time interval of interest	Under (1.5)
t_f	Final time of the time interval of interest	Under (1.5)
$\phi(\cdot)$	Cost function for the control system of interest	(1.6)
$\langle x, y \rangle$	Scalar product between elements x and y	Section 2.1
$\mathscr{C}_{[a,b]}$	Space of continuous functions over the closed interval $[a, b]$	Section 2.1
$[a, b]$	Closed interval in $\mathscr{R}^1$ beginning at a and ending at b	Section 2.1
$\|x\|$	Norm of vector x	(2.4)
$\|\cdot\|_{\mathscr{E}}$	Norm of space $\mathscr{E}$	(2.10)
$\mathscr{E} \times \mathscr{L}$	Cross-product space obtained from space $\mathscr{E}$ and space $\mathscr{L}$	Under (2.7)
$\mathscr{L}^2_{[a,b]}$	Space of functions square integrable over $[a, b]$	Under (2.7)
A	Linear operator	(2.9)
sup	Supremum, least upper bound	(2.10)
$\lambda(A)$	Eigenvalue of matrix A	(2.11)
$\|A\|$	Norm of linear operator A	(2.10)
$\forall x$	For all x	(2.14)

$B(x, y)$	Value of bilinear operator B at (x, y)	(2.18)
$\langle \cdot, \cdot \rangle_{\mathscr{E}}$	Scalar product of space $\mathscr{E}$	(2.8)
A^*	Adjoint of linear operator A	(2.24)
A^+	Transpose of matrix A	(2.11)
x^+	Transpose of n-vector x	(2.15)
$F'(x_0)$	Derivative of operator F at x_0	Above Fig. 2.7
$o(\cdot)$	The order operator	(2.38)
$F''(x_0)$	Second derivative of operator F at x_0	(2.45)
A^{-1}	Inverse of the linear operator A	(2.68)
$\{p_i\}$	Set with typical element p_i	(3.2)
CGD	Conjugate gradient descent	Section 3.3
SCGD	Scaled conjugated gradient descent	Section 3.5
∇_x	Gradient operator on $\mathscr{R}^n$, w.r.t. x	(4.8)
$f(x, u)$	n-Valued function giving the dynamics ($\dot{x} = f(x, u)$) of the control system of interest	(4.9), (1.5)
$f_x(x, u)$	$n \times n$ matrix of partial derivatives of components of f w.r.t. components of x	(4.19a)
$f_u(x, u)$	$n \times m$ matrix of partial derivatives of components of f w.r.t. components of u	(4.19b)
$1(t)$	Unit step function $1(t) = 0, t < 0$ $1(t) = 1, t \geq 0$	(4.72)
$\mathscr{PC}^M_{[t_0, t_f]}$	Space of piecewise continuous functions over $[t_0, t_f]$ with M discontinuities	Above (5.4)
sgn τ	$\text{sgn } \tau = \dfrac{\tau}{\lvert\tau\rvert}$	(5.42)
$x_d(t)$	$x(t)$ delayed by T	(5.98)
x_p	Precondition on x	(5.125)
θ_{ss}	Steady-state value of θ	Under (6.8)
$y_d(t)$	The desired output signal	Above (6.1)
$\Phi_{rr}(s)$	Power spectral density of signal $r(t)$	Above (6.53)
σ_r^2	Variance of signal $r(t)$	Below (6.56)
$\bar{e}^2$	Mean square value of signal $e(t)$	Fig. 6.20

PART ONE

Chapter 1

Introduction

1.1. STATEMENT OF THE GENERAL PROBLEM

Our object here is to lay a general foundation in optimization theory that can be used to optimize the responses of the control systems that are considered subsequently. Our main aid is the digital computer, so our aim is to develop a theory that anticipates the use of the computer.

To this end we postulate a mapping, or an operator, or a transformation, or whatever from a Hilbert space $\mathscr{H}$ to the real line $\mathscr{R}^1$. This shown in Fig. 1.1. Let us call this operator F since it is a functional. F assigns to every element $x \in \mathscr{H}$ a real value $F(x)$. Now any two elements x_1 and x_2 can be ranked relative to one another depending on whether

$$F(x_1) < F(x_2)$$

or

$$F(x_1) > F(x_2) \tag{1.1}$$

Of course if $F(x_1) = F(x_2)$, then x_1 and x_2 rank equally. The functional F is thus seen to be simply a device for ordering the elements of $\mathscr{H}$.

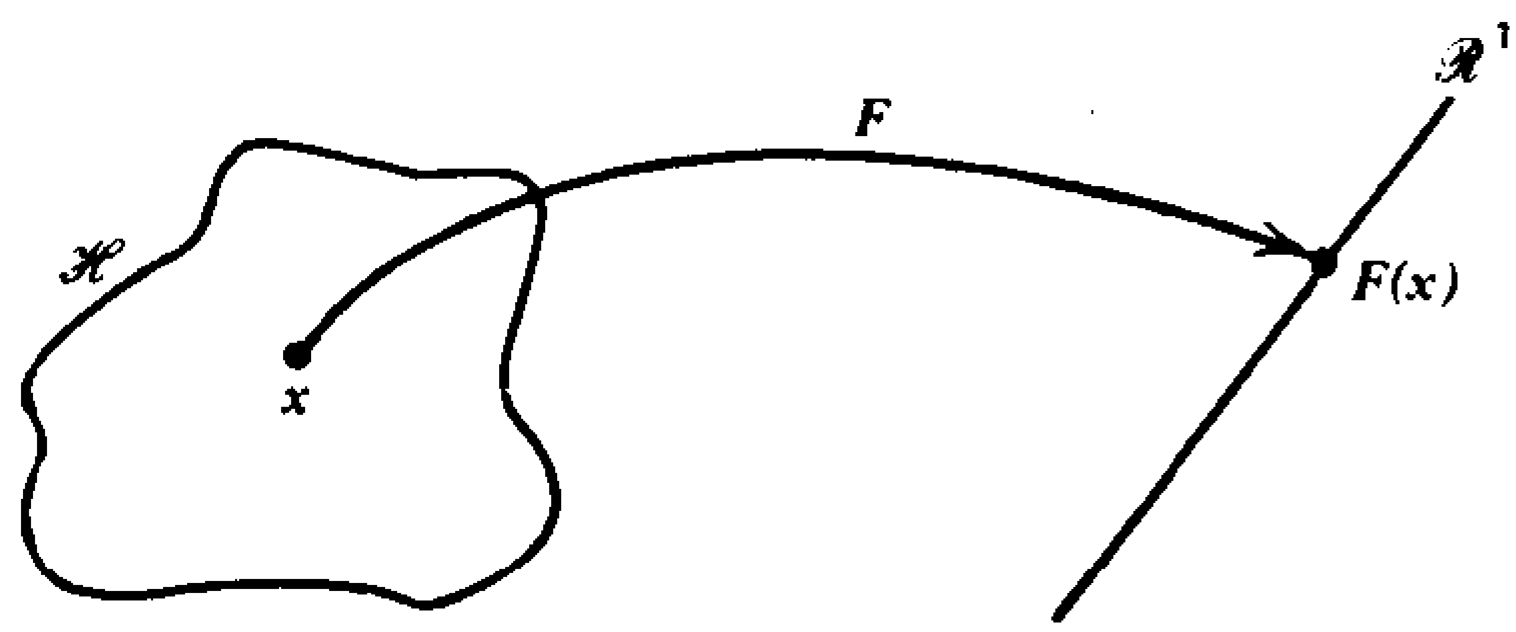

Figure 1.1 A functional F from Hilbert space $\mathscr{H}$ to the real line $\mathscr{R}^1$.

An additional and basic assumption here is that the functional F is a continuous, smooth operator. Therefore if x moves around at some fixed rate in $\mathscr{H}$, $F(x)$ does not change erratically or abruptly in $\mathscr{R}^1$.

The Problem

With these basic assumptions the basic problem is now

To find x^* in $\mathscr{H}$ such that

$$F(x^*) \leq F(x) \tag{1.2}$$

for all $x \in \mathscr{H}$ in a neighborhood of x^*.

It should be noted here that the problem is minimization. We can as well consider maximization by simply exchanging $-F$ for F and this is done in the sequel when a maximization problem arises.

Also, it can be appreciated that it would be better to choose as the problem finding the x^* such that (1.2) is satisfied for all $x \in \mathscr{H}$. That is, our problem states that we are seeking a relative minima when it would be better to seek an absolute minima. There is really no short, good explanation for this since one usually does want the absolute minimum. Suffice it to say that seeking the absolute minimum extends the problem beyond the range of the gradient methods studied here. From a practical standpoint, the relative minima techniques are usually sufficient for the control problem for which the theory is intended.

The Method of Attack

The basic method used here for finding a minimizing argument x^* of F in $\mathscr{H}$, termed descent, consists of the following steps:

1. Making an initial guess.
2. Constructing a sequence $x_0, x_1, x_2, \ldots$ such that

$$F(x_{i+1}) < F(x_i) \tag{1.3}$$

The sequence is extended until (1) no x_{i+1} can be found such that (1.3) is satisfied or (2) until the sequence $\{x_i\}$ approaches a limit. In case 1 $x_i = x^*$ and in case 2 x^* is the limit of the sequence. It is seen that we obtain either x^* or a value very near x^*.

This basic method has been chosen because it is particularly adaptable to the use of the computer. One has only to program the computer to find x_{i+1}

from x_i. The computer generally accomplishes the iteration very well for our control problems. And, though this method does introduce additional problems, it does avoid many of the problems of the more direct methods.

Thus it can be understood why a Hilbert space has been chosen as the space in which a minimizing argument for F is sought. The distinguishing features of a Hilbert space are as follows:

1. It is *linear.*
2. It is *complete.*
3. It has a *scalar product.*

1. We recall that if a space is *linear* and if x_1 and x_2 are in the space, then the element $y = ax_1 + bx_2$ is also in the space (a and b are scalar constants). Therefore our series is constructed so that

$$x_{i+1} = x_i + \alpha_i z_i \tag{1.4}$$

where z_i is also a member of the space under consideration and α_i is a scalar constant. Thus x_{i+1} is also a member of the space under consideration and the whole sequence is in the space.

2. A space is *complete* if every cauchy sequence in the space has a limit in the space. If we wish to construct a sequence that approaches a limit, it is desirable to have that limit in the space in which we are working.

3. The *scalar product* (see Chapter 2 for a definition of the scalar product) is very fundamentally involved in the way the direction of step is chosen in going from x_i to x_{i+1} (i.e., choosing z_i to be used in 1.4). How the scalar product comes into play becomes obvious in the following sections where expansion of a functional (such as F) about a given point is discussed.

Our plan for the chapters ahead is as follows:

Part One: To lay the mathematical foundation required and to develop the general algorithm to be used in constructing the sequence that is to converge on the solution to the general minimization problem stated above.

Part Two: To specialize the theory of Part One to the case of optimization of the inputs to a control system. Specifically we consider a control system whose dynamics are given by

$$\dot{x} = f(x, u) \tag{1.5}$$

where $x \in \mathscr{R}^n$, u is in Hilbert space $\mathscr{H}$, and f is a vector-valued function on $\mathscr{R}^n \times \mathscr{H}$. The initial time t_0 and final time t_f are assumed to be given. The cost of operating this system over the time interval $[t_0, t_f]$ is assumed to be given by

$$\text{cost} = J = \phi(x(t_f)) \tag{1.6}$$

where ϕ is a real-valued function on $\mathscr{R}^n$. The cost is thus a function of the final state reached. The problem then is to find $u \in \mathscr{H}$ using the general method developed in Part One, that produces minimum cost. The u considered can take many specific forms. It may be a set of initial conditions, an input control function that may take on several different forms, or a set of parameters, or it may be combinations of these.

Part Three: To use the techniques developed in Parts One and Two to attack the still more specialized problem of controller design for the class of control systems considered in Part Two.

Chapter 2

Basic Mathematical Concepts

2.1. SOME FUNDAMENTAL DEFINITIONS

We consider here definitions of the fundamental mathematical quantities and terms and introduce the notation that is used in all subsequent work. Our basic problem is finding the minimizing argument of a functional on a Hilbert space. Let us begin by defining the terms needed.

A Linear Space

A space $\mathscr{S}$ is a collection of elements; it is said to be linear if $x_1 \in \mathscr{S}, x_2 \in \mathscr{S}$ then $(ax_1 + bx_2) \in \mathscr{S}$ where a and b are scalar constants. We see in particular that if $x \in \mathscr{S}$, then $ax \in \mathscr{S}$ and so are $-x \in \mathscr{S}$ and $\boldsymbol{0} \in \mathscr{S}$, which correspond to $a = -1$ and $a = 0$, respectively. We shall in general adhere to the convention of denoting spaces by script capitals and elements of spaces by italic lowercase letters.

Examples of linear spaces are the real numbers, denoted by $\mathscr{R}^1$. Real, n-dimensional euclidean space is denoted by $\mathscr{R}^n$, and the space of continuous functions over an interval of the real line $[a, b]$, by $\mathscr{C}_{[a,b]}$. Examples of collections of elements that are not linear spaces are any finite set, any proper subset of $\mathscr{R}^n$, and the numbers in the interval $[0, 1]$.

The Scalar Product

A scalar product on a space $\mathscr{S}$ is an operation that assigns a real number to every pair of elements in $\mathscr{S}$. If the scalar product of two elements $x, y \in \mathscr{S}$ is denoted by $\langle x, y \rangle$, the scalar product has the following properties:

1. $\langle x, y\rangle \in \mathscr{R}^1$ for $x, y \in \mathscr{S}$.
2. $\langle x, y\rangle = \langle y, x\rangle$.
3. $\langle ax + bz, y\rangle = a\langle x, y\rangle + b\langle z, y\rangle$ (linear, where a and b are scalar constants).
4. $\langle x, x\rangle > 0$ if $x \neq 0$.
5. $\langle \mathit{0}, x\rangle = 0$ where $\mathit{0}$ is the null element. (2.1)

Examples of scalar products are as follows:

1. In $\mathscr{R}^1$, ordinary multiplication.
2. In $\mathscr{R}^n$ where

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{i=1}^{n} x_i y_i \tag{2.2}$$

3. In $\mathscr{C}_{[a,b]}$, with elements $x(t)$ and $y(t)$

$$\langle x, y\rangle = \int_a^b x(t)y(t)\, dt \tag{2.3}$$

Scalar Product Norm

The norm, is, obviously, a measure of the magnitude (or length) of an element in a space. For spaces on which a scalar product is defined, the most convenient and usual definition of norm for a typical element x, written $\|x\|$, is

$$\|x\| = \sqrt{\langle x, x\rangle} \tag{2.4}$$

This is the norm used in this text, since all interest centers on the Hilbert space, which by definition has a scalar product. Some examples of scalar product norms are as follows:

1. For $x \in \mathscr{R}^1$, $\|x\| = |x|$
2. For $x \in \mathscr{R}^n$, $\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$
3. For $x \in \mathscr{C}_{[a,b]}$, $\|x\| = \left[\int_a^b x^2(t)\, dt\right]^{1/2}$

We say two elements x and y of a space are orthogonal if $\langle x, y\rangle = 0$.

Two Norm Inequalities

Two very useful inequalities involving norms are Schwarz's inequality and the triangle inequality. *Schwarz's inequality* for two elements x and y from a *scalar product space is*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \tag{2.5}$$

The *triangle inequality is*

$$\|x + y\| \leq \|x\| + \|y\| \tag{2.6}$$

We offer no proof for these inequalities; any elementary text in functional analysis (see, for example, [1] or [2]) or linear spaces should have one. These inequalities, shown in Fig. 2.1, have a good geometrical interpretation in $\mathscr{R}^2$. It should be realized from the figure that the inequalities hold as well in $\mathscr{R}^n$ for the plane on which Fig. 2.1 is drawn can simply be taken as the plane in $\mathscr{R}^n$ determined by the two vectors (elements) x and y. As a simple exercise, these two inequalities may be shown to hold in $\mathscr{R}^1$, $\mathscr{R}^n$, and $\mathscr{C}_{[a,b]}$.

A Cluster Point

A point or element y is a cluster point of a sequence $\{x_n\}$ if every ε neighborhood of y contains an infinite number of members of the sequence. This means in our case, given any $\varepsilon > 0$ there are an infinite number of x_n for which

$$\|y - x_n\| < \varepsilon$$

where we assume the norm is as defined by (2.4). A cluster point is also commonly called a limit point if it is in the space in which the sequence occurs. If a sequence has a cluster point y in the sequence space, it is obvious that there is a sub-sequence that has limit y in the conventional sense of the limit.

A Complete Space

A space $\mathscr{S}$ is said to be complete if every sequence in that space with cluster point y has $y \in \mathscr{S}$. One might assume that every linear space that has a scalar product defined upon it and a corresponding norm as in (2.4) would be complete. This is true for the very important practical case of $\mathscr{R}^n$ with scalar products as in (2.2) and norm as in (2.4). However an important exception, that is, a linear, scalar-product space that is not complete, is $\mathscr{C}_{[a,b]}$ with scalar product as in (2.3) and corresponding norm as in (2.4).

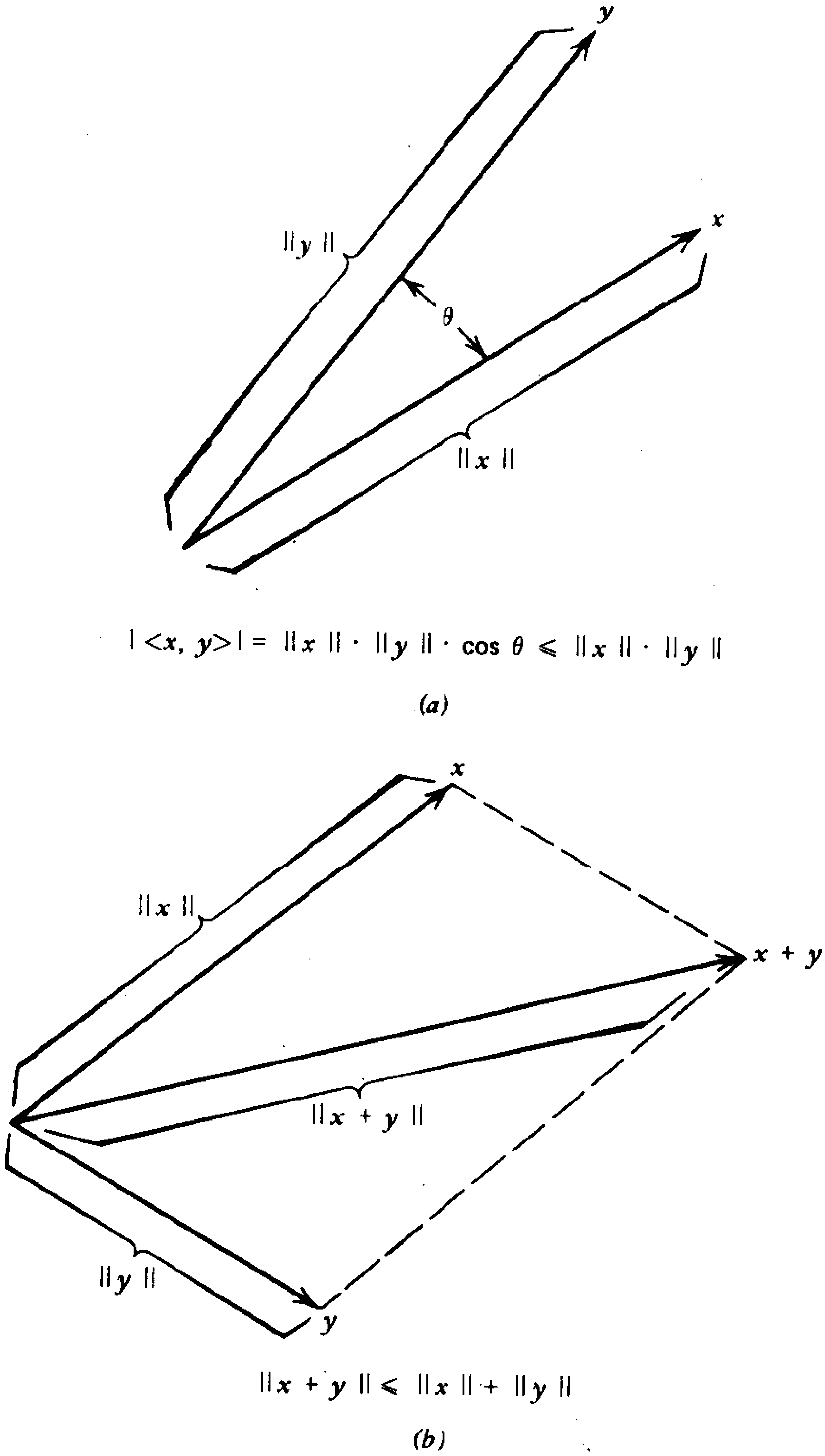

Figure 2.1 (*a*) Schwarz's inequality in $\mathscr{R}^2$. (*b*) Triangle inequality in $\mathscr{R}^2$.

To see this, we consider the functions in Fig. 2.2. Let us consider the sequence of functions $\{x_n(t)\} \in \mathscr{C}_{[a,b]}$. It is obvious how to extend the sequence of continuous functions of which $x_1(t)$, $x_2(t)$, $x_3(t)$, and $x_4(t)$ are the starting members so that

$$\int_a^b [y(t) - x_n(t)]^2 \, dt < \varepsilon$$