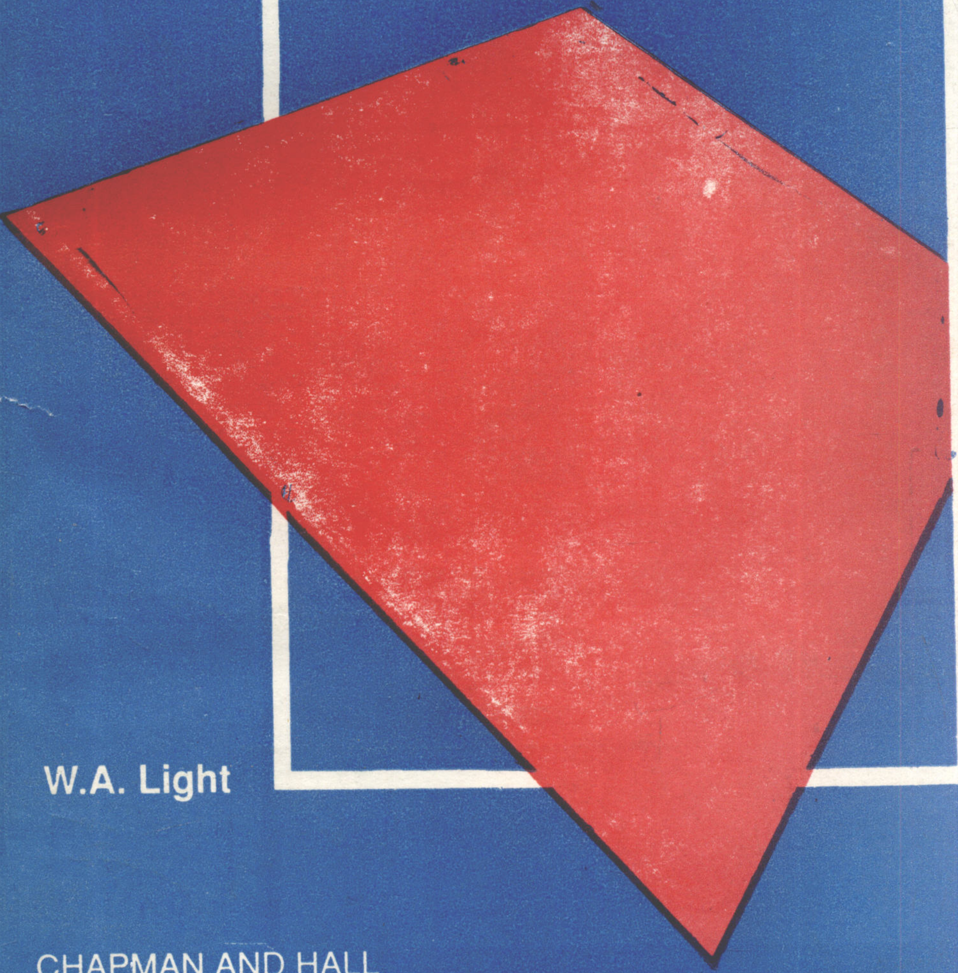
An Introduction to Abstract Analysis



W.A. Light

CHAPMAN AND HALL
World Publishing Corp

An Introduction to Abstract Analysis

W.A. Light

Senior Lecturer
Mathematics Department
Lancaster University

CHAPMAN AND HALL

World Publishing Corp

UK	Chapman and Hall, 11 New Fetter Lane, London EC4P 4EE
USA	Chapman and Hall, 29 West 35th Street, New York NY10001
JAPAN	Chapman and Hall Japan, Thomson Publishing Japan, Hirakawacho Nemoto Building, 7f, 1-7-11 Hirakawa-cho, Chiyoda-ku, Tokyo 102
AUSTRALIA	Chapman and Hall Australia, Thomas Nelson Australia, 480 La Trobe Street, PO Box 4725, Melbourne 3000
INDIA	Chapman and Hall India, R. Sheshadri, 32 Second Main Road, CIT East, Madras 600 035

First edition 1990

ISBN 0 412 31080 5 (HB)
0 412 31090 2 (PB)

British Library Cataloguing in Publication Data
Light, W.A.
An introduction to abstract analysis.
1. Linear metric spaces
I. Title
515.73

ISBN 0-412-31080-5
ISBN 0-412-31090-2 pbk

Library of Congress Cataloging-in-Publication Data
Available

This edition first published in the United Kingdom by Chapman & Hall, 1990
Reprinted by World Publishing Corporation, Beijing, 1991
Not for sale or distribution outside The People's Republic of China
ISBN 7-5062-1206-4

An Introduction to Abstract Analysis

List of notations

Introduction

There are two objectives in this introduction. Firstly, I want to give a brief overview of the history of analysis as it relates to the material in this book. Secondly, I want to explain how the book is constructed, and how it is intended to be read.

The modern theory of analysis has its roots in the work of Liebnitz and Newton as they studied and developed the theory of differential and integral calculus in the 17^{th} century. Central to this theory is the concept of a function. However, the modern definition of a function which every undergraduate mathematician encounters early in her undergraduate career is in fact the result of a long process of refinement. In the first half of the 17^{th} century, functions were defined only by algebraic operations. Thus $f(s) = s^2$ was a typical function. The importance of the range, the domain and the rule defining the function was at this stage not evident. Also, a consequence of the algebraic nature of the definition was that knowledge of the function on a small interval allowed one to deduce the nature of the function for arbitrary values of the argument. To see what this means, suppose we know that a function f mapping real numbers to real numbers is defined by an algebraic rule. If we also know that the rule for real numbers in $[0, 1]$ is given by $f(s) = s^2$, then, because the rule is known to be given by a single algebraic formula on the whole of $\mathbb{R}$, this rule must continue to hold for all real numbers. It must have universal truth. Even the idea that a function could be represented by different rules on different intervals of the real line was not thought of at this stage. The next significant step was the appearance of logarithmic and trigonometric functions. However, these functions still had a definition which had universal truth. Through the study of infinite series, the concept of a function began slowly to widen towards the end of the 18^{th} and the beginning of the 19^{th} centuries, and the idea of a function having several radically different forms of behaviour in different regions of the real line steadily gained acceptance.

In the 19^{th} century, men like Cauchy, Weierstrass, Dirichlet and Riemann began to place the subject of functions on a firm abstract foundation. Their concept of a function and associated ideas such as continuity and differentiablity were essentially those which we still use today. However, their work did not meet with uniform approval amongst mathematicians of the day. The problem was that the class of objects which were now refered to as functions was widened tremendously by these founding fathers of analysis. This process of abstraction was eventually found to have granted citizenship to some fairly bizarre functions – ones which not only had no simple rule defining them, but could not even be drawn graphically.

Once the concept of a function was firmly established, the ideas of continuity and differentiability could be placed on a firm footing. This led to another controversy. There had been a strong feeling, brought about no doubt by considering the sorts of functions which can be graphed, that every continuous function was differentiable everywhere, or at least with the exception of a small number of 'special' points. As the abstract theory became better understood, both Bolzano and Weierstrass realised independently that this feeling was badly incorrect. (The reader who is interested in knowing just how badly wrong this feeling is, need only read as far as chapter **5** in this book.) Not everybody was thrilled with this state of affairs, however. Their work caused Hermite to express himself as follows: "I turn away with fear and horror from this lamentable plague of functions which do not have derivatives".

By the time the 20^{th} century arrived, the foundations of the theory of functions of a real variable were well-understood, and analysis was ready for another change of pace and direction. Perhaps a major influence on this development was the pioneering work of G. Cantor at the end of the 19^{th} century. What Cantor may be credited with achieving is bringing the abstract concept of a set to the forefront of mathematics. The theory of functions could then take another quantum leap to the modern position where the domain and range were abstract sets. The pace now quickened, largely due to the French school of mathematicians. The work of Borel and Lebesgue on techniques for measuring sets opened the way for an abstract theory of integration which proved to be far superior to the original ideas of Riemann. Baire and Borel were involved in a theory of classification for functions, and many other able researchers contributed to making the beginning of the 20^{th} century a period of rapid development. Running parallel, and often inextricably intertwined with the real theory was the theory of functions of a complex variable.

With the publication of a book called *Théorie des opérations linéaires* [1] by a Polish mathematician, S. Banach, a new abstract framework in which many of the previous ideas could be discussed was created. It soon became apparent that this framework (that of normed linear spaces – see chapter **1**) offered powerful tools to the rest of analysis. In addition, there was considerable interest in the subject for its own sake, and this interest has given rise to important modern areas of analysis such as operator theory and the geometric theory of Banach spaces. At the same time, Banach's work gave mathematicians what amounts to a new language for analysis. Such was the power of this new language, that today it is well-understood by mathematicians whose involvement in abstract mathematics is quite small. For example, this language is prevalent in the field of numerical analysis – where techniques for using computers as aids to problem solving are studied.

This presents an interesting pedagogical problem. Anyone wishing to work in the analytical side of mathematics must be familiar with this modern language. But when is the correct time to introduce such material? A student who has only received an informal training in the methods of calculus (so that no formal proofs have been encountered) would find the combination of the formality of analysis and the strangeness of the abstract framework just too daunting. However, once a student has done the equivalent of one semester of univariate formal calculus, encountering the concepts of sequences, limit, continuity, differentiability and integration, and most important, has had plenty of practice with rigorous proofs using quantifiers, then the next step can involve some gentle introduction to normed linear spaces. Of course, a small amount of knowledge about linear structures is also necessary, but much can be done with nothing more than an understanding of the concept of a linear space together with the ideas of dimensionality and basis.

A major problem with this scheme of things is that most books in this field are simply too ambitious for the student who is still coming to terms with the formal notions of analysis. The majority of authors set out with the intention of covering a substantial portion of the theory at a late undergraduate or beginning postgraduate level. This often requires advanced technical machinery, such as topological spaces, and dictates a pace which is far too rapid for someone who is still coming to terms with the way in which analysts think, and the process by which results are established. This book attempts to provide a gentle introduction to the theory of normed linear spaces, while at the same time exposing the way in which the common arguments of analysis work. How is this achieved? Firstly, the pace of exposition at the start of the book is slow. (At least, in my opinion, it is slow. Experience has shown me that I do not always succeed in striking the right pace in the eyes of students!) I have tried to take considerable care in the initial chapters to point out the basic strategies of proof, and the common pitfalls. As the chapters progress, I presume that the reader is gradually becoming familiar with the subject, through her reading and through her working of the exercises. Consequently, the pace quickens, and the reader is left more and more to her own devices. Secondly, the choice of topics has been severely limited, so that this book is not of daunting length, neither does it contain many of the great landmarks in the subject. I considered most of these just too difficult to tackle in an introductory work. In addition, I have looked very much to the field of function theory for most of the examples. In doing so, a severe blow has been dealt to a large part of the subject, which deals with linear spaces where the objects are sequences or Lebesgue measurable functions. I feel that functions which have domain and range in the real numbers are objects which most readers will feel reasonably at home with, and so this is the best place to go for practical outlets of the theory. It has the added advantage that the applications in this book continue the themes that the student will have encountered in previous courses – continuity, ideas from calculus, polynomials, and so on.

One of the consequences of my approach is that the reader really should go on after reading this book to read other, more advanced books. I very much hope that she will do so. One of my main objects in writing this book has been to try to convey my enthusiasm for the subject to the reader, and I will regard myself as having failed if this is the last book on analysis which the reader studies.

Now a word about the structure of the book. Each chapter contains exercises liberally sprinkled throughout. My intention is that the reader should do most, if not all of these. None are really hard, and many round out the treatment of the theory. Some are even essential for certain future arguments. All theorems, lemmas, corollaries are numbered consecutively within the chapter they occur. The proofs are set off in the text by the head **Proof**, and are terminated by a small black box.

I have benefited greatly from the teaching I have received and from the many discussions I have had with colleagues about the correct way to teach analysis. It is a pleasure to acknowledge that assistance here. I am greatly indebted to my friend and colleague, Professor E. W. Cheney whose friendship and collaboration over many years has served to deepen my understanding of the subject, and to stimulate my own research interests in the area. Dr. G. J. O. Jameson has also had considerable influence on my teaching of analysis during the 18 years we have been colleagues, and this book owes much to my perusal of his analysis lecture notes. Professor W. Deeb was somehow persuaded to read the manuscript when it was almost in final form. I benefited greatly from his gentle but persistent criticisms.

Finally, I owe a deep debt of gratitude to my wife, Anita, who became a 'computer widow', while I struggled with the writing and typesetting of this book. This book is dedicated to her.

Will Light
Lancaster, 1990.

Contents

1

Basic ideas

We shall begin with some ideas which are fundamental to the rest of this book. Basic to all of mathematics is the idea of a *function* and the associated notation. We shall use the notation $f : A \longrightarrow B$ to denote the fact that A and B are two sets and f is a function (mapping) between them. It is usually understood that A is the domain of f so that $f(a)$ is the unique element of B corresponding to a via the mapping f. Things are a little different with regard to the set B, however. We normally consider B to contain the range of the function but strict containment is permissible, so that all of B need not necessarily be 'used up' by the function. Notice that each function has three pieces of information. Firstly, the domain A; secondly, the set B, which contains the range of the function; and, thirdly the rule – how to get from A to B using the mapping f. For example, the rule $f(x) = 1/x$ with $A = (0, \infty)$ and $B = \mathbb{R}$ is quite satisfactory, whereas the same rule with $A = [0, \infty)$ and $B = \mathbb{R}$ is not. (The destination of the point $x = 0$ under the rule is not defined.) Similarly, the rule $f(x) = \sin x$ with $A = B = \mathbb{R}$ is satisfactory, as is the same rule with $A = \mathbb{R}$ and $B = [-1, 1]$. In the second case the set B (or the 'target space') is exactly the range of the function.

Having made such a fuss over the care needed in talking about functions made up of the three ingredients rule, domain and target space, we often abuse the notation by refering to 'the function f'. Here the domain and target space are omitted, and this is frequently done when both are clearly understood from the context. This is usually the case when $A = B = \mathbb{R}$ or when A and B have been mentioned previously and it would be belabouring the point to continually write

$f : A \to B$. All careful mathematicians are aware of the fact that omitting the domain and/or the target space is fraught with danger and therefore try to abuse the notation carefully!

Our second fundamental ingredient is that of a linear space. The set X is a linear space if there is some method by which any two elements of X can be 'added together' to give a third member of X, and each element of X can be 'multiplied' by a real number to give another element of X. For example, the Cartesian plane $\mathbb{R}^2$ is a linear space. The points in $\mathbb{R}^2$ consist of coordinate pairs (s, t) and 'addition' means

$$(s, t) + (s_1, t_1) = (s + s_1, t + t_1).$$

Similarly, 'multiplication' by the real number (scalar) α means

$$\alpha(s, t) = (\alpha s, \alpha t).$$

Of course, the processes called addition and multiplication are not the simple concepts used in $\mathbb{R}$, although they are very similar. For example,

$$(s, t) + (s_1, t_1) = (s + s_1, t + t_1) = (s_1, t_1) + (s, t),$$

so that the new concept of addition is commutative, as is addition in $\mathbb{R}$. Furthermore, the origin $(0, 0)$ has the property that

$$(0, 0) + (s, t) = (s, t) = (s, t) + (0, 0).$$

Thus the point $(0, 0)$ in the linear space $\mathbb{R}^2$ plays an analogous role to that of 0 in $\mathbb{R}$. For this reason, it is often referred to as the *zero element* in the linear space $\mathbb{R}^2$. Given any point (s, t) in $\mathbb{R}^2$, the point $(-s, -t)$ has the property that

$$(s, t) + (-s, -t) = (0, 0) = (-s, -t) + (s, t).$$

The point $(-s, -t)$ is usually called the additive inverse of (s, t).

Consider the set X consisting of all mappings from $\mathbb{R}$ into $\mathbb{R}$. This is also a linear space. The elements in X consist of functions, and addition of two functions is effected by defining their sum to be the 'pointwise sum', so that if f and g are two points in X (i.e. two functions from $\mathbb{R}$ to $\mathbb{R}$) then the function $h = f + g$ is defined by

$$h(s) = f(s) + g(s), \quad s \in \mathbb{R}.$$

Similarly, the function y which is the result of multiplying f by the scalar α is defined by

$$y(s) = \alpha f(s), \quad s \in \mathbb{R}.$$

Faced with a concrete situation it is usually easy to say what is meant by addition and by multiplication by scalars. However, the description of these two properties in an abstract setting is rather harder. Notice that 'addition' associates with every pair of elements x, y in X a third element z which we call $x + y$. Association in mathematics usually involves mappings, and 'addition' is a mapping from $X \times X$ (ordered pairs of elements in X) into X with certain properties that make the mapping 'look like' the usual process of addition of real numbers. In a similar way scalar multiplication is a mapping from $\mathbb{R} \times X$ (ordered pairs of elements, the first element lying in $\mathbb{R}$ and the second in X) into X. It will have certain properties that make it resemble multiplication in $\mathbb{R}$. Now we are ready to say formally what constitutes a linear space.

Definition 1.1 *A linear space is a set X together with two mappings $\phi : X \times X \to X$ and $\psi : \mathbb{R} \times X \to X$ such that*

1. $\phi(x, y) = \phi(y, x)$ *for* $x, y \in X$
2. $\phi(x, \phi(y, z)) = \phi(\phi(x, y), z)$ *for* $x, y, z \in X$
3. *there exists a unique element θ in X such that $\phi(x, \theta) = \phi(\theta, x) = x$ for x in X*
4. *to each element x in X there corresponds a unique element y such that $\phi(x, y) = \theta$*
5. $\psi(\alpha, \phi(x, y)) = \phi(\psi(\alpha, x), \psi(\alpha, y))$ *for* $\alpha \in \mathbb{R}$, $x, y \in X$
6. $\phi(\psi(\alpha, x), \psi(\beta, x)) = \psi(\alpha + \beta, x)$ *for* $\alpha, \beta \in \mathbb{R}$, $x \in X$
7. $\psi(\alpha, \psi(\beta, x)) = \psi(\alpha\beta, x)$ *for* $\alpha, \beta \in \mathbb{R}$, $x \in X$
8. $\psi(1, x) = x$ *for* $x \in X$.

After the simplicity of the examples the formality looks quite daunting. However, we almost never use the mappings ϕ and ψ. Instead we always think of the mapping ϕ as 'addition' and write it as $\phi(x, y) =$

$x + y$. This is fine, as long as we remember that the sign '+' is now being used for a whole host of different meanings, depending on the context in which it is being employed. The above conditions now read

1. $x + y = y + x$ for $x, y \in X$
2. $x + (y + z) = (x + y) + z$ for $x, y, z \in X$
3. there exists a unique element θ in X such that $x + \theta = \theta + x = x$, for all $x \in X$
4. to each x in X there corresponds a unique element y (which we write as $-x$) such that $x + y = x + (-x) = \theta$
5. $\alpha(x + y) = \alpha x + \alpha y$ for $\alpha \in \mathbb{R}$, $x, y \in X$
6. $\alpha x + \beta x = (\alpha + \beta)x$ for $\alpha, \beta \in \mathbb{R}, x, y \in X$
7. $\alpha(\beta x) = (\alpha\beta)x$ for $\alpha, \beta \in \mathbb{R}, x \in X$
8. $1x = x$, for $x \in X$.

In these conditions we can see quite clearly the different meanings of 'addition'. For example, in condition 6 the addition on the left of the equality sign represents addition in the linear space, whereas the addition on the right represents the usual addition in $\mathbb{R}$. From now on we will always refer to $\phi(x, y)$ as $x + y$ and to $\psi(\alpha, x)$ as αx. Of course, in defining a linear space we need to decide three things; what choice we will make for the set of objects or points, how we will form the sum of two points and how we will take products of a point in the set and a real number. In fact the full impact of the axioms for a linear space as given in **1.1** is rarely realised in analysis. We usually work with linear spaces in which there is a simple 'natural' definition of addition and scalar multiplication, and so the problem of determining whether a given set is a linear space is rarely of great importance. Such problems belong more properly to the field of algebra. The words linear space convey the sense of the linear structure of these objects, but the historical development sometimes leads to the alternative phrase vector space.

Analysis is bound up with the idea of 'closeness' and our next concept introduces a measure of distance in a linear space. Let us return to our example $X = \mathbb{R}^2$. We are familiar with the idea that

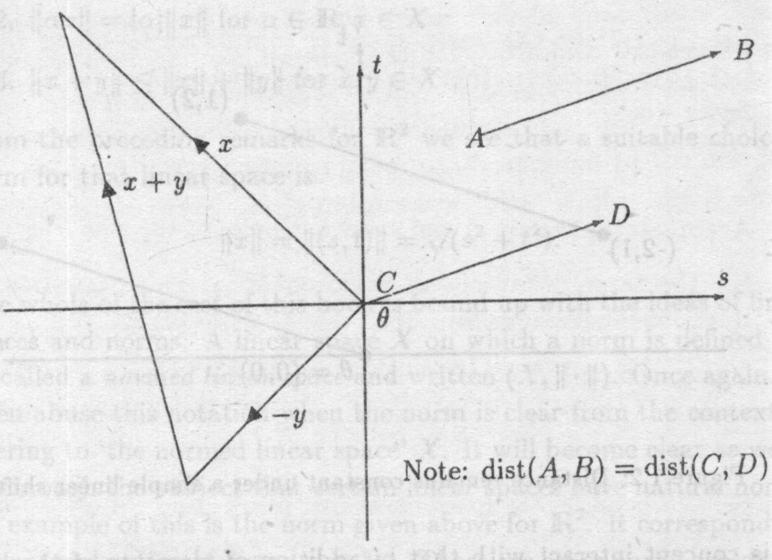


Figure 1.1: Distances in $\mathbb{R}^2$.

the distance between two points $x_1 = (-2,1)$ and $x_2 = (1,2)$ is $\sqrt{([1-(-2)]^2 + [2-1]^2)} = \sqrt{10}$. Notice in Figure 1.1 that $\sqrt{10}$ is also the distance from the point $(3,1)$ to the origin $\theta = (0,0)$, and that $(3,1) = (1,2) - (-2,1)$. This is not an accident! It will always happen that one of the properties we will require of our notion of distance is that the distance from a point x_1 to a point x_2 in $\mathbb{R}^2$ will be the same as the distance from $x_1 - x_2$ or $x_2 - x_1$ to the origin. Note that such a requirement rests partly on the fact that $\mathbb{R}^2$ is a linear space (otherwise we could never talk about $x_1 - x_2 = x_1 + (-x_2)$). The distance between points in $\mathbb{R}^2$ will therefore be completely determined once we know the distance of each individual point from the origin. There are two further properties of the distance which we will require. Both properties describe how the notion of distance ties in with the linear structure. For example, it would be nice if the distance of $10x$ from the origin was 10 times the distance of x from the origin. Thus if $\alpha \in \mathbb{R}$ then we shall demand that the distance of αx from the origin θ is $|\alpha|$ times the distance of x from θ. The modulus has appeared here because we always want distance to be a non-negative real number. This describes how the concept of distance interacts with that of multiplication by a scalar in the linear space structure. How should

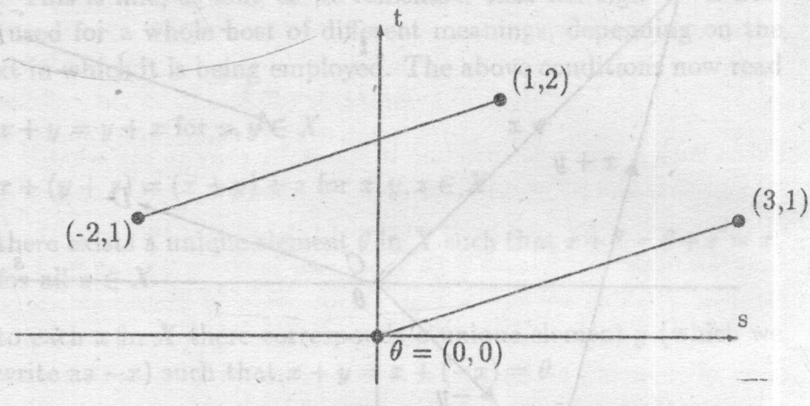


Figure 1.2: Distance remains constant under a simple linear shift.

this concept interact with that of addition of elements of the linear space? Roughly speaking, we want our measurement of distance to ensure that the direct distance between two points $x, y \in \mathbb{R}^2$ never exceeds the sum of the distances between x and any intermediate point z and z and y. Because of the linear space structure, this is the same as requiring that the distance of $x + y$ from θ is never bigger than the sum of the individual distances of x and y from θ. This last fact is illustrated in Figure 1.1. Notice that the concept of distance has associated each point in $\mathbb{R}^2$ with a real number - its distance from the origin. Once again this association is expressed mathematically by the idea of a mapping.

Definition 1.2 *Let X be a linear space. Then a norm is a mapping $\rho : X \to \mathbb{R}$ such that*

1. *$\rho(x) \geq 0$ for all x in X and $\rho(x) = 0$ if and only if $x = \theta$*
2. *$\rho(\alpha x) = |\alpha|\rho(x)$ for $\alpha \in \mathbb{R}, x \in X$*
3. *$\rho(x + y) \leq \rho(x) + \rho(y)$ for $x, y \in X$.*

As with the addition and multiplication maps it is unusual to encounter the mapping ρ. We nearly always write $\rho(x)$ as $\|x\|$ so that the above conditions read

1. $\|x\| \geq 0$ for all x in X and $\|x\| = 0$ if and only if $x = \theta$