Jerry Kazdan

THEORY OF FUNCTIONS OF A REAL VARIABLE

Peter D. Lax

Assisted by: H. Grossman
G. S. S. Avila

New York University
1958 - 1959

TABLE OF CONTENTS

Introduction

The theory of functions of real variables deals with various extensions, in the sense of enlargement*, of the classes of continuous and continuously differentiable functions of one or several variables. Such enlargements (and contractions) are needed to develop an adequate theory of those operators which occur in various problems of analysis** and for which the spaces of continuous or continuously differentiable functions do not form a natural domain. The prime examples of such operators are those which are formally symmetric or normal with respect to the ordinary scalar product for functions; e.g. the Fourier transformation and self adjoint differential operators.

The simplest and most generally useful enlargements (or contractions) of the space of continuous functions are: the L_1, L_2, L_∞ and L_p spaces; the spaces of functions whose derivatives up to a certain order belong to these spaces; and their duals. In these notes we shall define these spaces by the most direct method: By completion with respect to various metrics and by duality. Of course when introduced in this fashion the elements of these spaces are merely abstract entities - ideal elements, functions in name only. Nevertheless these abstract entities are easy enough to manipulate; e.g. functions can be formed of them, they can be differentiated and intepreted, etc. The view we wish to emphasize is that they behave sufficiently like functions to serve the purpose for which they were introduced. There are many examples illustrating the validity of this view; I regret that there was not enough time to include some of these in the lectures on which these notes are based.

---

* There are extensions of the function concept in different senses. To wit, one can take for the domain of the argument of the function something more general than Euclidean space, such as a topological space, metric space, or a differentiable manifold. Or in another sense, one can take for the range of the functional values something more general than the r eal numbers, such as Euclidean, or some more general, space.

** All problems of analysis involve in some form or another, operators. If e.g. the problem is to find a function with some prescribed properties, these properties are expressed as an equation involving an operator.

The concrete characterization of some of these classes of abstract functions as genuine functions (modulo the class of trivial functions) is presented in the last chapter. For the working analyst the significance of having such concrete realizations is twofold: First, one learns useful details about the structure of the generalized function-spaces; an example of such a useful result is Lebesgue's theorem about decomposing a measure into its singular and absolutely continuous part with respect to another. The second use is well illustrated by Lebesgue's Dominated Convergence Theorem; it is the most powerful and most often used criterion for showing the L_p convergence of a given sequence of functions.

In these notes we present the Lebesgue theory as giving a concrete realization of the abstractly defined L_1 space; much of the technique is borrowed from the Daniell approach. We have omitted a great many of the standard topics, such as the Baire classification of functions. Measure theory is kept to a minimum; e.g. product measure is not defined and Fubini's theorem is confined to a handwave.

The setting is a locally compact metric space; there is no special discussion of functions of a single variable, except in illustrations of results derived in a more general context.

Having stated our point of view, we give now a brief description of the content of these notes and indicate what additional material might have been included. First of all, the first three standard subjects of a course on real variables are not covered: the number system, the theory of sets and the rudiments of point set topology. My own lectures on these subjects were based on the outline notes (NYU, '56 - '57) prepared by J. Berkowitz. For introducing the real numbers I like to emphasize Cantor's method (equivalence classes of Cauchy sequences) since it serves as a model for the completion of metric spaces. The conclusion to be drawn is that the various function spaces, constructed by completion, are just as "real" as the real numbers; in fact, this is the most important lesson to be learned from a discussion of the number system at this level.*

* A serious discussion of the implications of the axiom of choice is not advisable.

In discussing point set topology one has the choice of sticking to metric spaces or taking into account more general topological spaces. It is desirable to acquaint students with both concepts as early as possible. Similarly, it is desirable to introduce briefly at this point the concept of a differentiable manifold. After all, it is important to know that not only continuous functions but also differentiable ones have meaningful generalizations to spaces more general than Euclidean.

In the first chapter we describe how to construct continuous functions with useful properties, i.e. the property of having the value 0 on one closed set, 1 on another, and others. This would be the place to present the simplest metrization theorem and the Tietze extension theorem. The Whitney extension theorem should at least be mentioned.

The bulk of the chapter contains a brief review of the concept of the Riemann integral, a discussion of convolution, the Weierstrass approximation theorem, and its generalization by Stone.

Chapter II is a brief introduction to functional analysis, i.e. the definition and elementary properties of linear spaces, normed linear spaces and continuous linear transformations of normed linear spaces. The principle of uniform boundedness is stated without proof. The rest of the chapter is devoted to the abstract L_p spaces; it is shown how to define integration and convolution, and a functional calculus is developed. In particular it is shown that the motion of positivity is meaningful, and the principle of monotone convergence is stated and proved. There is a brief discussion of the L_p space of functions whose values lie in a normed linear space.

In Chapter III the dual of a normed linear space is defined. The Hahn-Banach theorem is stated but not proved; it is shown that if L_p and L_g are dual, $1/p + 1/g = 1$. An axiomatic characterization of Hilbert space is given and the usual geometrical notions are developed.* Complete orthonormal sets are introduced and the convergence of Fourier series is proved. The Riesz Frechet representation theorem for linear functionals is proved and its relation to

* The weak compactness of the unit sphere ought to be mentioned here but it isn't.

the projection theorem is explained. The well known application of the projection theorem to deriving a criterion for completeness is given. Next the space H_m of functions with square integrable partial derivatives up to order m is defined by completion. The fundamental theorem of calculus is proved and the validity of the classical formula for integration by parts is noted. Sobolev's theorem (for m large enough H_m consists of continuous functions) is stated and a weak form of it is proved; Rellich's compactness criterion is stated. Next the H-m spaces are introduced as the duals of the H_m spaces with respect to the L_2 scalar product, and their relation to the space of distributions is discussed. The notion of the support of a distribution is defined.

Chapter IV presents the Lebesgue theory. The basic notion is the integral, i.e a positive linear functional I defined over the space C_o of all continuous functions with compact support over a locally compact* metric space. The volume of an open set G is defined as the supremum of the integral of all C_o functions which vanish outside G, and are ≤ 1 in G; the usual properties of volume are shown to hold. Then outer measure and sets of measure zero are defined in the usual fashion. Measurable sets are defined as those which can be approximated arbitrarily closely by open sets, i.e. the outer measure of the difference can be made arbitrarily small. The countable additivity of measure is demonstrated; that the complement of a measurable set is measurable is shown only in the section on measurable functions. Then σ- rings and Borel sets are defined, and related to measurable sets.

The usual example of a non-measurable set is presented and the Haussdorf paradox and related matters are briefly mentioned.

Next we define by completion the L_1 space with respect to an integral I. We show that every sequence of C_o functions which is a Cauchy sequence in the L_1 norm contains a sub-sequence which converges a.e. with respect to the measure induced by I. Such an a.e. limit is called an integrable function. The abstract L_1 space is shown to be in 1 - to - 1 correspondence with the equivalence class of integrable functions and it is shown that L_1 convergence and convergence a.e. are consistent. The principle of dominated convergence and Fatou's lemma are proved.

---

* By which we mean a metric space in which all bounded sets are compact

Measurable functions are defined in the usual fashion; it is shown that sums and limits of measurable functions are measurable. Esoroff's theorem is proved (but not used anywhere subsequently). It is shown that every integrable function is measurable, and that every dominated measurable function is integrable. The well-known formula for the value of the integral as the limit of approximate sums is derived as an afterthought.

Next abstract measures are defined as countably additive non-negative set functions. The Riemann-Stieltjes integral is defined; it is shown that the measure induced by an R-S integral is equal to the original measure. Signed measures are defined, and R-S integration with respect to them is also defined. Continuous linear functionals over C_o are defined and shown to be R-S integrals with respect to signed measures (Riesz representation theorem).

In the next section we define the concept of measures, singular or absolutely continuous with respect to another. The Lebesgue decomposition is given and the Radon-Nikodym theorem is proved.

There is a brief last section on differentiation, containing the classical example of a measure on the line without a discrete part which is absolutely continuous with respect to the Lebesgue measure.

There ought to be a last chapter on the basic theory of the Fourier transformation (the l_2 theory, tempered distribution and Fejer's summation), and giving a glimpse into the future. Suggested topic for brief mention: Arc length and surface area, principle value integrals and singular integral operators, projection valued measures, the Haar measure and problems in harmonic analysis.

Some Familiar Function Spaces

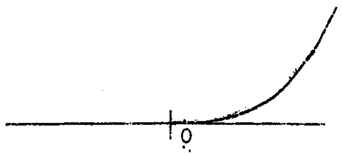
The space of real valued functions $f(x)$ defined for all real x which have derivatives of all order (are infinitely differentiable) is denoted by C^{∞}. So is the space of all functions of k real variables, defined in all R^k, which have partial derivatives of all order. The space of inifinitely differentiable functions defined in some open set D in denoted by C_D^{∞}; when there is no danger of confusion, the subscript D is dropped.

C_D^n denotes the space of all functions with continuous partial dirivatives up to order n.

These function spaces form an algebra, i.e., constant multiples, sums and products of its elements belong to the same space. Likewise, if a function in C_D^{∞} does not vanish in D, its reciprocal belongs to the same class.

Examples of infinitely differentiable functions:

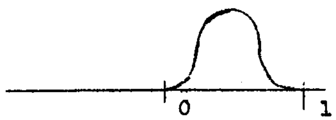
i) $a(x) = \begin{cases} 0 & x \text{ negative} \\ e^{-1/x} & x \text{ positive} \end{cases}$

i)


ii) $b(x) = a(x)a(1-x)$

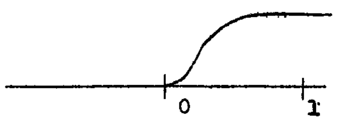
iii) $c(x) = \dfrac{a(x)}{a(x)+a(1-x)}$,

iv) $B(x) = \sum_{n=-\infty}^{\infty} b(x-n/2)$

ii)


v) $p(x) = \dfrac{b(x)}{B(x)}$

Define $p_n(x) = p(x-n/2)$; then, by construction,

iii)


$$\sum_{-\infty}^{\infty} p_n(x) \equiv 1 .$$

This is called a (smooth) partition of unity.

By multiplication we obtain a partition of unity in k dimensional space:

$$\sum p_{n_j}(x^1) \dots p_{n_k}(x^k) \equiv 1 .$$

Theorem: Let K_1 and K_2 be two closed disjoint pointsets in R^k. Then these exists an infinitely differentiable function $f(x)$ in R^k which is equal to zero on K_1, equal to one on K_2.

Proof: We take the case first that one of the sets, say K_1, is bounded. Since no point of K_1 belongs to K_2, and K_2 is closed, each point of K_1 lies at some positive distance from K_2. Draw around each point an open sphere with radius less than half that distance, the set of all these spheres constitutes an open covering of K_1. Since K_1 is a closed bounded subset of R^k, it is compact, so by the Heine-Borel theorem a finite subcovering can be selected. Let $S_1,\dots,S_n$ be the spheres in this finite covering, with centers $x_1,\dots,x_n$ and radii $r_1,\dots,r_n$. Define

$$f(x) = \prod_{i=1}^{n} c\left(\frac{|x-x_i|}{r_i} - 1\right) ;$$

here $|x-x_i|$ denotes the distance of x to x_i. Clearly, f has the desired properties.

If K_1 is unbounded, we write it as a union of bounded sets:

$$K_1 = UH_j$$

where

$$H_j = \left\{ x \,|\, x \epsilon K_1, j \le |x| < j+1 \right\} .$$

Denoting by f_j the function vanishing on H_j one on K_2 we put

$$f(x) = \prod_{j=1}^{\infty} f_j(x) .$$

For each x lying in a bounded set, this is a finite product, since all the functions $f_j(x)$ are $=1$ for $j > |x|+2$, if we make sure that the radii r used in the previous construction do not exceed 1.

Summary of Results about Riemann Integration:

Let f belong to the totality of continuous functions in R^k which vanish outside some bounded set. To each such function the definite integral

$$I(f) = \int_{x \varepsilon R} f_k(x)dx \quad = \int_{x\in R}^{(k)} \int f(x_1, \dots, x_k)\, dx_1\, dx_2 \dots dx_k$$

is defined. The functional I(f) has the following properties

i) Linearity:* $I(af+bg) = aI(f)+bI(g)$ for any real numbers a, b.

ii) Translation Invariance: $I(Tf) = I(f)$ for any translate $Tf = f(x+x_o)$. (Since we have integrated over the whole space or at least all portions of the space where the function is not zero)

iii) Positivity: $I(f) \geq 0$ if $f \geq 0$.

From the positivity of I we can deduce its boundedness: Suppose that f vanishes for $|x| \geq R$, and denote $\text{Max}|f(x)|$ by M. Then

(*) $\quad I(f) \leq c(R)M$,

c(R) a constant depending only on R.

Proof is left to the reader; of course the best value of c(R) is the volume of k-dimensional sphere with radius R.

From the boundedness of I(f) follows its continuity:

Let $f_1, f_2, \dots$ be a sequence of continuous functions, all vanishing for $|x| \geq R$, (with compact support) which converges uniformly to a function f(x):

$$\text{Max}|f_n(x)-f(x)| \longrightarrow 0 .$$

Then $I(f_n) \longrightarrow I(f)$.

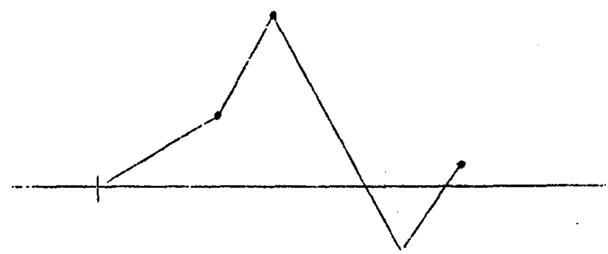
Again, proof is left to the reader.

Theorem: The three properties i) - iii) characterize the Riemann integral, i.e., any functional I(f) defined for all f of the above class and satisfying i) - iii) is a constant multiple of the Riemann integral.

Proof: We shall give the proof in one dimension. We shall operate with piecewise linear functions, i.e., functions f defined by

$$f(x) = a_i X + b, \quad x_i \leq x \leq x_{i+1} ;$$

The numbers a_i and b_i are required to be such that at the corners x_i the function f(x) is continuous, and that it is identically zero for (x) large. In what follows we shall deal with piecewise linear functions whose corners x_i are rational numbers. A piecewise linear function looks as follows:



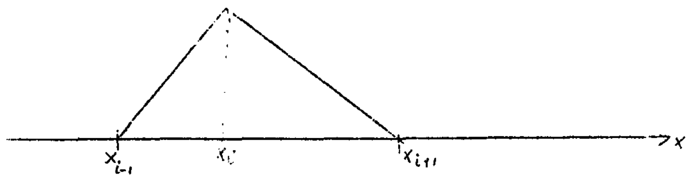
A piecewise linear function is uniquely characterized by specifying the position of its corners and its value at these corners.

Constant multiples and sums of piecewise linear functions are again piecewise linear.

---

* We are using here the fact that the class of functions considered forms a linear space, i.e., that linear combinations of its elements again belong to the class.

The simplest piecewise linear function is one with only three corners, a so-called "roof function."



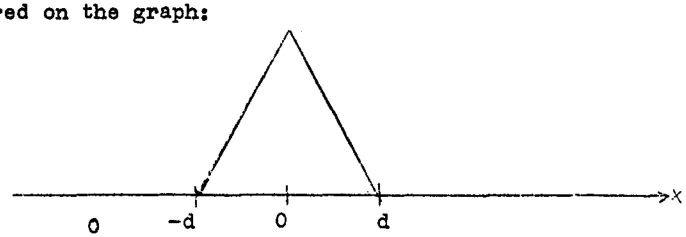
<u>Lemma</u>: Every piecewise linear function is the sum of roof functions:

$$f(x) = \sum r_i(x)$$

where $r_i(x)$ is the roof function equal to $f(x)$ at x_i, zero for $x \le x_{i-1}$ and $x_{i+1} \le x$. Here x_i is a finite, monotonic sequence of points which <u>includes</u> the corners of f. If the corners of f are rational, we can choose the x_i to be a set of equidistant rational points. Denoting by r_d the normalized symmetric roof function

$$r_d(x) = \begin{cases} 0 & x \le -d \\ d+x & -d \le x \le 0 \\ d-x & 0 \le x \le d \\ 0 & d \le x \end{cases}$$

pictured on the graph:



we can say:

<u>Every piecewise linear function with rational corners is the linear combination of translates of</u> r_d, d <u>the reciprocal of a suitably chosen integer</u>:

$$f = \sum a_i r_d(x-x_i) \ .$$

In particular, the function $r_1(x)$ can be written as the linear combination of translates of r_d if d is the reciprocal of an integer. Using linearity and translation invariance of I we see that $I(r_d)$ is uniquely determined once $I(r_1)$ is specified, and consequently also I(f) is uniquely defined. Next we use the result :

<u>The set of piecewise linear functions with rational corners is dense (in the maximum distance) in the set of all continuous functions which vanish outside a finite interval.</u>

We sketch the proof, which is quite simple: Let f(x) be any continuous function which is zero outside a finite interval. On the finite interval on which f is different from zero, f is uniformly continuous, * i.e., given any ε, we can find δ such that $|f(x)-f(y)| < \varepsilon$ when $|x-y| < \delta$. Now divide the interval in question by a finite number of rational subdivisions x_i into subintervals of length less than δ, and construct the piecewise linear function g(x) equal to f(x) at the points x_i. It is clear that $|f(x)-g(x)| < \varepsilon$. Letting ε tend to zero, we obtain f as the uniform limit of piecewise linear functions g_ε. Since we have already shown that $I(g_\varepsilon)$ is uniquely determined in terms of $I(r_1)$, by using the continuity of I we see that I(f) too is determined.

Here are some simple applications of the aforementioned uniqueness theorem for the integral:

---

* This is a special instance of the result that a continuous function on a compact set is uniformly continuous.

1) For any function $f(x)$, define $\tilde{f}(x)$ as $f(-x)$. Clearly, if $f(x)$ is continuous and vanishes outside a bounded pointset, so does $\tilde{f}(x)$. Define the functional $\tilde{I}$ as follows:

$$\tilde{I}(f) = I(\tilde{f}) \quad ,$$

where I denotes the Riemann integral. It is easy to verify that the functional $\tilde{I}$ has the three properties characteristic of the Riemann integral. Therefore from the uniqueness theorem it follows that $\tilde{I}(f) = cI(f)$. Taking f to be symmetric, i.e., $f(x) = f(-x)$ we see that $c = 1$. So we have the (trivial) result:

$$\tilde{I}(f) = I(f) \ .$$

2) Let $f(x,y)$ be a continuous function in R^2, which vanishes outside some bounded set. Define the function $g(y)$ by

$$\int f(x,y)dx = g(y) \quad ,$$

From the continuity property of the Riemann integral it follows that $g(y)$ is a continuous function of y. Since f vanishes outside some bounded set, so does g. Therefore $\int g(y)dy$ exists; denote the value of this integral by $I'(f)$. It is easy to show that this functional I' satisfies the three properties that were shown to be characteristic of the Riemann integral. Therefore $I'(f) = cI(f)$. Using a single special example, we can show that the value of c is 1. Therefore we have shown that, for the class of functions considered, <u>the repeated integral and the multiple integral are equal</u>.

<u>The Convolution</u>

Let f and g be a pair of continuous functions defined in R^k <u>of which at least one vanishes outside a bounded set</u>. For such a pair of functions we define the convolution product $f*g = h$ by the formula

$$f*g = h(x) = \int f(x-y)g(y)dy \ .$$

The convolution product has a number of important properties. These can be classified as <u>analytic</u> and <u>algebraic</u>:

<u>Analytic Properties</u>:

i) $f*g$ is continuous.

<u>Proof</u>: Applying inequality (*) expressing the boundedness of the integral:

$$h(x_1)-h(x_2) = \int [f(x_1-y)-f(x_2-y)]g(y)dy \leq \text{constant}\cdot G\varepsilon \quad ,$$

where G is the maximum of $g(y)$, ε the maximum deviation of $f(x_1-y)$ from $f(x_2-y)$ over the set where the integrand $\neq 0$. Here we make use of the uniform continuity and boundedness of continuous functions over compact sets.

ii) If both f and g vanish outside of a bounded set, so does $f*g$.

This can be read off from the formula defining the convolution. The relation of the sets outside of which f, g and $f*g$ vanishes will be examined in greater detail later.

iii) If both f and g vanish outside of bounded sets,

$$I(f*g) = I(f)I(g) \ .$$

This result is easily obtained by integrating with respect to x both sides of the equation defining $h(x) = f*g$, reversing the order of x and y integration on the right and using the translation invariance of the integral.

From the above result and the obvious inequality

$$|f*g| \leq |f|*|g| \quad ,$$

we get

$$I(|f*g|) \leq I(|f|)I(|g|) .$$

This inequality (and more general ones like it) will turn out to be very important in extending the notion of convolution to more general function classes.

Algebraic Properties

i) Commutativity:

$$f*g = g*f$$

Proof: Introduce $y' = x-y$ as new variable of integration.

ii) Associativity:

$$f*(g*h) = (f*g)*h ,$$

if two of the three functions f,g, h vanish outside of some bounded set.

The verification involves interchanging orders of integration.

iii) Linearity (≡ distributive law)

$$(a_1f_1+a_2f_2)*g = a_1(f_1*g)+a_2(f_2*g) ,$$

where a_1, a_2 are constants.

iv) Convolution commutes with translation:

$$T(f*g) = (Tf)*g .$$

Verification is immediate. A closely related analytic result is:

v) If f has continuous partial derivatives of order n and g of order m, then f*g has continuous partial derivatives of order n+m. Furthermore

$$D^nD^m(f*g) = (D^nf)*(D^mg) ,$$

where D^n and D^m denote partial differentiations to order n and m.

By induction it is sufficient to prove the result for $n+m = 1$, a verification of this case is left to the reader.

vi) Suppose now that g has the following properties:

i) g is nonnegative, and $I(g) = 1$.

ii) g vanishes outside a sphere of radius δ around the origin.

Suppose further that f has the following property: for all pairs of points x_1, x_2 lying in a certain set K, $|f(x_1)-f(x_2)| < \varepsilon$ whenever $|x_1-x_2| < \delta$.

Assertion:

$$|f(x)-h(x)| < \varepsilon \quad \text{i.e.} \quad |f(x)-f*g| < \varepsilon$$

for all x in K; here h is f*g.

Proof:

$$f(x)-h(x) = \int[f(x)-f(x-y)]g(y)dy$$

$$\leq \int|f(x)-f(x-y)|g(y) \leq \varepsilon \int g(y)dy = \varepsilon .$$

Now let K be any bounded set, f(x) any given continuous function. Since f(x) is uniformly continuous over any compact set, the above condition imposed on f is satisfied if δ is small enough. We choose g_ε to satisfy the conditions imposed on it, and in addition so that it is infinitely differentiable; such a function can be constructed from the second example listed at the beginning. Then $f*g_\varepsilon$ approximates f within ε on K; according to analytic property iv), $f*g_\varepsilon$ is infinitely differentiable if g_ε is. So we have this

<u>Theorem</u>: Given a continuous function defined in the whole space K, an arbitrary bounded set, there exists a sequence of infinitely differentiable functions f_ε which converges to f uniformly on K.

By a simple selection process we can achieve that f_ε tends to f on <u>every</u> bounded set.

Likewise, if f has continuous partial derivatives of order n, the same choice of f_ε ($= f*g_\varepsilon$) yields a sequence all of whose derivatives up to order n approximate the corresponding derivative of f uniformly on any bounded set.

Suppose we replace condition ii) on g by the following weaker one:

ii') The integral of g outside the sphere of radius δ is less than ε,

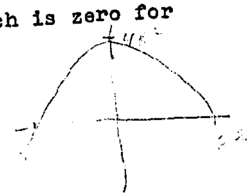
and impose the following additional condition on f; $|f(x)| \leq M$ for all x, M some number.

Then we get the following estimate for f-f*g:

$$f(x)-h(x) = \int [f(x)-f(x-y)]g(y)dy =$$

$$= \int_{|y|<\delta} + \int_{|y|>\delta} \leq \varepsilon+2M\varepsilon = (2M+1)\varepsilon .$$

Take now f(x) to be any continuous function which is zero for $|x| \geq R$, and introduce the auxiliary functions



$$g(x) = 4R^2 - |x|^2 ,$$

and

$$\hat{g}(x) = \begin{cases} g(x) & \text{for } |x| \leq 2R \\ 0 & \text{for } 2R \leq |x| . \end{cases}$$

The function $\hat{g}(x)$ has unique maximum at x = 0. From this it follows easily that its n^{th} power, $[\hat{g}(x)]^n$, is strongly peaked at x = 0, and therefore that $c_n[\hat{g}]^n = \hat{g}_n$ satisfies the conditions i) and ii'), provided that the constant c_n is so chosen that $I(\hat{g}_n) = 1$. Therefore we have

$$|f-f*\hat{g}_n| \leq (2M+1)\varepsilon .$$

Introduce now the auxiliary function

$$g_n(x) = c_n[g]^n .$$

Since g(x) and $\hat{g}(x)$ are equal for $|x| \leq 2R$, it follows that also $g_n(x)$ and $\hat{g}_n(x)$ are equal for $|x| \leq 2R$. Compare now $f*g_n$ and $f*\hat{g}_n$; their difference is $f*(g_n-\hat{g}_n)$, since the first factor for $|x| \leq 2R$, it follows that their convolution product vanishes for $|x| \leq R$. Thus we can replace in the foregoing result $\hat{g}_n$ by g_n:

$$|f-f*g_n| \leq (2M+1)\varepsilon$$

for $|x| \leq R$.

The function g_n is a <u>polynomial</u>, i.e., the sum of products of powers of the independent variables $x^1, x^2, \ldots, x^k$. The convolution of a polynomial g_n with any function f which vanishes outside of a bounded set is, as can be seen immediately from the definition of convolution, again a polynomial. Therefore we conclude from our last inequality:

<u>Any continuous function can be uniformly approximated by polynomials on any compact set.</u>

This is called the <u>Weierstrass approximation theorem</u>. The restriction that the function vanish for $|x|$ sufficiently large was used in the proof but can be easily achieved artificially.

<u>The Stone-Weierstrass Theorem</u>: Let X be a compact space,* $F = \{f\}$ a set of continuous real-valued functions on X with the following properties:

i) F <u>separates</u> the points of X, i.e., given two distinct points x and y of X, there exists a function f of F such that $f(x) \neq f(y)$,

ii) The functions $\{f\}$ of F do not have a common zero.

<u>Assertion</u>: Every continuous function on X can be approximated uniformly by polynomials of F.

By a polynomial of F we mean a linear combination of products of functions in F.

<u>Remark</u>: Both assumptions are obviously necessary; for if all functions of F agree at x and y, so does any polynomial in F and therefore any limit of polynomials. On the other hand there exist continuous functions which take on different values at x and y. Likewise, if all functions in F vanish at some point x, so will all polynomials of F, and all their limits. On the other hand, not all continuous functions need vanish at x; e.g., the constant function does not.

Denote by G the set of functions which are uniform limits of polynomials of F. We wish to show that G comprises all continuous functions on X. First we show that the <u>constant function</u> belongs to G.

<u>Lemma 1</u>: There exists a polynomial of F which is positive on X.

---

* So far only metric spaces have been discussed; but this theorem and the proof given here are valid for the more general class of compact topological spaces.

<u>Proof</u>: Given any point x of X, there exists a function of F which is not zero at x; then by continuity it is not zero in some neighborhood of x. These neighborhoods cover X; since X was assumed compact, a finite number of these neighborhoods cover X. Let $N_1, \ldots, N_k$ be these neighborhoods, and $f_1, \ldots, f_k$ the corresponding non-vanishing functions. Then the quadratic polynomial

$$f_1^2 + \ldots + f_k^2 = f$$

is positive on X.

<u>Remark</u>: A result of this kind can be proved for spaces which satisfy a condition weaker than compactness, requiring that any open covering of the space contains a <u>locally finite</u> subcovering. A covering is called locally finite if every point has a neighborhood which intersects only a finite number of the covering sets. This property is called <u>para-compactness</u>. The usefulness of locally finite coverings lies in this:

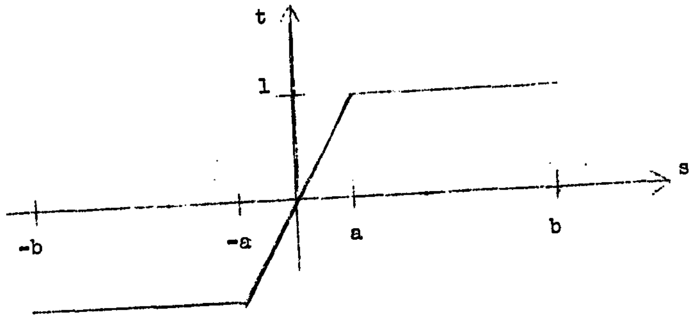
$$\sum f_i(x) \quad ,$$

where the functions f_i vanish outside of open sets which constitute a locally finite covering, is effectively a finite sum.

Since the space X is compact, the positive function f defined above achieves its minimum and maximum, a and b, which are positive quantities.

<u>Lemma 2</u>: Given an interval $a \leq s \leq b$ on the positive real axis, there exists a sequence of polynomials $p_n(s)$ <u>without constant term</u> which tends to 1 uniformly on the interval $a \leq s \leq b$.

<u>Proof</u>: Consider the function t(s) defined on the interval $-b \leq s \leq b$:



This $t(s)$ is equal to 1 in $a \leq s \leq b$, is continuous and <u>odd</u>:

$$t(-s) = -t(s) \quad .$$

According to the Weierstrass approximation theorem, there exists a polynomial $g(s)$ within ε of $t(s)$:

$$|t(s)-g(s)| < \varepsilon \quad , \qquad -b \leq s \leq b \quad .$$

From the symmetry of the interval of definition and of t around the origin, it follows that the <u>odd part</u> of g, i.e., the function

$$p(s) = \frac{g(s)-g(-s)}{2}$$

also approximates $t(s)$ within ε. Since $t(s) = 1$ in $a \leq s \leq b$, it follows that $p(s)$ differs from 1 by less than ε in that interval. But $p(s)$ being an odd polynomial, is without constant term, as asserted in Lemma 2.

Since the function $f(x)$ constructed before lies within the range $a \leq f \leq b$, it follows that $|p(f(x))-1| < \varepsilon$ on X. But $p(f)$, being a polynomial of a polynomial, is itself a polynomial. This completes the proof of Lemma 2.

<u>Lemma 3</u>: If g belongs to G and $p(s)$ is a continuous function on the range of g, then $p(g)$ belongs to G.

<u>Proof</u>: By the Weierstrass approximation theorem, $p(s)$ can be approximated uniformly in the interval which is the range of g, by polynomials $p_n(s)$. So $p_n(g)$ approximates $p(g)$; but $p_n(g)$ is the sum of a constant plus powers of g; all of these functions belong to G, and therefore so does their sum $p_n(g)$.

<u>Lemma 4</u>: Let y and z be two distinct points of X; there exists a function g in G with the following properties:

$$0 \leq g(x) \leq 1 \quad \text{for all } x \text{ in } X \quad ,$$
$$g(x) = 0 \quad \text{in some neighborhood of } y \quad ,$$
$$g(x) = 1 \quad \text{in some neighborhood of } z \quad .$$

<u>Proof</u>: According to assumption, there exists a function f in F which separates y and z:

$$a = f(y) \neq f(z) = b \quad .$$

Define $p(s)$ as some continuous function which is equal to 0 in some neighborhood of a, equal to 1 in some neighborhood of b. Clearly, $g(x) = p(f(x))$ has the required properties and, according to Lemma 3, belongs to G.

<u>Lemma 5</u>: Let K be a closed set in X, z a point not in K. There exists a function g in G with the following properties:

$$0 \leq g(x) \leq 1 \quad \text{for all } x \text{ in } X \quad ,$$
$$g(x) = 0 \quad \text{for } x \text{ in } K \quad ,$$
$$g(z) = 1 \quad .$$

To construct such a function, we consider the functions g_y of the previous lemma, for all y in K; g_y vanishes in some neighborhood S_y of y. By the Heine-Borel theorem a finite number of these neighborhoods suffice to cover K; the <u>product</u> of the corresponding functions g_y has the required properties.

<u>Lemma 6</u>: Let K_1 and K_2 be two closed, disjoint sets in X; then there exists a function g in G with the following properties:

$$0 \le g(x) \le 1 \quad \text{for all } x \text{ in } X \ ,$$
$$g(x) = 0 \quad \text{on } K_1 \ ,$$
$$g(x) = 1 \quad \text{on } K_2 \ .$$

Let z be any point of K_2, g_z the function constructed for the previous lemma; g is positive in some neighborhood of z. By Heine-Borel a finite number of these neighborhoods suffice to cover K_2; their <u>sum</u> $\sum g_z$ vanishes on K_1, is positive on K_2. Let a be the minimum of $\sum g_z$ on K_2, and construct an auxiliary function p(s) which is continuous, $p(0) = 0$ and $p(s) = 1$ for $a \le s$. Clearly, $g = p(\sum g_z)$ has the desired properties.

We turn now to the Stone-Weierstrass theorem. Let h(x) be any continuous function on X; we shall write

$$h = \sum_0^{\infty} g_n \ ,$$

where g_n belongs to G and tends to zero uniformly; this would prove the theorem.

We take g_0 to be a constant less than the minimum of h; g_n will be non-negative, satisfying the inequality

$$\sum_0^N g_n \le h \ ,$$

$$0 \le g_N \le \varepsilon_n \ ;$$

$$\varepsilon_N = K(\tfrac{2}{3})^N \ .$$

We construct the g_n recursively: By previous construction; $\sum_0^{N-1} g_n \le h$. Let K_1 be the set where $h - \sum_0^{N-1} g_n < \frac{1}{3}\varepsilon_{N-1}$, K_2 the set where $\frac{2}{3}\varepsilon_{N-1} \le h - \sum g_n$. Let g be the function described in Lemma 6, and put $g_N = (1/3)\varepsilon_{N-1} g$. It is clear that g_N has the required properties.

Application of the Stone-Weierstrass theorem. Let X be the <u>multitorus</u> or <u>period parallelogram</u>, defined as follows:

Let $R = R^k$ denote k-dimensional Euclidean space, L the subset of point $x = (x^1, \ldots, x^k)$ whose coordinates x^j are integers. Define two points x and y of R <u>congruent</u>:

$$x \equiv y$$

if x-y belongs to L; here x-y denotes the usual vector difference. Under vector addition, R forms a commutative group and L is a subgroup * of it; from this it follows, it is easy to show, that congruence as defined above is an equivalence relation, i.e.,

i) $x \equiv x$. (Reflexivity)

ii) If $x \equiv y$, $y \equiv x$. (Symmetry)

iii) If $x \equiv y$, $y \equiv z$, $x \equiv z$. (Transitivity)

The set of all equivalence classes X forms the multitorus. Sums of equivalent elements are equivalent; ** this permits the definition of addition of equivalence classes: the sum of two classes is defined in the obvious way as the class to which the sum of two arbitrary members of each class belongs. It is easy to show that the multitorus is a group, the so-called <u>quotient group</u> R/L.

We introduce the following <u>norm</u> among equivalence class:

$$|X| = \inf_{x \text{ in } X} |x| \ ,$$

---

* I.e., sums and differences of vectors in L lie in L.

** This is not necessarily so if the group operation in question is noncommutative; subgroups for which this is nevertheless true are called invariant.

where $|x|$ denotes the usual Euclidean length of the vector x. Actually, in this case the infimum is assumed, i.e., it is a true minimum; we have stated the definition in this form to anticipate more general situations where no minimum is assumed.

Using the fact that L is a closed set in R, one can show:

i) $|X|$ is positive, unless X is the zero class, i.e., the class containing the zero vector.

Let X and Y be two classes, x and y members of these classes whose length is within ε of the norm of the class:

$$|x| \leq |X| + \varepsilon \quad ,$$
$$|y| \leq |Y| + \varepsilon \quad .$$

By the triangle inequality for vectors

$$|x+y| \leq |x| + |y| \quad .$$

According to the above inequalities, the right side is not greater than

$$|X| + |Y| + 2\varepsilon \;;$$

on the other hand, by definition as infimum, $|X+Y|$ is bounded by the left side. So we have

$$|X+Y| \leq |X| + |Y| + 2\varepsilon \quad .$$

Since this is valid for any positive ε, it must hold with $\varepsilon = 0$:

ii) $|X+Y| \leq |X| + |Y|$. We define now as distance on the multitorus

$$d(X,Y) = |X - Y| \quad .$$

It follows from the above two properties of the norm that this distance satisfies the three postulates of a distance function. It follows easily from Bolzano-Weierstrass:

Theorem: The multitorus, metrized in the above fashion, is compact.

A function $f(x)$ defined in R^k is called periodic if $f(x) = f(y)$ whenever $x \equiv y$. Such a periodic function in R^k can be interpreted as a function on the multitorus; if f is continuous as function on R^k, it is continuous on the multitorus.

Examples of periodic functions are

$$(*) \qquad \sin 2\pi x^j \;,\quad \cos 2\pi x^j \;,\quad j = 1,\ldots,k \;,$$

and more generally,

$$(**) \qquad \sin 2\pi(n_1x^1+\ldots+n_kx^k) \;,\quad \cos 2\pi(n_1x^1+\ldots+n_kx^k) \;,$$

n_j integer. The set F of 2k functions (*) satisfies the assumptions of the Stone-Weierstrass theorem: they separate points and do not vanish simultaneously. Therefore every continuous function can be approximated by polynomials in them. Using the identity

$$2 \sin \alpha \sin \beta = \sin(\alpha+\beta)+ \sin(\alpha-\beta)$$

and similar ones for the products of sines and cosines and of cosines - we can express a polynomial in the functions (*) as linear combinations of functions (**), called trigonometric polynomials. So we have proved the theorem, also due to Weierstrass:

Every continuous periodic function can be approximated uniformly by trigonometric polynomials.

It is also possible to give a direct construction of the approximating trigonometric polynomials by a convolution.