VAN NOSTRAND REINHOLD MATHEMATICAL STUDIES #24

SMOOTHING and APPROXIMATION of FUNCTIONS

Harold S. Shapiro
The University of Michigan

LD MATHEMATICAL STUDIES

RES ON BOOLEAN ALGEBRAS

.LIPTIC BOUNDARY VALUE PROBLEMS

N DIFFERENTIAL GEOMETRY

-TOPOLOGY AND ORDER

IOTES ON SPECTRAL THEORY

n—NOTES ON LOGIC

RES ON CHOQUET'S THEOREM

CISES IN SET THEORY

George W. Mackey—**LECTURES ON THE THEORY OF FUNCTIONS OF A COMPLEX VARIABLE**

Lars V. Ahlfors—**LECTURES ON QUASICONFORMAL MAPPINGS**

J. Peter May—**SIMPLICIAL OBJECTS IN ALGEBRAIC TOPOLOGY**

W. H. J. Fuchs—**TOPICS IN THE THEORY OF FUNCTIONS OF ONE COMPLEX VARIABLE**

Lennart Carleson—**SELECTED PROBLEMS ON EXCEPTIONAL SETS**

Leopoldo Nachbin—**ELEMENTS OF APPROXIMATION THEORY**

S. S. Chern—**COMPLEX MANIFOLDS WITHOUT POTENTIAL THEORY**

F. P. Greenleaf—**INVARIANT MEANS ON TOPOLOGICAL GROUPS**

H. Widom—**LECTURES ON INTEGRAL EQUATIONS**

A. K. Aziz, Gen. Ed.—**LECTURE SERIES IN DIFFERENTIAL EQUATIONS, Vol. I**

A. K. Aziz, Gen. Ed.—**LECTURE SERIES IN DIFFERENTIAL EQUATIONS, Vol. II**

Harold Widom—**LECTURES ON MEASURE AND INTEGRATION**

Shaul R. Foguel—**THE ERGODIC THEORY OF MARKOV PROCESSES**

Makoto Ohtsuka—**DIRICHLET PROBLEMS, EXTREMAL LENGTH AND PRIME ENDS**

Hans Samelson—**NOTES ON LIE ALGEBRAS**

Harold S. Shapiro—**SMOOTHING AND APPROXIMATION OF FUNCTIONS**

SMOOTHING AND APPROXIMATION of FUNCTIONS

by

HAROLD S. SHAPIRO
The University of Michigan

VAN NOSTRAND REINHOLD COMPANY
NEW YORK CINCINNATI TORONTO LONDON MELBOURNE

Van Nostrand Reinhold Company Regional Offices:
Cincinnati, New York, Chicago, Millbrae, Dallas

Van Nostrand Reinhold Company Foreign Offices:
London, Toronto, Melbourne

 Manufactured in the United States of America.

Published by Van Nostrand Reinhold Company
450 West 33rd Street, New York, N.Y. 10001

Published simultaneously in Canada by
D. Van Nostrand Company (Canada), Ltd.

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

PREFACE

The present book is a revised and expanded edition of our notes [47], based on lectures given at the Institute of Mathematical Sciences, Madras, in January and February of 1967. The unifying theme is the relation between a given function and its "smoothed" version, obtained by forming its convolution with $K_\lambda(x) = \lambda K(\lambda x)$ where the "kernel" K is an integrable function satisfying $\int_{-\infty}^{\infty} K(x)dx = 1$ and λ is a "large" parameter. Special as it is, this problem nevertheless subsumes much attractive material, e.g., the limiting behavior of harmonic functions in a half plane and of solutions of the heat equation, as well as the classical theory (Jackson, S. Bernstein) of the approximation of periodic functions by trigonometric polynomials.

(Incidentally, it seems to us that the infinite line, on which we shall work exclusively, is in many respects a more favorable carrier on which to study the approximation of periodic functions than the circle group. This purely formal difference is not so trivial as one might suppose since, as de la Vallée Poussin seems first to have observed, certain fundamental smoothing operations (notably Fejér's) are representable as convolutions with much simpler kernels on the line than on the circle.)

The subject of smoothing by convolution is a very rich one, and we have been unable to touch upon a number of interesting topics. In these notes we concentrate exclusively on "degree of approximation" results. We have not discussed the extent to which the smoothed function enjoys various qualitative properties (e.g., convexity, general shape of the "graph") of the original, variation-diminishing kernels in the sense of Pólya and Schoenberg, nor the behavior of the iterates of a smoothing transformation. The paper of Schoenberg [44] may serve to orient the interested reader in some of these questions. Neither have we singled out for special study kernels K for which the K_λ form a convolution semigroup, nor various specific kernels (e.g., that of Weierstrass) associated with specific partial differential equations, etc.,except insofar as they serve to illustrate the general theory. For these topics the reader may consult the recent book of Butzer and Berens [3]. Also, we do not discuss the convolution operation as such (i.e., apart from its role in generating approximations), the solution of convolution equations, Wiener's Tauberian theorem, etc. Good guides to the vast literature on these questions are the books of Widder [57], Pitt [42] and Hirschman and Widder [30]. We also do not discuss approximations (e.g., Bernstein polynomials) generated by other kinds of "smoothing" transformations than our canonical one of convolution with K_λ.

Within the confines of the area we *have* singled out, we believe we have given a fuller treatment than is to be found elsewhere. We have thus rounded out in some essential respects a study pioneered by de la Vallée Poussin, Bochner, and Ahiezer, and greatly furthered in recent years especially by P. L. Butzer and his students (for references, see "Notes and Comments"). The most distinguishing characteristic in this line of study is the considerable role played by Fourier transforms.

We have in some measure preserved the discursive style of the lecture hall rather than write a systematic treatise, and allowed ourselves the luxury of proving some theorems twice, and even (horrors!) presented an unsuccessful attempt to prove one theorem (Theorem 26), the better to appreciate the ingenuity of Serge Bernstein. We have tried to keep the material accessible to a student who has had a course in Lebesgue integration, except that in Chapter V a fair amount of harmonic analysis is needed. The relevant chapters of Rudin's "Real and Complex Analysis" should adequately prepare a prospective reader, and for our first four chapters, familiarity with Chapters X, XI and XII of Titchmarsh's "Theory of Functions" (the "bible" of my generation) is more than sufficient. We think the subject treated here should prove rewarding to a student interested in analysis, who has had some basic courses and wants to see what one can "do" with the standard tools. Here he can gain a proficiency in handling the basic convergence theorems for the Lebesgue integral, Fubini's theorem, and Fourier transforms of functions and measures. At some points (especially in §5.4), he will catch glimpses of the top of an iceberg marked "distributions." In the "Notes and Comments" sections he can find references to related literature.

Since we have stressed *methods* more than encyclopedic completeness we have avoided inessential complications. Our discussion is limited to functions on the real line, and (except for a few scattered results) to *uniform*, or sup norm, approximation. Nevertheless, much of the theory presented here can quite easily be extended both to L^p norms and to several variables; indeed to prepare the way for such extensions we have sought to present the results in the most unified possible way. Doubtless there is much fundamental research yet to be done in this field, and we have indicated here and there what seem to us to be promising areas.

Acknowledgements. The author wishes to thank Dr. Alladi Ramakrishnan, director of the Institute of Mathematical Sciences in Madras, for his kind invitation to the author to lecture at the Institute. That lecture series was the initial stimulus to writing this book and our sojourn in Madras provided the tranquillity to think through some thorny points. Mr. G. N. Keshavamurthy, Mr. M. R. Subrahmanya, and especially Dr. K. R. Unni graciously assisted the author in the preparation of the original notes. We wish to thank them, as well as Dr. R. J. Nessel, who kindly read the manuscript and made a number of corrections. The revision was prepared during the summer of 1968, with financial assistance from the National Science Foundation.

CONTENTS

CHAPTER I

INTRODUCTION

§1.1. *What is "approximation theory?"*

The theory of approximation of functions, even if we restrict attention to functions of one real variable, is an extremely rich and variegated one. We may separate it into at least three main divisions, which can be termed "best approximation," "good approximation," and "possible approximation (or, problems of completeness)." Problems of the first kind deal with the search for a function in a prescribed class which has the least deviation from a given function, as measured in a prescribed metric, for example, one may seek a polynomial of degree n whose maximum deviation from the exponential function on $[-1, 1]$ is as small as possible. The main techniques available for the solution of such problems are *variational*, as one would expect with extremal problems. As a rule, explicit solutions are not possible, and the theory of "best approximation" is largely concerned with questions of existence, uniqueness, and qualitative properties of the minimizing function. There is, however, a computational side of the theory which seeks algorithms that grind out a sequence of functions which converge to a minimizing function, estimates the rapidity of convergence, etc. Much of the theory of "best approximation" lends itself nicely to an abstract treatment, based on Banach spaces.

Concerning problems of "good approximation," one usually studies the degree of approximation to a given function by functions of a prescribed class which are generated in some systematic way by operations performed on the given function, e.g., processes of interpolation, integral transforms, expansion in a Fourier series, etc. Whereas an *absolute* best approximating function from the prescribed class may be very difficult to find, experience shows that an approximation to within, say, twice as large an error as the optimum possible may often be generated rather easily by the aid of operations of the above kinds. Moreover, the behavior of systematically generated approximations, such as the Fourier sums, is an important object of study in its own right. To the area of "good approximation" we should also reckon the "inverse theorems" of approximation theory, on the one hand "saturation theorems," which demarcate the intrinsic limitations of specific approximation processes, and the "Bernstein-type" theorems where one infers a certain degree of smoothness of a given function from the fact that it is approximable to such and such a degree by functions in prescribed classes. In discussing "good approximation," we should also mention problems of goodness in a *qualitative* sense, i.e., generating approximations which "conform to" a given function in a sense beyond mere proximity, e.g., exhibit convexity in the same sense, approach from above, etc.

The third area, completeness problems, deals with the mere *possibility* of approximating, i.e., the question whether a certain class of functions is dense in some larger class. The prototypical theorem here is that of Weierstrass (and its generalization by Stone), that with polynomials one can uniformly approximate an arbitrary continuous function on a bounded interval. The proofs of completeness theorems are often non-constructive, based on the use of linear functionals (e.g. the complex variables proof of Müntz's theorem).

In such cases it is an interesting task to look for constructive proofs, estimate the degree of approximation, etc.

Needless to say, the above separation is only schematic and should be taken with the proverbial grain of salt. Thus, for instance, Favard, Achieser and Krein showed how to prove Jackson's theorem in the uniform metric (the prototypical theorem on "good approximation") even in a much sharpened form, by first finding the *best* approximation, in the dual (L^1) metric, of a specific kernel. In fact, the cited investigations opend up a new chapter in approximation theory, that of the best approximation of a *class* of functions, rather than an individual function. For further details we must refer the reader to Chapter 8 of Lorentz's book [38].

Now that we have whetted the reader's appetite, we must unfortunately retire to the more confined area to which this book is devoted.

§1.2. *An example.*

Before going into the details of a general theory, we shall give an example to show how we can construct *continuous* functions to approximate a given integrable function. We formally state our problem as follows:

Problem. Given $f \in L^1(-\infty, \infty)$, can we approximate f by continuous functions in the L^1 metric?

The answer to this question is in the affirmative:

THEOREM 1. *Let $f \in L^1(-\infty, \infty)$. Then, given $\varepsilon > 0$ there exists a continuous function $g \in L^1(-\infty, \infty)$ such that*

$$\|f-g\|_1 = \int_{-\infty}^{\infty} |f(x)-g(x)|\, dx < \varepsilon .$$

Proof: Let a be a positive real number. Define f_a, the *symmetric moving average* of f, by the formula

$$f_a(x) = \frac{1}{2a}\int_{x-a}^{x+a} f(t)dt = \frac{F(x+a)-F(x-a)}{2a}$$

where

$$F(x) = \int_{-\infty}^{x} f(t)dt \ .$$

It is easy to check that $f_a \in L^1(-\infty, \infty)$ (see Theorem 2 below) and it is continuous (in fact, it is absolutely continuous) and from integration theory we know that $\lim_{a \to 0} f_a(x) = f(x)$ a.e. Thus we already see that f is the limit a.e. of a sequence of continuous functions. To prove the theorem, it is convenient to rewrite f_a as follows:

$$f_a(x) = \frac{1}{2a}\int_{-a}^{a} f(x+t)dt$$

so that

(1) $$f_a(x) = \int_{-\infty}^{\infty} f(x+t)G_a(t)dt$$

where

$$G_a(x) = \begin{cases} \dfrac{1}{2a} & \text{for } |x| \leq a \\ 0 & \text{for } |x| > a \end{cases}$$

Since $\int_{-\infty}^{\infty} G_a(t)dt = 1$, it follows that

$$f_a(x) - f(x) = \int_{-\infty}^{\infty} f(t+x)\, G_a(t)dt - \int_{-\infty}^{\infty} f(x)\, G_a(t)dt$$

$$= \int_{-\infty}^{\infty} [f(x+t) - f(x)]\, G_a(t)dt$$

so that

$$\|f_a - f\|_1 = \int_{-\infty}^{\infty} |f_a(x) - f(x)|\, dx$$

$$\leq \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(x+t) - f(x)|\, G_a(t)dt \right) dx$$

$$= \int_{-\infty}^{\infty} G_a(t)dt \left(\int_{-\infty}^{\infty} |f(x+t) - f(x)|\, dx \right)$$

(by Fubini's theorem)

$$= \int_{-\infty}^{\infty} G_a(t)\, \omega(t)dt$$

where

(2) $$\omega(t) = \int_{-\infty}^{\infty} |f(x+t) - f(x)|\, dx \ .$$

It is easy to verify that $\omega(t) \to 0$ as $t \to 0$. (As this is a standard exercise from real analysis we omit the proof.) Thus we have

$$\|f_a - f\|_1 \leq \int_{-\infty}^{\infty} \omega(t)\, G_a(t)dt = \frac{1}{2a} \int_{-a}^{a} \omega(t)dt < \varepsilon$$

if a is small. This completes the proof of Theorem 1. A more general result will be proved in the next chapter (Theorem 2).

§1.3 *Convolutions*

Let us now generalize from this example. It is convenient to make a change of variable and write (1) in the form

$$f_a(x) = \int_{-\infty}^{\infty} f(x-t)\, G_a(t)\, dt \ .$$

This suggests the following definition

DEFINITION. Let $f, g \in L^1(-\infty, \infty)$. We define the *convolution* $f*g$ of f and g by the formula†

$$(f*g)(x) = \int_{-\infty}^{\infty} f(x-t)\, g(t)dt \ .$$

We can easily verify the following properties of the convolution product

(i) $f*g \in L^1(-\infty, \infty)$. In fact, $\|f*g\|_1 \leq \|f\|_1 \, \|g\|_1$;

(ii) $f*g = g*f$;

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

for $f, g, h \in L^1(-\infty, \infty)$.

From the definition, we immediately notice that f_a is the convolution product of f and G_a. Moreover, if we set

$$K(x) = \begin{cases} \frac{1}{2}, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases}$$

then

$$\int_{-\infty}^{\infty} K(x)dx = 1$$

† This formula is meaningful also with other hypotheses, for example $f \in L^1$ and g bounded, and we continue to speak of it as a convolution in those cases. Properties (ii), (iii) below remain valid.

and writing, for $\lambda > 0$,

$$K_\lambda(x) = \lambda K(\lambda x) \quad ^\dagger$$

we have

$$G_a(x) = K_\lambda(x) \,, \qquad \lambda = \frac{1}{a} \,.$$

Thus the "moving average" method of approximation is characterized by a certain function ("kernel") $K(x)$. Writing $f(x;\lambda)$ for $f_a(x)$, we have

$$f(x;\lambda) = (f * K_\lambda)(x) = \int_{-\infty}^{\infty} f(x-t)\,\lambda K(\lambda t)\,dt \quad ^{\dagger\dagger}$$

$$= \int_{-\infty}^{\infty} f\left(x - \frac{t}{\lambda}\right) K(t)\,dt \,.$$

In the remainder of these lectures we shall examine the behavior of $f(x;\lambda)$ in relation to f for quite general kernels $K \in L^1(-\infty,\infty)$. Notice that there is no need in general to assume $f \in L^1$ since $f * K$ is well defined if f is merely bounded when $K \in L^1$ and (in the above example where K has compact support) even if f is locally integrable. For definiteness, we shall mostly study the problem of approximating to a *continuous bounded* f by $f(x;\lambda)$ in the *supremum norm,* but the techniques developed below apply *mutatis mutandis* to L^p-norms as well.

Especially we shall be interested in estimating the error, or deviation, $f(x;\lambda) - f(x)$ under varying hypotheses on f and K (*direct theorems* of approximation theory); on intrinsic limits to the goodness of the approximation imposed by the peculiarities of the

† Throughout these notes the notation $K_\lambda(x) = \lambda K(\lambda x)$ will be adhered to.

†† We shall always denote $(f * K_\lambda)(x)$ either by $f^\lambda(x)$ or $f(x;\lambda)$.

kernel (*saturation theorems*); and on inference of smoothness properties of f from the smallness of the deviation $f(x;\lambda)-f(x)$ (*inverse theorems*). It will turn out that these questions are decisively influenced by the behavior of the Fourier transform $\hat{K}(x)$ of $K(t)$, notably its *flatness* at $x=0$. By specializing to the case where f has period 2π and K is a "low frequency function" (that is, $\hat{K}$ vanishes outside an interval) we shall obtain theorems on the approximation of periodic functions by trigonometric polynomials.

Examining the graph $y=\lambda K(\lambda x)$ for $\lambda\to\infty$ for typical K, we see that K_λ is a "peaking kernel" i.e., shows the qualitative behavior of a convolution identity or "Dirac delta function." (Of course, more general peaking kernels $K(x;\lambda)$ could be studied which have a more complicated functional dependence on the parameter λ. Our kernels have the advantage that they are generated by scale change from a function $K(x)$ of *one* variable.) Dually, observe that the Fourier transform of K_λ is $\hat{K}(x/\lambda)$, which for large λ is approximately equal to one over a very wide range, imitating the identically constant Fourier transform of the delta function.

§1.4 *Convolutions, cont'd.*

Our main purpose, as we have said, is to illustrate the use of convolutions in approximation problems. In particular, what is perhaps the most powerful and general known method for generating approximations to f may be summarized thus: "convolve f with a peaking kernel." The reasons for the dazzling versatility of this method may be summed up as follows:

a) Convolution is a *smoothness–increasing operation*. That is, if g is integrable and of norm one, $f*g$ is at least as smooth as f by just about any conceivable test (modulus of continuity, moduli of smoothness of higher order, number of derivatives, total variation, etc.). This isn't too surprising perhaps if we think of convolution as a (generalized) moving average.

b) Various special structural properties of a function f (e.g., having a given period, or being a trigonometric polynomial of degree not exceeding n) are likewise inherited by $f*g$.

At bottom a) and b) are the same: very roughly, they say that properties based on the *translation group* are *hereditary* under convolution. And because of the commutativity of convolution, their presence in *either factor* ensures their presence in the convolution product. Thus, convolution is like a marriage in which (unlike real life) the "best" properties of each parent are inherited by the offspring (i.e. differentiability, periodicity, etc. are "dominant genes"). Thus, suppose a lowly bounded measurable function f is convolved with an integrable function g which happens to have 100 derivatives. The resulting function has again 100 derivatives, but moreover resembles f if g is chosen to be a peaking kernel (for instance, if g is K_λ for large λ and suitable K, then $f*K_\lambda$ tends almost everywhere to f). If f moreover has period 2π, so have all the approximating functions. And if, in addition, $\hat{K}(x) = 0$ for $|x| \geq 1$ (so that K_λ has a Fourier transform vanishing for $|x| \geq \lambda$), this property too is inherited by $f*K_\lambda$ which must, therefore, be a trigonometric polynomial† of degree less than λ.

Moreover, convolutions have other properties which make them technically very nice to work with. For instance, if f and g are differentiable we can for the derivative of $f*g$ take our choice of the expressions $f'*g$ and $f*g'$. For higher order derivatives there is still greater freedom. Also the close tie-in with Fourier

† For precise formulation, see Chapter IV.

transforms puts powerful techniques from harmonic analysis at our disposal. Furthermore, the asymptotic behavior of convolutions is often easy to estimate.

In addition, it turns out (although this is far from obvious *a priori*) that under suitable restrictions the operations of passing from a function to its derivative or its (suitably normalized) primitive may be realized as convolutions with suitable kernels. These facts enhance the importance of convolutions, and play an essential role in the theory which follows.

Finally, although this plays only a minor role in the present book, the notion of convolution admits of far reaching and fruitful generalization: not only can functions be convolved with one another, but more general entities (functionals) can meaningfully be convolved with functions, and under suitable circumstances, with one another. Of course, such "convolutions" cannot any longer be interpreted as "smoothing" operations.

CHAPTER II

SOME ELEMENTARY THEOREMS ON APPROXIMATION

§2.1 *Some applications of convolutions to approximation.*

In this section we shall give some simple applications of the technique of smoothing a function by convolution. First of all, we shall generalize the result found for moving averages in Chapter I.

THEOREM 2. *Let $K \in L^1(-\infty, \infty)$ and suppose $\int_{-\infty}^{\infty} K(x)dx = 1$. If $K_\lambda(x) = \lambda K(\lambda x)$ where $\lambda > 0$ and $f \in L^1(-\infty, \infty)$ then*

(i)† $$f^\lambda = f * K_\lambda \in L^1$$

and

(ii) $$\|f^\lambda - f\|_1 \to 0 \text{ as } \lambda \to \infty \ .$$

Proof: (i) follows from property (i) of convolutions (§1.3).
Next, we have, setting $a = \lambda^{-1}$

$$f^\lambda(t) = \int f(t - au) K(u) du$$

$$\|f^\lambda - f\|_1 \le \iint |f(t - au) - f(t)| \cdot |K(u)| dt\, du$$

$$= \int \omega(au) |K(u)| du$$

† We wish to emphasize that the superscript λ is not an exponent: f^λ is just a designation for the function whose value at x is $f(x; \lambda) = (f * K_\lambda)(x)$. This notation is adhered to throughout these notes.

where ω is defined by 1.2(2). (Here and in the sequel the range of integration is $(-\infty, \infty)$ if not otherwise indicated.) Since $\lim_{a \to 0} \omega(au) = 0$ for each fixed u, and ω is bounded, the dominated convergence theorem implies that the last integral tends to zero as $a \to 0$, and the theorem is proved.

We can now greatly strengthen Theorem 1.

COROLLARY. *Under the hypotheses of Theorem 1, we may take for g the restriction to the real axis of an entire analytic function.*

Proof: (Sketched only.) We take $K(x) = (1/\sqrt{\pi})\, e^{-x^2}$. Then all the f^λ are (restrictions to the real axis of) entire analytic functions: for

$$F(z) = \frac{\lambda}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(t)\, e^{-\lambda^2(z-t)^2} dt$$

is defined as a Lebesgue integral for every complex z (as one easily checks), and $F(x) = f^\lambda(x)$ when x is real. Moreover, writing

$$F = F_n + G_n \quad \text{where } F_n = \int_{-n}^{n} \quad \text{and } G_n = \int_{|t|>n}$$

we see that F_n is entire, and moreover for every $R > 0$, $\max_{|z| \le R} |G_n(z)|$ tends to zero as $n \to \infty$. Hence F is entire, being a uniform limit of entire functions, on every compact set. In other words, f^λ is entire, and furnishes the required approximation.

Exercises

1. Prove that, for $f \in L^1$, $\lim_{t \to 0} \int |f(x+t) - f(x)|\, dt = 0$.

(Hint: The class of f for which this is true contains all step functions, and is a closed subset of L^1.)

2. Extend Theorem 2 and the corollary to L^p norm ($p > 1$).

THEOREM 3. *Let K be as in Theorem 2. If f is uniformly continuous and bounded on $(-\infty, \infty)$, then*

$$\sup_x |f(x) - f(x; \lambda)| \to 0 \quad \text{as } \lambda \to \infty .$$

Proof: As before, we have (with $a = \lambda^{-1}$)

$$f(x) - f(x; \lambda) = \int_{-\infty}^{\infty} [f(x) - f(x - at)] K(t) dt$$

so that

$$|f(x) - f(x; \lambda)| \leq \int_{-\infty}^{\infty} \omega(at) |K(t)| dt$$

where

$$\omega(t) = \sup_x |f(x-t) - f(x)| .$$

Since f is uniformly continuous and bounded, it follows that $\omega(t) \to 0$ as $t \to 0$ and $\omega(t) \leq C$. The proof is now completed by a reasoning similar to the one used in Theorem 2.

COROLLARY 1. *If f is uniformly continuous and bounded on $(-\infty, \infty)$, it can be uniformly approximated as closely as desired by (the restriction to the real axis of) an entire analytic function.*

REMARK 1. Actually the hypotheses of *uniform* continuity, and boundedness, in the corollary are not essential. Carleman has shown that any continuous function can be approximated uniformly on the whole real axis by (the restriction to the real axis of) an entire function. This theorem requires other methods for its proof.

On the other hand, a closer analysis shows that our approximating entire functions are of order at most two, and one can obtain a still stronger result of this kind, with approximating functions of *exponential type*. Such results are not valid for arbitrary continuous functions.

REMARK 2. If f is assumed only continuous (but not necessarily uniformly) and bounded, we can nevertheless assert $f^{\lambda} \to f$ uniformly on each bounded interval; we leave the necessary modification of the proof to the reader. (It also follows from the "localization theorem," Theorem 24 in §4.5 below.)

COROLLARY 2. (Weierstrass) *A continuous function defined on a closed and bounded interval can be approximated uniformly by means of polynomials.*

Proof: Extending the function so as to be constant outside its interval of definition, we get an *entire* approximation to it by Corollary 1 and the entire function can in turn be uniformly approximated by a partial sum of the Taylor series on a bounded interval.

§2.2 *A theorem on pointwise convergence.*

We start with a rather general result, involving a *Lebesgue-Stieltjes convolution,* which has useful applications.

THEOREM 4. *Let* σ *denote a function of finite total variation on* $(-\infty, \infty)$. *Suppose* K *is continuous and even, of finite total variation on* $(0, \infty)$, $\lim_{x\to\infty} xK(x) = 0$, *and* $\int_0^\infty x\,|dK(x)| < \infty$. *Let*

$$(1) \qquad f^{\lambda}(x) = \int_{-\infty}^{\infty} K_{\lambda}(t)\,d\sigma(x+t) .$$

Then

(i) *at every point* x *where* σ *possesses a finite symmetric derivative, i.e.,*

$$\lim_{t \to 0} \frac{\sigma(x+t) - \sigma(x-t)}{2t} = \sigma'(x)$$

exists and is finite, we have

$$\lim_{\lambda \to \infty} f^{\lambda}(x) = \sigma'(x) \int_{-\infty}^{\infty} K(t)dt \tag{2}$$

(ii) *If, moreover,* K *is non-increasing on* $(0, \infty)$ $K \not\equiv 0$, *then* (2) *holds also at points where* σ *is continuous and*

$$\sigma'(x) = \pm \infty .$$

REMARK 1. Wherever σ possesses a finite or infinite (ordinary) derivative, the symmetric derivative exists and has the same value (but not conversely). In particular, σ has almost everywhere a finite symmetric derivative.

REMARK 2. The integrability of K follows from the hypotheses.

REMARK 3. Important examples of kernels K satisfying the hypotheses are:

$$K(t) = \frac{1}{\pi} \cdot \frac{1}{1+t^2} \qquad \text{(Cauchy kernel)}$$

$$K(t) = \frac{1}{\sqrt{\pi}} e^{-t^2} \qquad \text{(Weierstrass kernel)}$$

$$K(t) = \frac{1}{2} e^{-|t|} \qquad \text{(Picard kernel)}$$

On the other hand,

$$K(t) = \frac{1}{\pi} \left(\frac{\sin t}{t}\right)^2 \qquad \text{(Fejér-de la Vallée Poussin kernel)}$$

fails to satisfy the hypotheses, and

$$K(t) = \frac{3}{\pi}\left(\frac{\sin t}{t}\right)^4 \qquad \text{(Jackson-de la Vallée Poussin kernel)}$$

satisfies the hypotheses for (i), but not for (ii).

Exercise. Compute the Fourier transform of each of these kernels (these will be needed later).

Proof of Theorem 4.

$$f^\lambda(x) = \left(\int_{-\infty}^{0} + \int_{0}^{\infty}\right) K_\lambda(t)\,d\sigma(x+t)$$

$$= \int_0^\infty K_\lambda(t)\,d[\sigma(x+t) - \sigma(x-t)]$$

$$= K_\lambda(t)\,[\sigma(x+t) - \sigma(x-t)]\Big|_0^\infty - \int_0^\infty [\sigma(x+t) - \sigma(x-t)]\,dK_\lambda(t)$$

In both cases (i) and (ii) the first term on the right in the last equation vanishes and

$$f^\lambda(x) = -\int_0^\infty [\sigma(x+t) - \sigma(x-t)]\,\lambda\,dK(\lambda t) \; .$$

A change of variables gives

$$\text{(3)} \qquad f^\lambda(x) = -2\int_0^\infty \frac{\sigma(x+au) - \sigma(x-au)}{2au} \cdot u\,dK(u) \qquad (a = \lambda^{-1}) \, .$$

We now distinguish cases, taking case (i) first. The first factor in the integrand tends to $\sigma'(x)$ as $a \to 0$, for each fixed u, and moreover remains less than a finite bound for all values of a and u. Moreover, $u\,dK(u)$ is, by hypothesis, a finite (signed) measure on

$(0, \infty)$. Therefore, by the bounded convergence theorem

$$\lim_{\lambda \to \infty} f^{\lambda}(x) = -2\sigma'(x) \int_0^{\infty} u dK(u) = \sigma'(x) \cdot 2 \int_0^{\infty} K(u) du ,$$

which proves (2).

Case (ii): now we suppose K non-increasing on $(0, \infty)$, and $\sigma'(x) = +\infty$. We have, from (3)

$$f^{\lambda}(x) = \int_0^{\infty} S_a(u)(-2u dK(u))$$

where $S_a(u)$ denotes $(\sigma(x+au) - \sigma(x-au))/2au$. Now, $S_a(u)$ is bounded *below* (for fixed x) uniformly in a and u, and so, for some positive constant A, $S_a(u) + A \geq 0$. Hence

$$f^{\lambda}(x) = A \int_0^{\infty} 2u dK(u) + \int_0^{\infty} [S_a(u) + A](-2u dK(u)) .$$

The first term on the right is finite. As for the second, $-2u dK(u)$ is a finite *positive* measure on $(0, \infty)$, and $S_a(u) + A$ is non-negative, and tends to $+\infty$ as $a \to 0$, for each fixed u. Therefore, by Fatou's lemma, the second integral tends to $+\infty$ as $a \to 0$, which establishes (ii).

COROLLARY 1. *If $f \in L^1$ and K has integral one and satisfies the hypotheses of the theorem, $f^{\lambda}(x) \to f(x)$ for almost all x, where $f^{\lambda} = f * K_{\lambda}$.*

Proof: Apply Theorem 4 to $\sigma(x) = \int_{-\infty}^{x} f(x) dt$.

REMARK. Since

$$\frac{\sigma(x+u) - \sigma(x-u)}{2u} = \frac{1}{2u} \int_{x-u}^{x+u} f(t) dt \to f(t)$$

as $u \to 0$ if f is continuous at t, we see that $f^{\lambda}(x) \to f(x)$ at each point x where f is continuous. It is also readily verified that if f is continuous on an open interval, the convergence is uniform on each compact subinterval.

Similarly, if f has limiting values at x both from the left and from the right, $f^{\lambda}(x)$ tends to half the sum of these values.

An important application of Theorem 4 is to the limiting behavior of the Poisson integral:

COROLLARY 2. *Let* $u(x, y)$ *be a harmonic function in* $y > 0$ *and non-negative. Then* $\lim_{y \to 0+} u(x, y)$ *exists and is finite a.e.*

Proof: (Note: For the preliminary notions concerning harmonic functions which we require in this proof, and in a few other illustrative remarks in these lectures, the reader may consult Tsuji [55], Chapter 4.) By a theorem of Herglotz there exists a non-decreasing function σ and a real number $c \geq 0$ such that $\int (d\sigma(t))/(1+t^2)$ is finite, and

$$(4) \qquad u(x, y) = cy + \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{d\sigma(t+x)}{t^2+y^2}$$

Let us first consider the case where $\int d\sigma < \infty$. To complete the proof for this case, it is enough to show that the integral in (4) is expressible in the form (1) by choice of a suitable kernel K; then Theorem 4 will give the result. To this end, we take $K(x) = [\pi(1+x^2)]^{-1}$, $\lambda = y^{-1}$. Then $K_\lambda(x) = \lambda K(\lambda x) = y[\pi(x^2+y^2)]^{-1}$, and we are done. In the general case, choose a finite interval (a, b) and split $\sigma = \sigma_1 + \sigma_2$ where σ_1 is constant outside (a, b) and σ_2 vanishes in (a, b). This splits the Poisson-Stieltjes integral in (4) into a sum of two harmonic functions $u_1 + u_2$, the second of which tends to zero as $y \to 0$ for $a < x < b$, and the first of which is covered by the case already handled. (The reader can fill in the details.)

REMARK. The theory of boundary behavior of harmonic functions in a half-plane is, in large measure, the study of the integral appearing in (4). It is of interest that many of the typical results of that theory carry over to the corresponding integral formed with other kernels (the carry-over is strongest to so-called semigroup kernels).

COROLLARY 3. *(Uniqueness theorem for the Fourier transform). Let*

$$f(x) = \int_{-\infty}^{\infty} e^{itx} g(t)dt \tag{5}$$

where g is any function in $L^1(-\infty, \infty)$. If $f(x) = 0$ for all x, then g is almost everywhere equal to zero.

Proof: Multiply (5) by $e^{-a|x| - ixu}dx$ where $a > 0$, u is real and integrate. Then

$$0 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{itx}e^{-a|x| - ixu}g(t)dt\,dx$$

$$= \int_{-\infty}^{\infty} G(t)\,g(t)dt, \text{ by Fubini's theorem}$$

where

$$G(t) = \int_{-\infty}^{\infty} e^{-a|x|}e^{ix(t-u)}dx$$

$$= 2\int_{0}^{\infty} e^{-ax}\cos x(t-u)dx$$

$$= \frac{2a}{a^2 + (t-u)^2}$$

Hence

$$\int_{-\infty}^{\infty} \frac{2a}{a^2+(t-u)^2}\, g(t)dt = 0 .$$

We may write this as

$$\int_{-\infty}^{\infty} K_\lambda(t)\, g(t+u)dt = 0 .$$

where $\lambda = 1/a$, $K(x) = (1/\pi)(1/(1+x^2)$. Now, let $\lambda \to \infty$. Corollary 1 gives $g(u) = 0$ almost everywhere.

Exercises 1. Prove, if $\int e^{itx} d\sigma(t) = 0$ for all x, where σ is a finite measure on the line, then σ is the zero measure.

2. Let f be non-decreasing on $(-\infty, \infty)$. Then, there exists a sequence of non-decreasing functions $f_n \in C^\infty$ such that $f_n(x) \leq f_{n+1}(x) \leq f(x)$ for all n, and $f_n(x) \to f(x)$ at all points of continuity of f; and a similar sequence of non-decreasing functions $g_n \in C^\infty$ such that $f(x) \leq g_{n+1}(x) \leq g_n(x)$.

3. Let f be continuous and convex on $(-\infty, \infty)$. There exists a sequence of convex functions $f_n \in C^\infty$ such that $f_n(x) \geq f_{n+1}(x) \geq f(x)$, and $f_n(x) \to f(x)$ for all x.

4. Can we get similar approximation by an *increasing* sequence in Exercise 3?

5. Can the sequences in Exercises 2 and 3 be chosen to consist of *entire* functions?

In concluding this section, we remark that for many kernels which do *not* satisfy the hypotheses of Theorem 4 (e.g., the Fejér-de la Vallée Poussin kernel) the relation $f * K_\lambda \to f$ a.e. none the less holds for integrable (and even, bounded measurable) functions f. The interested reader may consult Chapter I, §7 of Reference [7].

§2.3 *Degree of approximation.*

Having now illustrated various kinds of approximation (L^1, uniform, pointwise almost everywhere) we shall, in the remainder of these lectures, concentrate on uniform approximation of continuous functions.

In this section we shall study how *rapidly* the function f can be approximated by a given method, in terms of λ.

DEFINITION. We say that $f \in \text{Lip}\ \alpha$, $0 < \alpha \leq 1$, if there exists a constant C such that

$$|f(x) - f(y)| \leq C|x-y|^{\alpha} .$$

When $\alpha = 1$, this is equivalent to saying that f is absolutely continuous and $|f'(x)| \leq C$ a.e. The definition would be meaningful for $\alpha > 1$, but useless, since in that case f must be constant.

If $f \in \text{Lip}\ \alpha$, then

$$|f(x) - f^{\lambda}(x)| = \left| \int_{-\infty}^{\infty} [f(x) - f(x - \frac{t}{\lambda})] K(t) dt \right|$$

$$\leq C\lambda^{-\alpha} \int_{-\infty}^{\infty} |t|^{\alpha} |K(t)| dt .$$

Note. In the sequel, C, C_1, C_2, A, A_1, A_2, ... denote constants whose precise value shall not concern us. All norms $\|\cdot\|$ are *uniform* (sup) norms, unless otherwise indicated by a subscript. Moreover, we tacitly assume, unless the contrary is asserted, that f is *continuous* on $(-\infty, \infty)$ and *bounded.*

The preceding inequality may be formulated as

THEOREM 5. *If*

(6) $$M = \int_{-\infty}^{\infty} |x|^{\alpha} |K(x)| dx < \infty, \quad 0 < \alpha \leq 1$$

then $f \in \text{Lip}\ a$ *implies*

$$\|f^{\lambda} - f\| \leq \frac{CM}{\lambda^{a}} \quad .$$

REMARK. If we don't wish to assume K satisfies (6), we can proceed as follows:

$$|f^{\lambda}(x) - f(x)| \leq \left| \left(\int_{-T}^{T} + \int_{|t| \geq T} \right) \left[f\left(x - \frac{t}{\lambda}\right) - f(x) \right] K(t) dt \right|$$

$$\leq C\lambda^{-a} \int_{-T}^{T} |t|^{a} |K(t)| dt = C_1 \int_{|t| > T} |K(t)| dt$$

where T is a positive number which we are free to choose optimally. For example, if $K(t) = O(t^{-2})$ we see (choosing $T = \lambda$) that $f \in \text{Lip}\ 1$ implies $\|f^{\lambda} - f\| = O\left(\frac{\log \lambda}{\lambda}\right)$. This is a case of some importance, since it covers the Cauchy and Fejér-de la Vallée kernels. For these kernels, it can be shown by examples that the degree of approximation to a Lip 1 function needn't be of smaller order of magnitude than $\frac{\log \lambda}{\lambda}$ (for the Picard and Weierstrass kernels, on the other hand, it is $O(1/\lambda)$, in view of Theorem 5).

A more comprehensive discussion of degree of approximation, especially regarding functions of smoothness *beyond* Lip 1, will be given later, in Chapter 4.

Exercise. Establish the above assertions about the Fejér-de la Vallée Poussin and Cauchy kernels.

CHAPTER III

SOME SATURATION THEOREMS

§3.1 *An example*

One of the most important phenomena of approximation theory, although its consideration from a general viewpoint (by Favard) is of recent origin, is "saturation." It may happen for a fixed kernel that a degree of approximation beyond a certain critical level is possible only for "trivial" functions such as constants, linear functions etc. (the "trivial" class may vary from case to case). In general, as we have seen, the degree of approximation improves with the *smoothness* of the function. But there may be a limit beyond which even if we presuppose greater smoothness of the function being approximated we don't get better approximation. This phenomenon, called "saturation," is the subject matter of the present chapter.

Let us first illustrate the saturation phenomenon for moving averages. Suppose we seek to approximate the nice function

$$f(x) = e^{ix}$$

Then, if f_a denotes the symmetric moving average of f with span $2a$,

$$f_a(x) = \frac{1}{2a} \int_{x-a}^{x+a} e^{it} dt = e^{ix} \cdot \frac{e^{ia} - e^{-ia}}{2ia}$$

so that

$$f(x) - f_a(x) = e^{ix}(1 - \frac{\sin a}{a})$$

hence

$$\|f - f_a\| = 1 - \frac{\sin a}{a} \sim \frac{a^2}{6} \text{ for small } a.$$

Now, $f(x)$ is as smooth a function as we could wish, and yet we get an error in the approximation which has the order a^2. (We shall see that this "quadratic law" is shared by a large class of kernels).

We can also get at the saturation phenomenon in another way. Let us show:

Suppose $f''(x)$ exists. Then

$$\lim_{a \to 0} \frac{f_a(x) - f(x)}{a^2} = \frac{f''(x)}{6} .$$

Proof: By hypothesis, we have

(1) $$f(x+t) - f(x) = tf'(x) + \frac{t^2}{2!} f''(x) + \varepsilon_x(t)$$

where

$$\frac{\varepsilon_x(t)}{t^2} \to 0 \text{ as } t \to 0 .$$

Now

$$f_a(x) - f(x) = \frac{1}{2a} \int_{-a}^{a} [f(x+t) - f(x)]dt$$

$$= \frac{f''(x)}{2} \cdot \frac{1}{2a} \int_{-a}^{a} t^2 dt + o(a^2) \text{ (using (1))},$$

$$= \frac{a^2}{6} f''(x) + o(a^2) .$$

Now, dividing by a^2 and letting $a \to 0$, the result follows. It is not hard to deduce from this the following results:

(i) if $\|f_a - f\| = o(a^2)$, then f is a linear polynomial;

(ii) if $\|f_a - f\| = O(a^2)$, then for all $h > 0$ and real x

$$|f(x+h) - 2f(x) + f(x-h)| \leq Ah^2 \tag{2}$$

where A is a constant independent of x and h.

We shall not give these deductions now, because we shall shortly carry them out in a more general situation. The statements (i) and (ii) together give the "saturation theory" of the moving average method. Note, that the converse ("direct") assertions are also true, i.e., if f is linear, then $\|f_a - f\| = o(a^2)$, in fact $f_a - f$ vanishes identically in this case; and if f satisfies (2), it is easy to show that $\|f_a - f\| = O(a^2)$. Thus, (i) and (ii) characterize precisely the limitations of the "moving average" method, and the functions for which the optimal degree of approximation (namely, a^2) is attained. We shall (without troubling here to give a *formal* definition of "saturation" but using it merely as a descriptive word), remark only that one can rephrase the results thus:

> "The moving average method is saturated with order a^2 (or λ^{-2}, in terms of our usual parameter $\lambda = a^{-1}$), and the *saturation class* is the set of f satisfying (ii) (which, as we will see later, is the same as the class of absolutely continuous f with $f' \in$ Lip 1)."

Exercise. Show that for the *asymmetric* (or one-sided) moving average (determined by the kernel which is the characteristic function of the interval $[0, 1]$) the order of saturation is λ^{-1} and the saturation class is Lip 1.

We now generalize the preceding considerations. If K is an L^1 kernel and $f(t) = e^{it}$, then

$$(f * K_\lambda)(t) = e^{it}\hat{K}(\tfrac{1}{\lambda})$$

$$\|f^\lambda - f\| = |1 - \hat{K}(\tfrac{1}{\lambda})|$$

where $\hat{K}(x)$ denotes the Fourier transform $\int_{-\infty}^{\infty} K(t)e^{-itx}dt$.

Since it is reasonable to expect that the exponential function is well behaved enough to gain entrance to the saturation class of K (whatever it shall turn out to be), we should speculate that the kernel K is saturated with the order $|1 - \hat{K}(\frac{1}{\lambda})|$. This heuristically arrived at conclusion turns out to be in essence correct (and rigorous formulations will be stated and proved later). The qualitative formulation is the principle: *the "flatter" the graph of $\hat{K}$ is at $x = 0$, the smaller the order of magnitude at which saturation occurs.*

In particular, if $\hat{K}$ has many vanishing derivatives at $x = 0$ (equivalently, if many *moments of* K vanish) K is a "relatively unsaturated kernel." We shall therefore expect such kernels to play a central role in generating optimal or near-optimal approximations to functions of high smoothness. The best behaved kernels from the standpoint of saturation ought to be those for which $\hat{K}(x)$ is identically equal to one in a neighborhood of $x = 0$. On the other hand, a *positive* kernel cannot have a vanishing second moment, so we would suspect that approximation to an order less than λ^{-2} is possible only for trivial functions using such a kernel.

§3.2 *Saturation theorems, "regular" kinds; mollifiers.*

Now, let us prove some theorems which support the preceding rather vague speculations. Our first theorems, which make no use of Fourier transform considerations, generalize the λ^{-2} result noted above for the moving average approximation to a wide class of kernels. We begin with the "local" result.

THEOREM 6. (Pointwise saturation theorem). *Let* $K(x)$ *be integrable,* $\int_{-\infty}^{\infty} K(x)dx = 1$ *and further*

$$x^2 K(x) \in L^1(-\infty, \infty)$$

and

$$\int_0^{\infty} xK(x)dx = 0.$$

Let

$$A = \int_0^{\infty} t^2 K(t)dt .$$

If f *is a bounded measurable function and if* $f''(x)$ *exists, then*

$$\lim_{\lambda \to \infty} \lambda^2 [f^{\lambda}(x) - f(x)] = \frac{Af''(x)}{2}$$

REMARK. By the hypothesis that $f''(x)$ exists, i.e., that f has a second derivative *at* the point x, we mean that $f(x+t)$, as a function of t, has for small $|t|$ the asymptotic behavior of a quadratic polynomial; the coefficient of $t^2/2!$ in this polynomial is then by definition $f''(x)$. (A similar remark applies to a derivative of any order.) In view of Taylor's formula, this definition coincides with the usual in case $f \in C^2$ in the neighborhood of x.

Proof:

$$
\begin{aligned}
f^{\lambda}(x) - f(x) &= \int [f(x - \tfrac{t}{\lambda}) - f(x)]K(t)dt \\
&= \int [f(x-at) - f(x)]K(t)dt \qquad (a = \lambda^{-1}) \\
&= \int [f(x-at) - f(x) + atf'(x)]K(t)dt \\
&= a^2 \int \left[\frac{f(x-at) - f(x) + atf'(x)}{(at)^2}\right] t^2 K(t)dt \quad .
\end{aligned}
$$

Now, the bracketed term in the last integral is bounded for all values of a and t. Moreover, for fixed $t \neq 0$, it tends to $(f''(x))/2$ as $a \to 0$. The theorem now follows by dominated convergence.

DEFINITION. Let S denote the class of functions f such that

$$f(x+h) - 2f(x) + f(x-h) = O(h^2) ,$$

uniformly with respect to x.[1]

THEOREM 7. *Let K be as in Theorem 6, $A \neq 0$, and let f be a bounded continuous function on $(-\infty, \infty)$.*

(i) *If* $\|f - f^{\lambda}\| = o(1/\lambda^2)$ *then f is constant.*

(ii) *If* $\|f - f^{\lambda}\| = O(1/\lambda^2)$ *then $f \in S$.*

This follows from the next theorem in which the situation is localized to an arbitrary interval.

THEOREM 8. (Localized version of Theorem 7). *Let K and f be as in Theorem 7 and let $-\infty \leq a < b \leq \infty$.*

(i) *If* $\max\limits_{a \leq x \leq b} |f(x) - f^{\lambda}(x)| = o(1/\lambda^2)$ *then f is linear in (a, b).*

(ii) *If* $\max\limits_{a \leq x \leq b} |f(x) - f^{\lambda}(x)| = O(1/\lambda^2)$ *then $f \in S$ in (a, b).*

---

[1] See p. 46.

REMARK. Thus, for these kernels (which include all non-negative even kernels with $x^2K \in L^1$, in particular those of Weierstrass, Picard, and Jackson-de la Vallée Poussin) we find the identical saturation behavior as in the case of the symmetric moving average.

Proof: Let J be a kernel in the class C^2 with support contained in $(-1, 1)$. That is, $J(x) = 0$ for $|x| > 1$ and $\int_{-\infty}^{\infty} J(x)\,dx = 1$. We set

$$g^\lambda(x) = f^\lambda(x) - f(x) .$$

Then from (i) it follows that

$$|g^\lambda(x)| \le \frac{\varepsilon(\lambda)}{\lambda^2} \quad \text{for } a \le x \le b$$

where $\varepsilon(\lambda) \to 0$ as $\lambda \to \infty$.

The idea of the proof which follows is, we want to use Theorem 6, except that *a priori* we do not know that f has any smoothness properties, let alone a second derivative. Therefore, we shall first smooth out f by convolving it with a smooth "approximate identity," apply Theorem 6 to the smoothed function, then pass to a limit to obtain the desired conclusion about f. This useful technique is widely employed in the theory of partial differential operators ("method of mollifiers") and other branches of analysis.

We define[2] (where, as usual, $J_\mu(x) = \mu J(\mu x)$)

$$f_\mu = f * J_\mu, \quad f_\mu^\lambda = f^\lambda * J_\mu = (f * K_\lambda) * J_\mu = f_\mu * K_\lambda .$$

Then $f_\mu \in C^2$ and for any $\delta > 0$ we have, when $\mu > 1/\delta$,

$$|(g^\lambda * J_\mu)(x)| = \left| \int_{-\delta}^{\delta} g^\lambda(x-t)\mu J(\mu t)dt \right| \le \frac{\varepsilon(\lambda)}{\lambda^2}$$

$$\text{if } a + \delta \le x \le b - \delta .$$

This implies

$$|f_\mu^\lambda(x) - f_\mu(x)| \le \frac{\varepsilon(\lambda)}{\lambda^2} \quad , \quad a+\delta \le x \le b-\delta$$

which, by Theorem 6, implies that f_μ has a vanishing second derivative, and hence is a linear function on $(a+\delta,\ b-\delta)$. Since f_μ converges uniformly to f, the same is true of f, and finally since δ is arbitrary (i) is proved.

To prove (ii) we proceed similarly, except now we start from

$$|g^\lambda(x)| \le \frac{B}{\lambda^2} \qquad \text{for } a \le x \le b$$

and we conclude, as before,

$$|f_\mu^\lambda(x) - f_\mu(x)| \le \frac{B}{\lambda^2} \, , \quad a+\delta \le x \le b-\delta \, .$$

Therefore, by Theorem 6, $|f_\mu''(x)| \le 2B/|A| = B_1$, hence by the mean value theorem, if $x-h$ and $x+h$ lie in $[a+\delta,\ b-\delta]$,

$$(3) \qquad |f_\mu(x+h) - 2f_\mu(x) + f_\mu(x-h)| \le B_1 h^2 \, .$$

Letting $\mu \to \infty$ we see that f satisfies this last inequality and since δ is arbitrary (ii) is proved.

REMARK. We have already pointed out earlier that S is identical with the class of f having a Lip 1 derivative. To obtain the conclusion in (ii) in this form, observe that f is a uniform limit of the functions f_μ which have uniformly bounded second derivatives on $[a+\delta,\ b-\delta]$. Now, instead of using (3), observe that the derivatives of the f_μ are uniformly equicontinuous; hence some subsequence of $\{f_\mu'\}$ converges uniformly to a function ϕ which is easily seen to be of class Lip 1. It is now easy to check that

$$\int_{x_1}^{x_2} \phi(t)dt = f(x_2) - f(x_1) .$$

The details are left to the reader.

Exercise. Prove that S is identical with the class of absolutely continuous f satisfying $f' \in \text{Lip } 1$. (In one direction this is trivial, because of the identity

$$f(x+h) - 2f(x) + f(x-h) = \int_x^{x+h} [f'(t) - f'(t-h)]dt .$$

In the opposite direction, convolve f with a smooth peaking kernel, and proceed as in the above remark.)

Theorems 6, 7 and 8 can obviously be generalized to the case where K has an n-th moment $(n \geq 1)$ (we have formulated the case $n = 2$ explicitly because of the many kernels of practical interest which are subsumed in this case). Thus we have

THEOREM 9. *Let K be integrable, $\int K dx = 1$, and further*

$$x^n K \in L^1(-\infty, \infty)$$

$$\int x^i K(x)dx = 0, \qquad i = 1, 2, \ldots, n-1$$

$$\int x^n K(x)dx = A \neq 0 .$$

If f is a bounded measurable function and if $f^{(n)}(x)$ exists, then

$$\lim_{\lambda \to \infty} \lambda^n [f^\lambda(x) - f(x)] = \frac{Af^{(n)}(x)}{n!} .$$

THEOREM 10. *Let K be as in Theorem 9, and let $-\infty \leq a < b \leq \infty$.*

(i) *If* $\max_{a \leq x \leq b} |f(x) - f^\lambda(x)| = o(\lambda^{-n})$, *$f$ coincides on $[a, b]$*

with a polynomial of degree at most n.

(ii) *If* $\max_{a \le x \le b} |f(x) - f^{\lambda}(x)| = O(\lambda^{-n})$, *then* $f \in S_n$ *in* (a, b).

(Here S_n *denotes the set of functions whose* n*-th difference formed with span* h, *is* $O(h^n)$. *)*

The proofs are nearly identical with the preceding, and we leave them to the reader.

There is still one point to settle before the saturation theory of these kernels is complete: if $f \in S_n$, can we assert that $\|f - f^{\lambda}\| = O(\lambda^{-n})$? The answer is in the affirmative (we shall settle this point in Chapter IV; see Theorem 20). Anticipating this result, we may assert that *for kernels which satisfy the hypotheses of Theorem 9, the saturation class is precisely* S_n.

Exercise. Prove that S_n is identical with the class of functions $f \in C^{n-1}$ such that $f^{(n-1)} \in \mathrm{Lip}\ 1$.

Thus, the question of saturation is completely settled for *regular kernels*, i.e., those such that *for some integer* $n \ge 1$, *the n-th moment of* K *exists as a Lebesgue integral and is not zero.* Choosing the smallest such n, we may assert that K is saturated with order λ^{-n} and S_n is the saturation class.

§3.3 *Saturation theorems, "regular" kernels.*

For kernels which are not regular, for instance if xK is not integrable (or, is xK is integrable and has integral equal to zero, but x^2K is not integrable) we have still no result. In particular, determination of the saturation order and class for the Cauchy and Fejér-de la Vallée Poussin kernels is unsettled. For the Cauchy kernel we prove first the following:

THEOREM 11. *Let* $K(x) = 1/\pi \cdot 1/(1+x^2)$. *Suppose* f *is bounded and measurable on* $(-\infty, \infty)$ *and*

$$J(f;\, x) = -\frac{1}{\pi} \int_0^\infty \frac{f(x+t) - 2f(x) + f(x-t)}{t^2}\, dt$$

exists as a Lebesgue integral for some particular value of x. *Then*

$$\lim_{\lambda \to \infty} \lambda[f(x) - f(x;\, \lambda)] = J(f;\, x) \; . \tag{4}$$

Proof:

$$f(x;\, \lambda) = \int_{-\infty}^{\infty} f(x-t)\, K_\lambda(t)\, dt$$

$$f(x;\, \lambda) - f(x) = \int_0^\infty [f(x+t) - 2f(x) + f(x-t)]\, K_\lambda(t)\, dt$$

$$\lambda[f(x;\, \lambda) - f(x)] = \frac{1}{\pi} \int_0^\infty \left[\frac{f(x+t) - 2f(x) + f(x-t)}{t^2} \cdot \frac{\lambda^2 t^2}{1+\lambda^2 t^2} \right] dt$$

and since the second factor in the integrand is bounded by one, and tends to one as $\lambda \to \infty$ for each positive t, the result now follows by the Lebesgue dominated convergence theorem.

Similarly we have

THEOREM 12. *Theorem 11 holds with* K *replaced by* $\dfrac{\sin^2 x}{\pi x^2}$, *except that the right side of (4) must be replaced by* $\frac{1}{2}J(f;\, x)$.

Proof: The proof proceeds as before, but now in (5) the second factor is replaced by $\sin^2 \lambda t$, and instead of dominated convergence we use the fact that

$$\lim_{\lambda \to \infty} \int_0^\infty F(t) \sin^2 \lambda t \, dt = \frac{1}{2} \int_0^\infty F(t) dt$$

holds for any $F \in L^1$. (This is a consequence of the identity $\sin^2 \lambda t = ½(1 - \cos 2\lambda t)$, together with the Riemann-Lebesgue lemma.)

From the two previous theorems one can deduce

THEOREM 13. *Let K denote either the Cauchy or the Fejér-de la Vallée Poussin kernel, and let f be bounded and continuous on $(-\infty, \infty)$.*

(i) *If $\|f - f^\lambda\| = o(1/\lambda)$, then f is constant;*

(ii) *If $\|f - f^\lambda\| = O(1/\lambda)$, then f is the uniform limit of a sequence $\{f_n\}$ of functions of class C^2 such that $J(f_n; x)$ remains uniformly bounded.*

Proof: By Theorems 11 and 12, and the "mollifier technique", the proof is reduced to demonstrating

THEOREM 14. *Let f be bounded and of class C^2 on $(-\infty, \infty)$, and suppose f'' bounded. Then, if*

$$J(f; x) = -\frac{1}{\pi} \int_0^\infty \frac{f(x+t) - 2f(x) + f(x-t)}{t^2}$$

vanishes identically, f is constant.

Proof: The following proof, kindly communicated to us by D. J. Newman, requires some knowledge of harmonic functions. However, no really elementary proof of Theorem 14 is known to us (Another proof, based on Fourier transforms, shall be given in Chapter 5.)

We consider the Poisson integral

$$u(x, y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y}{(x-t)^2 + y^2} f(t)dt .$$

u is harmonic in the half-plane $y > 0$. Now, the partial derivative $u_y = \partial u/\partial y$ is also harmonic for $y > 0$, and

$$u_y(x, y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(x-t)^2 - y^2}{((x-t)^2 + y^2)^2} \cdot f(t)dt$$

$$= \frac{1}{\pi} \int_0^{\infty} [f(x+t) + f(x-t)] M(t, y)dt$$

where

$$M(t, y) = \frac{t^2 - y^2}{(t^2 + y^2)^2}$$

Since

$$\int_0^{\infty} M(t, y)dt = 0$$

we have

$$(6) \quad u_y(x, y) = \frac{1}{\pi} [f(x+t) - 2f(x) + f(x-t)] M(t, y)dt$$

$$= \frac{1}{\pi} \int_0^{\infty} \frac{f(x+t) - 2f(x) + f(x-t)}{t^2} \cdot \frac{t^2(t^2 - y^2)}{(t^2 + y^2)^2} dt .$$

Let $C = \sup |f''(x)|$, then from (6)

$$|u_y(x, y)| \leq \frac{1}{\pi} \int_0^{\infty} \frac{|f(x+t) - 2f(x) + f(x-t)|}{t^2} dt$$

$$\leq \frac{C}{\pi} + \frac{4\|f\|}{\pi} \cdot \int_1^{\infty} \frac{dt}{t^2}$$

showing that u_y is bounded in the half-plane $y > 0$. Now, from (6), by dominated convergence

$$\lim_{y \to 0+} u_y(x, y) = \frac{1}{\pi} \int_0^\infty \frac{f(x+t) - 2f(x) + f(x-t)}{t^2} \, dt$$

$$= 0$$

by hypothesis, for each x. Therefore u_y vanishes identically, hence u is constant,[3] and so finally

$$f(x) = \lim_{y \to 0+} u(x, y)$$

is constant, and the Theorem is proved.

To complete the discussion of the saturation behavior of these kernels (i.e., to identify precisely the class of functions for which approximation to within an error of $O(\lambda^{-1})$ is possible) involves some harmonic analysis, and we shall return to this in Chapter 5.

§3.4 *Application of harmonic analysis.*

Thus far we have not made use of Fourier methods in our analysis, except heuristically, It is, however, advantageous to do so (even indispensable for some of the deeper inverse theorems, see Chapter 5). Let us give here a very simple application of Fourier methods to a saturation problem involving *periodic* functions.

THEOREM 15. *Let f be continuous on $(-\infty, \infty)$ and of period 2π, with the Fourier series*

$$f \sim \sum_{-\infty}^{\infty} c_n e^{int} .$$

Let $K \in L^1$, then

$$(7) \qquad |c_n| \cdot |1 - \hat{K}(\tfrac{n}{\lambda})| \le \|f - (f * K_\lambda)\| .$$

Proof: Denote the right side of (7) by M. Then

$$\left| \int_{-\infty}^{\infty} f(x-t)\, K_\lambda(t) dt - f(x) \right| \leq M .$$

Hence

$$\left| \frac{1}{2\pi} \int_0^{2\pi} e^{-inx} \left(\int_{-\infty}^{\infty} f(x-t)\, K_\lambda(t) dt \right) dx - c_n \right| \leq M$$

and the iterated integral equals

$$\int_{-\infty}^{\infty} K_\lambda(t) \left(\frac{1}{2\pi} \int_0^{2\pi} f(x-t)\, e^{-inx} dx \right) dt = c_n \hat{K}(\tfrac{n}{\lambda})$$

completing the proof.

Examples. If K is the Fejér-de la Vallée Poussin kernel, $1-\hat{K}(n/\lambda) = |n|/\lambda$ for $|\lambda| \geq n$. If therefore $\lambda \| f-(f*K_\lambda)\| \to 0$ as $\lambda \to \infty$, we deduce from (7) that nc_n vanishes for each n, hence f is constant. (It is easily verified that for integral λ, $f*K_\lambda$ is a classical Fejér sum formed from the Fourier series of f.) A similar argument applies to the Cauchy kernel (here $\hat{K}(x) = e^{-|x|}$). Moreover, for general K, if $|1-\hat{K}(x)| \geq A|x|^p$ for $|x| \leq a$, where $A > 0$, then one shows similarly that no non-constant (2π-periodic) function can have degree of approximation of smaller order than λ^{-p}.

The Fourier method does not, however, directly give the saturation *class*, for instance in the Fejér case just discussed a degree of approximation $O(1/\lambda)$ implies, using (7), that the Fourier coefficients of f are $O(1/n)$. This is an imperfect result however, since the latter condition does not imply the former.

Further results, giving more perspective on saturation, will be proved in Chapter V. There does not seem to be any royal road, however, to the determination of saturation classes for arbitrary kernels (assuming that this problem could even be posed in a precise way).

§3.5 *Digression: further applications of mollifiers.*

To illustrate further the use of the mollifier technique, we shall use it to prove two theorems of Hardy and Titchmarsh, which are very close in spirit to those of this chapter.

THEOREM 16. (Titchmarsh). *Let* $f \in L^1(-\infty, \infty)$ *and*

$$\int_{-\infty}^{\infty} |f(x+t) - f(x)|\,dx = o(t), \qquad t \to 0+ \ .$$

Then f *is (almost everywhere) equal to zero.*

Proof: 1) Suppose first $f \in C^2$ and f'' is bounded. Let $M = \sup |f''(x)|$. By Taylor's formula

$$|f(x+t) - f(x) - tf'(x)| \le \frac{Mt^2}{2} \ .$$

Therefore $f(x+t) - f(x) \ge tf'(x) - Mt^2/2$. Now, for any numbers a, b with $a < b$, by hypothesis

$$\int_a^b |f(x+t) - f(x)|\,dx \le \phi(t)$$

where $(\phi(t))/t \to 0$ as $t \to 0+$. Hence

$$\int_a^b \left(tf'(x) - \frac{Mt^2}{2}\right)dx \le \phi(t)$$

$$t(f(b) - f(a)) \le \phi(t) + \frac{Mt^2}{2}(b-a) \ .$$

Dividing by t, and letting $t \to 0$ we conclude $f(b) \le f(a)$. Since the same argument applies to $-f$, we conclude $f(a) \leq f(b)$. Hence $f(a) = f(b)$, and since a and b are arbitrary, f is constant.

2) In the general case, let K denote (say) the Cauchy kernel, and let $f^\lambda = f * K_\lambda$. Then f^λ has bounded derivatives of all orders. Moreover,

$$\begin{aligned}\int |f^\lambda(x+t) - f^\lambda(x)|dx &= \int \left| \int (f(x+t-u) - f(x-u))\, K_\lambda(u)du \right| dx \\ &\leq \int K_\lambda(u) \left[\int |f(x+t-u) - f(x-u)|dx \right] du \\ &= \int |f(x+t) - f(x)|dx\end{aligned}$$

and this is $o(t)$ by hypothesis, hence f^λ is constant. By Theorem 2, f^λ converges to f in L^1 norm, which implies that f is a constant, apart from a set of measure zero. Since $f \epsilon L^1$ this constant is zero.

THEOREM 17. (Hardy and Littlewood) *Let* $f \epsilon L^1(-\infty, \infty)$ *and*

$$\int_{-\infty}^{\infty} |f(x+t) - f(x)|dx = O(t), \qquad t \to 0+ \; . \tag{8}$$

Then f *coincides almost everywhere with a function of finite total variation on* $(-\infty, \infty)$.

Proof: 1) Suppose $f \epsilon C^2$ and $|f''(x)| \leq M$, and $\int |f(x+t) - f(x)|dx \leq At$ for $t > 0$. Let E denote any finite union of disjoint intervals. Then (reasoning just as in the preceding proof)

$$\int_E (tf'(x) - \frac{Mt^2}{2})dx \leq At$$

which implies, letting $t \to 0$,

$$\int_E f'(x)dx \leq A$$

and so (since we can write $-f$ for f)

$$\left| \int_E f'(x)dx \right| \le A .$$

Since E is arbitrary, f has total variation not exceeding $2A$.

2) If now f is integrable and satisfies the hypotheses, we get for $f^\lambda = f * K_\lambda$,

$$\int |f^\lambda(x+t) - f^\lambda(x)|dx \le \int |f(x+t) - f(x)|dx \le At$$

and so $\{f^\lambda\}$ have uniformly bounded variation. Therefore f is the limit almost everywhere of a sequence of functions of uniformly bounded variation, which implies the desired result (the reader may easily supply the missing details).

Exercises. a) Formulate and prove the appropriate corresponding theorems for L^p norm ($p > 1$).

b) The same, replacing $f(x+t) - f(x)$ by $f(x+t) - 2f(x) + f(x-t)$.

The converse of Theorem 17 is also true: *if f is integrable and coincides a.e. with a function of finite total variation, then* (8) holds.

To see this, consider first an integrable function g which is absolutely continuous and has an integrable derivative. Then,

$$\int_{-\infty}^{\infty} |g(x+t) - g(x)|dx = \int_{-\infty}^{\infty} \phi_t(x)(g(x+t) - g(x))dx$$

where ϕ_t is measurable and bounded by one. This equals

$$\int_{-\infty}^{\infty} \phi_t(x)\left(\int_0^t g'(x+y)dy\right)dx = \int_0^t \left(\int_{-\infty}^{\infty} \phi_t(x)g'(x+y)dx\right) dy$$
$$\leq t \int_{-\infty}^{\infty} |g'(x)|\,dx \quad .$$

To pass to the general case, we first make a preliminary smoothing of f and apply the preceding inequality, then pass to a limit. The details should by now be familiar, and may be left to the reader (a very short direct proof is possible too; we give this proof only for its methodological value).

REMARK. We have proved, in passing

THEOREM 18. *The following two classes of functions are identical:*

(i) *The set of* f *which are integrable and coincide a.e. with a function of finite variation.*

(ii) *The set of* f *for which there exist a sequence* f_n *of integrable* C^∞ *functions such that* $\|f - f_n\|_1 \to 0$ *and* $\sup \int |f_n'| < \infty$. *Moreover, the total variation of the B. V. function equal a.e. to* f *is the* inf *of the numbers* $\sup \int |f_n'|$ *over all such sequences* $\{f_n\}$.

A similar identification can be made between bounded functions of bounded variation, and *pointwise* limits of C^∞ functions f_n such that $\sup \int |f_n'|$ is finite. This theme has many variations.

Results of this general character† are very important in analysis— we have already seen evidence of this in the proof of the converse of Theorem 17, and in the identification of the class S above with the pointwise limits of sequences $\{f_n\}$ having uniformly bounded second derivatives. By analogy with a language employed in the

† often called *regularisation theorems*

theory of partial differential equations, where one speaks of *weak solutions* of differential equations, we may say (in a language that can be made precise in various equivalent ways) that functions of total variation $\le M$ are, apart from adjustment on a set of measure zero, weak solutions of the *inequality* $\int|f'| \le M$. Likewise, Lip 1 (with Lipschitz constant M) is the set of weak solutions of the inequality $|f'(x)| \le M$, S above is the set of weak solutions of $\sup|f''(x)| < \infty$, non-decreasing functions are weak solutions of $f'(x) \ge 0$, convex functions of $f''(x) \ge 0$, subharmonic functions of $u_{xx} + u_{yy} \ge 0$, etc., etc. The "mollifier technique" often enables us to prove theorems about these "weak solutions" by operating with the "strong" (i.e., genuine) solutions, and passing to a limit, i.e., the "weak solutions" are the limits, in some ("weak") topology, of the strong solutions.

§3.5.1 Another method of regularization.

We have proved many of the theorems of this chapter by using "mollifiers," i.e., smoothing a given function by convolving it with a smooth kernel. There is another variant of this technique for accomplishing the same purpose, based upon the formula for integration by parts. Central here is that one fixes in advance a class of "test functions"; for our purposes here we may define a "test function" to be a C^∞ function of compact support.

Let us illustrate the technique by giving an alternate proof of Theorem 16. We have, by hypothesis

$$\int |f(x+t) - f(x)|dx \le t\varepsilon(t), \quad \lim_{t\to 0+} \varepsilon(t) = 0 .$$

Therefore, for any test function u,

$$\left|\int [f(x+t) - f(x)]\, u(x)dx\right| \leq t\,\varepsilon(t) \max |u| \;.$$

Now, the integral on the left equals

$$\int f(y)\,[u(y-t) - u(y)]dy \;,$$

and so

$$\left|\int f(y)\left[\frac{u(y) - u(y-t)}{t}\right] dy\right| \leq \varepsilon(t) \max |u| \;.$$

Now let $t \to 0$; applying the dominated convergence theorem we see that the integral on the left tends to $\int f(y)u'(y)dy$, and so

$$\int f(y)u'(y)dy = 0$$

for all test functions. From this, it is a simple exercise to show f is constant a.e. Replacing $u(y)$ by the test function $u(x-y)$ gives for every x,

$$0 = \int f(y)u'(x-y)dy = \frac{d}{dx} \int f(y)\, u(x-y)dy \;.$$

Therefore $f * u$ is a constant, for every test function. Choosing, in particular, $u = K_\lambda$ where K is a test function with $\int K = 1$ gives, since $\|f - f * K_\lambda\| \to 0$ (by Theorem 2) that f coincides a.e. with a constant.

To get Theorem 17, we proceed in the same way and obtain the conclusion

$$(9) \qquad \left|\int f(y)u'(y)dy\right| \leq C \max |u|$$

for all test functions u, where C is a constant. Now if one can show that f coincides a.e. with a function of total variation $\leq C$, the proof is complete. This is not too hard, and we leave it as an exercise.

Exercises

1. Deduce from (9) that f coincides a.e. with a function of total variation $\leq C$.

2. Suppose f is continuous, and for all test functions u,

$$\left| \int f(x)u'(x)dx \right| \leq C \int |u(x)|dx \quad .$$

Show $f \in$ Lip 1, with Lipschitz constant C, and conversely.

3. Suppose f is continuous, and for all test functions u,

$$\left| \int f(x)\, u''(x)dx \right| \leq C \int |u(x)|dx \quad .$$

Show f is of class S (see §3.2), and conversely. Generalize this result to higher derivatives.

4. Suppose f is continuous, and $\int f(x)\, u''(x)dx \geq 0$ for all non-negative test functions u. Prove f is convex, and conversely.

5. Prove the following version of "Weyl's lemma": Let f be a continuous function on R^2 such that $\int f \Delta u = 0$ for every C^∞ function u with compact support in R^2 (Δ denotes the Laplacian). Then f is a harmonic function. (Hint: The hypotheses imply $f * u$ is harmonic for all test functions. Choosing $u(x, y) = \lambda^2 K(\lambda x, \lambda y)$ for suitable K shows that f is, on each bounded set, a uniform limit of harmonic functions).

Generalize to f which are merely assumed integrable on bounded rectangles, and to higher dimensional space. (Can you extend the result also to other differential operators?)

6. The following two conditions on a function f are equivalent if $p > 1$:

(i) $\int |f(x+h) - f(x)|^p dx \le h^p$ for $h > 0$;

(ii) f coincides almost everywhere with an absolutely continuous function g such that $\int |g'(x)|^p dx \le 1$.

(Hint: If g is absolutely continuous, then

$$\begin{aligned}\int_{-\infty}^{\infty} |g(x+h) - g(x)|^p dx &= \int_{-\infty}^{\infty} \left| \int_0^h g'(x+t)dt \right|^p dx \\ &\le h^{p-1} \int_{-\infty}^{\infty} \left(\int_0^h |g'(x+t)|^p dt \right) dx \\ &= h \int_{-\infty}^{\infty} |g'(x)|^p dx \ .\end{aligned}$$

In the other direction, consider first an f satisfying (i) which is absolutely continuous; an application of Fatou's lemma gives (*) $\int |f'(x)|^p dx \le 1$. For the general case, observe that (i) continues to hold with f replaced by $f^\lambda = f * K_\lambda$ where K is (say) the Cauchy kernel, therefore the f^λ satisfy (*). This implies, by a simple application of Hölder's inequality, that if E is any finite union of disjoint intervals (x_i, y_i), $\Sigma |f^\lambda(y_i) - f^\lambda(x_i)| \le (mE)^{(p-1)/p}$ (m: Lebesgue measure), and letting $\lambda \to \infty$ gives the absolute continuity of f, after adjustment on a set of measure zero.)

7. If f is absolutely continuous and $f' \in L^p$ $(p > 1)$, then the modulus of continuity of f is $o(h^r)$, where $r = p/(p-1)$.

8. Suppose $f \in L^p$ $(p > 1)$, and $\int |f(x+h) - f(x)|^p dx = O(h^{ps})$ where $1/p < s \le 1$. Then, f coincides almost everywhere with a continuous function, whose modulus of continuity is $o(h^r)$, $r = s - (1/p)$.

9. A function satisfying the hypotheses of the preceding exercise is said to be of class Lip(s, p). Prove that if $f \in$ Lip(s, p) and $ps \le 1$, then also $f \in (s-(1/p) + (1/q); q)$ for all q such that $p < q < p/(1-sp)$.

Chapter III: Footnotes added in proof

Page 28. [1] The corresponding class of functions which belong to S *in an interval* (a, b) is defined in the obvious way.

Page 29 [2] Here, and only here, we depart from our usual superscript notation to avoid ambiguity.

Page 36 [3] For, u is a function of x alone, hence it is a linear function of x, and it is bounded (being the Poisson integral of f).

CHAPTER IV

DIRECT THEOREMS, DEGREE OF APPROXIMATION

§4.1 *Introduction.*

By way of orientation, let us examine some well-known facts concerning trigonometric approximation. Let $\mathcal{T}_n$ denote the set of trigonometric polynomials of degree not exceeding n, i.e.,

$$\mathcal{T}_n = \left\{ T \mid T(x) = \sum_{k=-n}^{n} c_k e^{ikx} \right\}.$$

Let $C_{2\pi}$ denote the class of continuous functions of period 2π. If $f \in C_{2\pi}$ and $f \in \text{Lip}\, \alpha$ $(0 < \alpha \leq 1)$ then, as is well known, f can be approximated by an element of $\mathcal{T}_n$ with an error $O(n^{-\alpha})$. The usual proof is to convolve f with a suitable peaking kernel K (here 'convolution' is meant relative to the circle group:

$$(f * g)(x) = \frac{1}{2\pi} \int_0^{2\pi} f(x-t)\, g(t)\, dt \quad),$$

where K is a trigonometric polynomial. The choice

$$J_n(x) = A_n \left(\frac{\sin nx}{\sin x} \right)^4$$

(Jackson kernel) for K, which is easily seen to be a trigonometric polynomial of degree not exceeding $4n-2$, is adequate to prove the stated result. The choice of the Fejér kernel $B_n((\sin nx)/(\sin x))^2$

is not sufficient–it works for each $a < 1$, but fails at $a = 1$. This reflects the fact that the Fejér kernel is saturated with order $1/n$, and the saturation class does *not* contain all Lip 1 functions. The Jackson kernel, on the other hand, is saturated with order $1/n^2$.

It is important to understand what structural property of the Jackson kernel makes it 'less saturated' than the Fejér kernel, and more generally to be in possession of a method for constructing kernels which generate approximations with desired characteristics. The situation for periodic functions becomes most transparent when we look at the problem on the infinite line (then the considerations of the previous chapter can be applied). It is a remarkable fact, apparently first noticed by de la Vallée Poussin, that the study of approximation of functions on the circle group is very much simplified by considering them as functions on the infinite line, of period 2π, and constructing the desired approximations by convolving with L^1 kernels on $(-\infty, \infty)$. For example, the Fejér and Jackson kernels pass over to kernels of the type $\lambda K(\lambda x)$ determined by mere scale change from one single generating function K; this is not true for the corresponding kernels for the circle group.

From this point of view, in order to generate *trigonometric polynomial* approximations to functions of period 2π, we are led to look for L^1 kernels K such that $f * K \in \mathcal{T}_n$ whenever $f \in C_{2\pi}$. We have

THEOREM 19. *Let* $f \in C_{2\pi}$ *(i.e., continuous and of period* 2π*), and let* $K \in L^1$, $\hat{K}(x) = 0$ *for* $|x| \geq 1$. *Then* $f * K_n \in \mathcal{T}_{n-1}$.

Note. From here on, Fourier transforms will occur quite often in the analysis. We understand by $\hat{K}$, the ("direct") *Fourier transform* of K, the function

$$\hat{K}(x) = \int_{-\infty}^{\infty} K(t)e^{-itx}dt .$$

We will sometimes speak of K as the '*inverse transform*' of $\hat{K}$. When $\hat{K} \in L^1$, we have a.e. the Fourier inversion

$$K(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{K}(x)e^{itx}dx .$$

We shall have to presuppose some knowledge of Fourier transforms (the more so in the following chapter). Especially we require sufficient conditions on a function f in order that it be the Fourier transform of an L^1 function. One useful criterion is f integrable and absolutely continuous with a square integrable derivative. Another is f absolutely continuous with f and f' square integrable. (We leave the proofs as an exercise.)

We give two proofs of Theorem 18, and outline a third.

First Proof. Let us first consider $f(t) = e^{imt}$, where m is an integer. Then

$$f * K_\lambda = \hat{K}\left(\frac{m}{\lambda}\right)e^{imt} .$$

If λ is a positive integer n, this belongs to $\mathcal{T}_{n-1}$ since $\hat{K}(m/n) = 0$ for $|m| \geq n$. Therefore, the conclusion of the theorem is correct when f is a single exponential, and so by linearity when f is a trigonometric polynomial.

To complete the proof we use the Weierstrass (trigonometric) polynomial approximation theorem (this slight aesthetic defect to the development given here will be more than compensated by the ease with which the deeper degree-of-approximation theorems are obtained). For $f \in C_{2\pi}$, we can find a sequence of trigonometric polynomials $\{T_m\}$ with $\|f - T_m\| \to 0$. Hence

$$\lim_{m \to \infty} \|f * K_n - T_m * K_n\| = 0 .$$

This shows that $f * K_n$ is the uniform limit of a sequence of elements of $\mathcal{T}_{n-1}$, hence itself belongs to $\mathcal{T}_{n-1}$ since this set is closed.

Second Proof: If $f \in C_{2\pi}$, the same is true of $g = f * K_n$. Therefore, to prove $g \in \mathcal{T}_{n-1}$ it is enough to prove the Fourier coefficients of g having rank exceeding $n-1$ vanish. Now

$$\begin{aligned}\int_0^{2\pi} g(x)e^{-imx}dx &= \int_0^{2\pi} e^{-imx}\left(\int_{-\infty}^{\infty} f(x - \tfrac{t}{n})\, K(t)dt\right)dx \\ &= \int_{-\infty}^{\infty} K(t)\left(\int_0^{2\pi} f(x - \tfrac{t}{n})e^{-imx}dx\right)dt \\ &= \hat{K}(\tfrac{m}{n}) \int_0^{2\pi} f(u)e^{-imu}du \\ &= 0 \text{ for } |m| \geq n\end{aligned}$$

completing the proof.

Third Proof: (Outline) K is (the restriction to the real axis of) an entire function of exponential type $\leq n$. $g = f * K_n$ "inherits" this property from K_n, as well as 2π-periodicity from f. But, a beast with *both* of these properties must be a trigonometric polynomial of degree $\leq n$. (The pleasure of proving this, via Cauchy's theorem, we leave as an exercise, as well as showing that g really has degree *less* than n.)

REMARKS. If we want $f * K_n \in \mathcal{T}_{n-1}$ for *all* n (or, for arbitrarily large n) then we must, conversely, require that $\hat{K}(x)$ vanish for $|x| \geq 1$. On the other hand, if we confine ourselves to a *fixed* value of n, the weaker requirement that $\hat{K}(m/n) = 0$ for $|m| \geq n$ suffices. We can also express the conclusion in this form: if $J \in L^1$, the necessary and sufficient condition that $f \in C_{2\pi}$ implies $f * J \in \mathcal{T}_{n-1}$ is that $\hat{J}(x) = 0$ for all *integral* x such that $|x| \geq n$. We remark also that the conclusion of Theorem 19 holds (as the second proof shows) under weaker hypotheses, for example if f (of period 2π) is only bounded and measurable; or if $f \in L^1(0, 2\pi)$ and $\Sigma_{r=-\infty}^{\infty} M_r < \infty$, where $M_r = \operatorname*{ess\,sup}_{2\pi r \leq t \leq 2\pi(r+1)} |K(t)|$. Thus, a great deal of classical Fourier theory (for example, almost everywhere summability theorems for integrable functions of period 2π) can be deduced from corresponding, more general, theorems which are valid for non-periodic functions as well.

The result of the computation in the second proof is also worth emphasizing: if f is of period 2π, then $f * K_\lambda$ is the function whose Fourier coefficients are obtained from the corresponding coefficients c_m of f upon multiplication by the factor $\hat{K}(m/\lambda)$. In particular, we see that when K is the Fejér-de la Vallée Pousin kernel $f * K_n$ is the classical Fejér sum of order $n-1$.

In the remainder of this chapter, we shall illustrate some techniques for proving "direct" theorems. Another, more powerful technique, will be discussed in the next chapter.

§4.2 *Fundamental direct theorem—first variant.*

In keeping with the approach outlined in §4.1, we wish to prove a 'direct theorem' for functions on the line, which when specialized to functions of period 2π gives us the fundamental degree-of-approximation theorem, that of Jackson. We require first

DEFINITION. *The modulus of continuity* $\omega(f; t) = \omega(t)$ of a continuous function f is defined by

$$\omega(t) = \sup_{|x_1 - x_2| \leq t} |f(x_1) - f(x_2)| .$$

The following facts are easily verified:

(i) $\lim_{t \to 0} \omega(t) = 0$ if and only if f is uniformly continuous.

(ii) $\omega(t_1 + t_2) \leq \omega(t_1) + \omega(t_2)$.

(iii) $\omega(nt) \leq n\omega(t)$ (n positive integer).

(iv) $\omega(at) \leq (a+1)\omega(t)$ (a positive real).

(v) $\omega(t) = O(t^{\alpha})$ if and only if $f \in \mathrm{Lip}\,\alpha$.

THEOREM 20. *Hypotheses: f continuous and bounded on $(-\infty, \infty)$, and for some $r \geq 0$, $f \in C^r$ and $f^{(r)}$ has modulus of continuity ω,*

$$t^m K \in L^1(-\infty, \infty), \quad m = 0, 1, \ldots, r+1$$

and

$$\int_{-\infty}^{\infty} t^m K(t)\,dt = \begin{cases} 1, & m = 0 \\ 0, & m = 1, \ldots, r . \end{cases}$$

Conclusion: $\|f * K_\lambda - f\| \leq A\lambda^{-r}\omega(\frac{1}{\lambda})$, *where*

$$A = \frac{1}{r!} \int_{-\infty}^{\infty} (|t|^r + |t|^{r+1})\,|K(t)|\,dt .$$

COROLLARY. (Jackson's theorem). *If f has period 2π and satisfies the hypothese of Theorem 20, there exists $T \in \mathcal{T}_{n-1}$ such that*

$$\|f - T\| \leq A_r n^{-r} \omega(\frac{1}{n})$$

where A_r is a constant depending only on r.

DEDUCTION OF COROLLARY FROM THEOREM 20. In view of what has been said in 4.1, it is enough to produce a 'low frequency function' K (that is, one with $\hat{K}(x) = 0$ for $|x| \geq 1$) satisfying the integrability and moment conditions in Theorem 20. To this end, let $H \epsilon C^{r+3}$ and satisfy

(i) $H(0) = 1, \quad H'(0) = \cdots = H^{(r)}(0) = 0$.

(ii) $H(x) = 0$ for $|x| \geq 1$.

Such H (even infinitely differentiable ones) obviously exist. Then $K(t) = (1/2\pi) \int_{-\infty}^{\infty} H(x)e^{itx}dx$ is $O(1/|t|^{r+3})$ at infinity, and satisfies the required moment conditions.

REMARK. We shall not at this stage concern ourselves with a good estimate for A_r. In Jackson's original proof, an A_r was obtained which grows exponentially with r. Actually, A_r may be taken to be an absolute constant† independent of r, as was discovered by Favard, and independently by Ahieser and Krein. In fact, Favard and Ahieser-Krein found the *best possible* value for A_r in the case $\omega(t) = t$!

For the proof of Theorem 20 we will require a slightly non-standard estimate for the remainder in Taylor's formula.

LEMMA. *Under the hypotheses of Theorem 20, for all* x, t

$$\left| f(x+t) - \sum_{k=0}^{r} \frac{f^{(k)}(x)t^k}{k!} \right| \leq \frac{t^r \omega(|t|)}{r!}$$

Proof: Suppose first $t > 0$; we have the elementary identity

$$f(x+t) - \sum_{k=0}^{r} \frac{f^{(k)}(x)t^k}{k!} = \frac{1}{(r-1)!} \int_x^{x+t} [f^{(r)}(y) - f^{(r)}(x)](x+t-y)^{r-1} dy$$

† See §4.4.2 below.

and the right side is bounded by

$$\frac{\omega(t)}{(r-1)!} \int_x^{x+t} (x+t-y)^{r-1} dy = \frac{t^r \omega(t)}{r!} .$$

The case $t \leq 0$ is treated similarly.

Proof of Theorem 20.

$$f(x;\lambda) - f(x) = \int [f(x - \tfrac{t}{\lambda}) - f(x)] K(t) dt ,$$

and

$$f(x - \tfrac{t}{\lambda}) - f(x) = \sum_{k=1}^{r} \frac{f^{(k)}(x)}{k!} \left(-\frac{t}{\lambda}\right)^k + R(t) ,$$

where, by the lemma,

$$|R(t)| \leq \frac{1}{r!} \cdot \left(\frac{|t|}{\lambda}\right)^r \cdot \omega\left(\frac{|t|}{\lambda}\right) .$$

Hence

$$|f(x;\lambda) - f(x)| = \left| \int R(t) K(t) dt \right| \leq \frac{1}{r!\,\lambda^r} \int |t|^r \omega\left(\frac{|t|}{\lambda}\right) |K(t)| dt$$

and since $\omega(|t|/\lambda) \leq (|t| + 1)\omega(1/\lambda)$ the theorem is proved.

REMARK. Application of Theorem 20, with $r = n - 1$ and $\omega(t) = t$ shows (in the notation of §3.2) that for the kernels of Theorem 9, $f \in S_n$ is approximable to the order λ^{-n}. This remark completes the saturation theory of regular kernels

§4.3 *Fundamental direct theorem—Second variant.*

One way to express quantitatively that a function has a certain degree of smoothness is to specify that it possesses a *derivative* of some order, which in turn may have a certain modulus of continuity.

Another way, in some respects more convenient, to express smoothness is by means of *differences*, rather than derivatives. We have already remarked that the first difference $f(x+t)-f(x)$ cannot be uniformly $o(t)$ unless f reduces to a constant. The second difference $f(x+2t) - 2f(x+t) + f(x)$ can, however, non-trivially be $O(t^2)$ (but not $o(t^2)$), and so on for higher order differences. If we introduce, therefore, the operator

$$\Delta_t : \ (\Delta_t f)(x) = f(x+t) - f(x)$$

and its powers

$$\Delta_t^r : \ (\Delta_t^r f)(x) = \sum_{m=0}^{r} \binom{r}{m}(-1)^r f(x+mt)$$

we can define

$$\phi_r(t) = \phi_r(f;\, t) = \sup_x |(\Delta_t^r f)(x)|$$

$$\omega_r(t) = \omega_r(f;\, t) = \sup_{|u|\le t} \phi_r(f;\, u)$$

ω_r is called the r^{th} *order modulus of smoothness of* f. For $r = 1$ we get the modulus of continuity defined earlier (and shall write ω for ω_1).

The following properties of ω_r can be verified without much difficulty (for details we must refer the reader to the textbooks of Timan or Lorentz).

(i) $\omega_r(t) \le 2\omega_{r-1}(t)$

(ii) If f is r times differentiable, then

$$\omega_r(t) \leq t^r \sup_x |f^{(r)}(x)| .$$

(iii) $\omega_{k+r}(f; t) \leq t^r \omega_k(f^{(r)}; t)$ for $f \in C^r$.

The case $k = 1$ is of special importance:

(iv) $\omega_{r+1}(f; t) \leq t^r \omega(f^{(r)}; t)$.

(v) $\omega_r(nt) \leq n^r \omega_r(t)$, n positive integer

$\omega_r(tu) \leq (n+1)^r \omega_r(t)$, u positive real number.

There are also inequalities enabling one to estimate ω_r in terms of ω_s when $r < s$, but these lie deeper and shall be discussed in Chapter V.

Actually, it will be convenient notationally to define a somewhat more general modulus of smoothness. Let m be any finite measure on $(1, \infty)$ of total mass one. For $f \in C$, we define

$$\begin{aligned} \psi_m(u) &= \psi_m(f; u) \\ &= \sup_t |f(t) - \int f(t - uy)dm(y)| \end{aligned}$$

$$\omega_m(t) = \sup_{|u| \leq t} \psi_m(u) .$$

Examples. If m is a single point mass at $y = +1$, of mass one, we get

$$f(t) - \int f(t-uy)dm(y) = f(t) - f(t-u)$$

and ω_m is the ordinary modulus of continuity. If m consists of point masses $+2$ at $y = 1$ and -1 at $y = 2$, ω_m is the modulus of smoothness ω_2. Similarly, the higher order moduli of smoothness are special cases of the above.

LEMMA. *Let f, m, ψ_m be as above, and* p $\in L^1(-\infty, \infty)$, $\int p\, dt = 1$. *Define*

(1) $$K(t) = \int p(\tfrac{t}{y})\, \frac{dm(y)}{y}$$

Then, $K \in L^1$, *and* $\|f-(f*K)\| \le \int \psi_m(u)|\,p(u)|\,du$.

Proof: That $K \in L^1$ follows from

$$\int_{-\infty}^{\infty} |K(t)|\,dt \le \int_{-\infty}^{\infty} \left(\int |p(\tfrac{t}{y})| \frac{|dm(y)|}{y} \right) dt = \|p\|_1 \int |dm| \ .$$

Write $g = f*K$. Now,

$$g(t) = \int f(t-v)\,K(v)\,dv = \int f(t-v) \left(\int p(\tfrac{v}{y})\, \frac{dm(y)}{y} \right) dv$$

$$= \int dm(y) \int f(t-v)\,p(\tfrac{v}{y})\,\frac{dv}{y}$$

$$= \int dm(y) \int f(t-uy)\,p(u)\,du = \int p(u) \left(\int f(t-yu)\,dm(y) \right) du$$

$$f(t) - g(t) = \int p(u) \left[f(t) - \int f(t-yu)\,dm(y) \right] du$$

$$|f(t)-g(t)| \le \int \psi_m(u)\,|p(u)|\,du$$

proving the lemma.

THEOREM 21. *If* $f \in C_{2\pi}$, *then for every positive integer* r *one can find* $T \in \mathcal{T}_{n-1}$ *such that*

$$\|f-T\| \le B_r \omega_r(f;\ \tfrac{1}{n})$$

where B_r *is a constant depending only on* r.

Proof: Choose $p_0 \in L^1$, such that $t^r p_0 \in L^1$, and $\hat{p}_0(x) = 0$ for $|x| \ge 1$. Write $p(t) = np_0(nt)$, and define K by (1). Finally, let $T = f*K$. We claim this T fulfills the requirements.[1] First of all, note that

$$\hat{K}(x) = \int e^{-itx}\left(\int p\left(\frac{t}{y}\right) \frac{dm(y)}{y}\right) dt$$

$$= \int \hat{p}(xy)\, dm(y) .$$

Now, $\hat{p}(x) = 0$ for $|x| \geq n$, and since the support of m lies in $(1, \infty)$, also $\hat{K}(x) = 0$ for $|x| \geq n$, therefore $T \in \mathcal{T}_{n-1}$ by Theorem 19. Now, by suitable choice of the measure m, $\omega_m(t)$ becomes the modulus of smoothness $\omega_r(t)$, and finally, using the Lemma

$$\|f - T\| \leq \int_{-\infty}^{\infty} \omega_r(|u|)\, |p(u)|\, du$$

$$= \int_{-\infty}^{\infty} \omega_r\left(\frac{|t|}{n}\right) |p_0(t)|\, dt$$

$$\leq \omega_r\left(\frac{1}{n}\right) \int_{-\infty}^{\infty} (|t| + 1)^r |p_0(t)|\, dt$$

(using property (v) of moduli of smoothness) and the theorem is proved.

REMARK. Theorem 21 implies Jackson's theorem in the formulation of §4.2 because of the inequality

$$\omega_{r+1}\left(f; \frac{1}{n}\right) \leq n^{-r} \omega\left(f^{(r)}; \frac{1}{n}\right) .$$

§4.4 *Fundamental direct theorem—Third variant.*

§4.4.1 Use of $2n$ "integrating kernel".

In this section we confine ourselves to the trigonometric case, and present a method which enables us to get concrete estimates of the constant in Jackson's theorem. We shall restrict ourselves first to a special form of Jackson's theorem (from which the general form can easily be deduced, however).

THEOREM 22. *Let $f \in C^r$ have period 2π, and $|f^{(r)}(t)| \le 1$. Then $T \in \mathcal{T}_{n-1}$ can be found such that*

$$\|f - T\| \le A_r n^{-r}$$

where A_r depends only on r, and is $O(r)$.[2] *For even r, we may take $A_r = r + 1$.*

The proof is based on two lemmas. The first demonstrates a particular instance of the phenomenon mentioned already in the Introduction, that primitives are under certain conditions obtainable as convolutions.

LEMMA 1. *Let $f \in C^r$ have period 2π, and mean value zero. Let $J_r \in L^1(-\infty, \infty)$, and $\hat{J}_r(x) = (ix)^{-r}$ for $x = \pm 1, \pm 2, \ldots$. Then $f = f^{(r)} * J_r$.*

Proof: Let f have the Fourier series $\Sigma_{-\infty}^{\infty} c_n e^{int}$, then

$$f^{(r)} \sim \sum_{-\infty}^{\infty} (in)^r c_n e^{int}$$

hence

$$f^{(r)} * J_r \sim \sum (in)^r c_n \hat{J}_r(n)\, e^{int}$$

and the right side is just the Fourier series of f, since $c_0 = 0$ by hypothesis.

LEMMA 2. *Given $N > 0$, and a positive integer r, there exists*

$p = p_{r,N} \in L^1$ *such that*

$$\hat{p}(x) = x^{-r}, \qquad |x| \geq N,$$

$$\|p\|_1 \leq A_r N^{-r}$$

where A_r *depends only on* r, *and is* $O(r)$. *For even* r, *we may take* $A_r = r + 1$.

Proof: Let $g \in L^1$, $\hat{g}(x) = x^{-r}$ for $|x| \geq 1$ (such g can be constructed by smoothly completing the graph of x^{-r} between -1 and $+1$, and taking this as the graph of $\hat{g}$. For the time being, we will leave the choice of $\hat{g}(x)$ on $[-1, 1]$ open.)

Now, set $p(t) = g(Nt)/N^{r-1}$. We have

$$\|p\|_1 = N^{-r}\|g\|_1,$$

and

$$\hat{p}(x) = N^{-r}\hat{g}\left(\frac{x}{N}\right) = x^{-r}, \text{ for } |x| \geq N.$$

Now, in case r is *even*, we can complete the graph of $\hat{g}$ by drawing tangent lines to the curve $y = x^{-r}$ at $x = \pm 1$. With this choice,[3] $\hat{g}(x)$ is even, and *convex* on $(0, \infty)$ hence by a known theorem on Fourier integrals, (Reference: Titchmarsh, *Fourier Integrals*, Theorem 124, p. 170) $g(t) \geq 0$, and so $\|g\|_1 = \hat{g}(0)$, which we readily compute to be $r+1$. The modification of the construction to get $A_r = O(r)$ for r *odd* is left to the reader.

REMARK. Lemma 2 deals with a special case of a fundamental *extension problem* of harmonic analysis: given a function on a subset E of the reals, to determine whether it is the restriction to E of an L^1 Fourier transform $\hat{p}$, or more generally of a Fourier-

Stieltjes transform; and if so, to obtain a good upper bound for $\|p\|_1$ (for the total variation of the measure in the latter case). This problem was formulated and studied by Beurling [3b]. The search for best possible constants in Jackson's theorem leads naturally to this problem (see Lemma 3, below).†

Proof of Theorem 22. Let $p = p_{r,N}$ be the function constructed in Lemma 2 (where now $N = n$), and let J_r be any integrable function such that

(i) $\hat{J}_r(x) = i^{-r}\hat{p}(x), \quad |x| \geq n; \; i = \sqrt{-1}$.

(ii) $\hat{J}_r(k) = (ik)^{-r}, \qquad k = \pm 1, \pm 2, \ldots, \pm (n-1)$.

Clearly such J_r exist; we have only to make a smooth modification of $p_{r,n}$ in $|x| \leq n$ so that the modified function takes the required values at $x = \pm 1, \ldots, \pm (n-1)$. Now, by Lemma 1, $f = f^{(r)} * J_r$. Define $T = f^{(r)} * (J_r - (i^{-r})\, p)$. Since the Fourier transform of $J_r - (i^{-r})\, p$ is $\hat{J}_r - (i^{-r})\,\hat{p}$, which vanishes for $|x| \geq n$, $T \in \mathcal{T}_{n-1}$. Moreover,

$$\|f - T\| = \|f^{(r)} * p\| \leq \|p\|_1 \leq A_r n^{-r}$$

and the theorem is proved.

The function J_r gives us a convenient tool for obtaining an approximating polynomial for f from one for its derivative of order r:

THEOREM 23. *Let $f \in C^r$ have period 2π, and suppose there exists $T \in \mathcal{T}_{n-1}$ such that $\|f^{(r)} - T\| \leq M$. Then there exists $U \in \mathcal{T}_{n-1}$ such that*

$$\|f - U\| \leq A_r n^{-r} M$$

where A_r is as in Theorem 22.

---

† Also, paper [74].

Proof: Preserving the notations of the preceding proof, we have

$$\|(f^{(r)} * (i^{-r}) p) - (T * (i^{-r}) p\| \leq M\|p\|_1 = MA_r n^{-r}.$$

Now,

$$T_1 = T * ((i^{-r} p) \in \mathcal{T}_{n-1}$$

and $f^{(r)} * (i^{-r}) p = f^{(r)} * J_r + T_2 = f + T_2$, where

$$T_2 = f^{(r)} * ((i^{-r}) p - J_r) \in \mathcal{T}_{n-1}$$

therefore setting $U = T_1 - T_2$, we are done.

COROLLARY. *Let $f \in C^r$ have period 2π, and let $\omega(f^{(r)}; t)$ denote the modulus of continuity of $f^{(r)}$. Then, there exists $T \in \mathcal{T}_{n-1}$ such that*

$$\|f - T\| \leq BA_r n^{-r} \omega(f^{(r)}; t) .$$

Here A_r is as in Theorem 22, and B is an absolute constant.

This follows at once from Theorem 23, combined with the case $r = 0$ of the corollary to Theorem 20.

§4.4.2 Best possible constants.

Thus, we can now assert that the constant A_r in the general Jackson theorem (as formulated in the corollary to Theorem 20) is $O(r)$. In order to obtain a sharper result, we must refine Lemma 2 of the present section. Now, it is actually possible to assert that the A_r in Lemma 2 may be assumed *bounded*†, as a function of r. However, it is clear that we do not need $\hat{p}(x) = x^{-r}$ for *all* real x

† See "Notes and Comments."

satisfying $|x| \geq N$, but only for *integer* values. Indeed, it is possible to prove

LEMMA 3. *Given positive integers n, r there exists $p = p_{r,n} \in L^1(-\infty, \infty)$ such that*

(i) $\hat{p}(x) = x^{-r}$ *for integral* x, $|x| \geq n$.

(ii) $\|p\|_1 \leq B_r n^{-r}$, *where*

$$B_r = \begin{cases} \dfrac{4}{\pi} \displaystyle\sum_{m=0}^{\infty} \frac{1}{(2m+1)^{r+1}} & \text{if } r \text{ is odd} \\ \dfrac{4}{\pi} \displaystyle\sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)^{r+1}} & \text{if } r \text{ is even} \end{cases} \tag{2}$$

In particular, $B_1 = \pi/2$, *and* $B_r \leq \pi/2$, $r = 1, 2, \ldots$. *The conclusion becomes false if any smaller number is substituted for B_r in* (ii).

We shall not give the proof. Details may be found in Chapter 8 of Lorentz's book (where, however, the point of view is slightly different than in the present section).

Using Lemma 3 in place of Lemma 2 in the above reasoning, leads to this result:

Theorems 22 and 23 remain valid if A_r is replaced by B_r, where B_r is given by (2).

§4.4.3 Approximation of classes of functions.

As remarked in the Introduction, the fundamental problem of approximation theory is to approximate to a given function by a function in some prescribed class. For any concretely specified function, the search for a *best* approximation is, in general, fruitless. Suppose, however, that we do not specify the function but only give partial information about it. How closely can we assert that the function is approximable on the basis of this information? In other words, denoting by $E(f)$ the infimum of the distances from f to the allowable approximating functions we are asking about $\sup E(f)$ as f ranges over the class of functions determined by the specified information.

Thus, for example, if we know only that f belongs to the class D_r of 2π-periodic, r times differentiable functions having r-th derivative bounded by one, and the allowable approximating class is $\mathcal{T}_{n-1}$, we can assert (on the basis of the discussion in §4.4.2) that f is approximable to within $B_r n^{-r}$. Here it can be shown that B_r may not be replaced by any smaller constant; thus, we may say that the error of approximation of the *class* D_r by $\mathcal{T}_{n-1}$ is precisely $B_r n^{-r}$.

If D_r is enlarged slightly so as to include all functions such that $f^{(r-1)}(x)$ exists for all x, and $f^{(r-1)} \in$ Lip 1 with Lipschitz constant 1, (in short, all "weak solutions" of $|f^{(r)}(x)| \leq 1$) then D_r contains a "worst function," i.e., one whose distance from $\mathcal{T}_{n-1}$ is *exactly* $B_r n^{-r}$. This function has the property that its r-th derivative is alternately equal to plus and minus one on consecutive intervals of length π/n (details in Lorentz).

§4.5 *Localization theorem.*

We conclude this chapter with a simple but useful result which enables us to estimate, if f has a certain smoothness only on some interval (a, b) (but not necessarily on $(-\infty, \infty)$), the error $f-(f*K_\lambda)$ at points of (a, b). Not surprisingly, the crucial factor is how fast K falls off at ∞. Since we can, generally speaking, extend f from (a, b) to $(-\infty, \infty)$ with preservation of smoothness, it is enough (considering the difference between the original function and the altered one) to treat the case where f vanishes identically on (a, b).

THEOREM 24. *Let f be bounded and measurable, and $f(t) = 0$ for $a < t < b$. Then, for $a + \delta \leq t \leq b - \delta$ we have*

$$|f(t;\lambda)| \leq \phi(\lambda\delta)$$

where

$$\phi(u) = \int_{|t|>u} |K(t)|\,dt \,.$$

Proof: We can write $f = f_1 + f_2$, where

$$f_1(t) = 0, \quad t > a\,,$$

$$f_2(t) = 0, \quad t < b\,.$$

Then $f(t;\lambda) = f_1(t;\lambda) - f_2(t;\lambda)$

$$|f_1(t;\lambda)| = \left| \int_{-\infty}^{a} f_1(u)\,\lambda K(\lambda(t-u))\,du \right|$$

$$\leq M \int_{-\infty}^{a} \lambda\,|K(\lambda(t-u))|\,du = M \int_{\lambda(t-a)}^{\infty} |K(v)|\,dv$$

$$\leq M \int_{\lambda\delta}^{\infty} |K(v)|\,dv$$

if $t \geq a+\delta$, where $M = \sup|f|$. We obtain a similar estimate for $f_2(t;\lambda)$, and the theorem is proved.

Example. Let K = Cauchy kernel, and $f \in \text{Lip}\,\alpha (\alpha < 1)$ on $(-1, 1)$. Then f can be modified outside of $(-1, 1)$, so that the modified function f^* is in Lip α on $(-\infty, \infty)$, and so $|f^*(t;\lambda) - f^*(t)| < C\lambda^{-\alpha}$.

Moreover, since here $\phi(u) = O(1/u)$, we have for $|t| < 1-\delta$

$$|f^*(t;\lambda) - f(t;\lambda)| \leq C_1\lambda^{-\delta}$$

so that $f(t;\lambda) - f(t)$ is $O(\lambda^{-a} + \lambda^{-\delta})$, uniformly for $|t| < 1-\delta$.

Chapter IV: Footnotes added in proof

Page 57. [1] providing m is chosen so that $\omega_m = \omega_r$, i.e., m consists of atoms with "masses" $(-1)^{j-1}\binom{r}{j}$ at the points $j = 1, 2, \ldots r$.

Page 59. [2] As we have already indicated, $O(r)$ can be replaced here by $O(1)$ (see §4.4.2).

Page 60. [3] This choice is a clumsy one! For a more clever one, due to Kahane, see p. 118.

CHAPTER V

INVERSE THEOREMS

§5.1 Orientation, "classical" theory

Let us, in the first instance, talk about trigonometric polynomial approximation. Already in the last chapter (§4.4.3) we remarked that Jackson's theorem is, in a sense, best possible. Another kind of result in this direction is the famous theorem of S. Bernstein, that if $f \in C_{2\pi}$ is approximable to the order n^{-a} $(0 < a < 1)$ then $f \in \text{Lip}\ a$. This theorem was the starting point for a series of investigations as to which structural properties of a function can be inferred from a given degree of approximation (by, say, trigonometric polynomials). Here "structural properties," is meant in a quite general sense and can comprise such properties as Lipschitz smoothness, bounded variation, absolute continuity, analyticity, etc. In the present chapter we discuss some of these generalizations, and at the same time extend the theory to approximation of functions on $(-\infty, \infty)$ generated by an arbitrary L^1 kernel. First, for orientation, we will prove the cited theorem of S. Bernstein.

We require first

THEOREM 25. (Bernstein's inequality). *Let* $T \in \mathcal{T}_n$, $|T(x)| \leq 1$. *Then* $|T'(x)| \leq n$.

REMARK. *A weak form* of this is, under the same hypothesis, $|T'(x)| \leq An$, where A is independent of n. We prove only this weaker version, which is sufficient for the applications we have in mind. One may note that an upper bound of the order of n^2 is quite trivial to prove, but useless.

Proof (of weak version). Let $K \in L^1$ and $\hat{K}(x) = 1$ for $|x| \leq 1$. Moreover, suppose $\hat{K}$ is smooth enough so that dK/dt is integrable. Then, one verifies easily the identity $T = T * K_n$, i.e.,

$$T(t) = n \int_{-\infty}^{\infty} T(u)\, K(n(t-u))\, du$$

whenever $T \in \mathcal{T}_n$. Hence, for $T \in \mathcal{T}_n$ bounded by one,

$$|T'(t)| = \left| n^2 \int T(u)\, K'(n(t-u))\, du \right|$$

$$\leq Bn, \text{ where } B = \|K'\|_1 .$$

REMARK. To obtain the sharp constant (one) in place of B we are led to the following extremal problem: Find K having the smallest total variation such that $\hat{K}(m/n) = 1$ for all integers m such that $|m| \leq n$. (This is easily transformed into an instance of the Beurling minimization problem described in 4.4.1.)

Exercises.

1. Let f be bounded and measurable on $(-\infty, \infty)$, $K \in L^1$, K absolutely continuous with an L^1 derivative, and $\hat{K}(x) = 0$ for $|x| \geq A$. Let $g = f * K$ (g is, in a certain sense, a "low frequency function").[1] Prove $\|g'\| \leq BA\|g\|$ where B is an absolute constant (the norms are sup norms). (*Hint:* let h be a smooth integrable function such that $\hat{h}(x) = 1$ for $|x| \leq 1$, and establish the identity $g'(t) = \int g(t-u)\, A^2 h'(Au)\, du$.)

2. Extend Theorem 26 (weak version), and the previous exercise,

to L^p norms; to entire functions of exponential type $\leq n$, bounded on the real axis.

Notation. For any $f \in C_{2\pi}$, we define (the "degree of approximation"):

$$E_n(f) = \inf \|f - T\|, \quad T \in \mathcal{T}_n .$$

THEOREM 26. (S. Bernstein). *Let* $0 < a < 1$. *If* $E_n(f) = O(n^{-a})$, *then* $f \in \text{Lip}\ a$.

REMARK 1. The theorem is not true for $a = 1$ (see below).

REMARK 2. To appreciate the subtlety of Bernstein's proof, it is worthwhile to see first what a routine approach yields. If $\|f - T_n\| \leq An^{-a}$, the T_n are uniformly bounded, say $\|T_n\| \leq B$. Then, by Bernstein's inequality (weak form), $\|T_n'\| \leq Cn$ (as usual C, C_1, C_2 etc. denote constants).

Therefore

$$\begin{aligned} |f(x+h) - f(x)| &\leq |T_n(x+h) - T_n(x)| + 2An^{-a} \\ &\leq C_1(nh + n^{-a}) \\ &\leq C_2 h^{\frac{a}{1+a}} \end{aligned}$$

(if we choose n around $h^{-1/(1+a)}$). This shows $f \in \text{Lip}\ a/(1+a)$ rather than Lip a. The reason why we obtain an imperfect result is that the estimate $T_n'(x) = O(n)$ is too crude. The extra information that $\{T_n\}$ *converges with a certain rapidity* must be exploited to produce a better estimate for T_n'. If we can show that in fact $|T_n'(x)| \leq Cn^{1-a}$, then the sharp result will follow.

Proof of Bernstein's Theorem. By hypothesis

$$|T_n(x) - T_{2n}(x)| \le |T_n(x) - f(x)| + |f(x) - T_{2n}(x)|$$
$$\le 2An^{-a} .$$

By Bernstein's inequality, this gives

$$|T_n'(x) - T_{2n}'(x)| \le B(2n)^{1-a} .$$

Put $n = 2^{k-1}$. Then if $\beta = 1-a$, this inequality gives

$$|T_{2^k}'(x) - T_{2^{k-1}}'(x)| \le B(2^k)^\beta .$$

Now changing k to $k-1$, $k-2, \dots$ we obtain a system of inequalities

$$|T_{2^{k-1}}'(x) - T_{2^{k-2}}'(x)| \le B(2^{k-1})^\beta$$

$$|T_2'(x) - T_1'(x)| \le B2^\beta .$$

Now, adding these inequalities, we obtain, by virtue of the triangle inequality

$$|T_{2^k}'(x) - T_1'(x)| \le B2^{k\beta}[1 + \frac{1}{2^\beta} + \dots +] \le C2^{k\beta} .$$

Therefore

(1) $$|T_n'(x)| \le C_1 n^{1-a} \text{ if } n = 2^k, \ k = 1, 2, \dots .$$

Now, as in Remark 2, we obtain, using (1),

$$|f(x+h) - f(x)| \le C_2(n^{1-a}h + n^{-a}), \text{ if } n = 2^k .$$

Now choose n so that $1/h \leq n < 2/h$ (clearly, this interval contains a power of two). We obtain finally

$$|f(x+h) - f(x)| \leq C_2((\frac{2}{h})^{1-a}h + h^a) = C_3 h^a .$$

This proves the theorem.

REMARK 3. Case $a = 1$. The same reasoning as above shows that

$$|f(x+h) - f(x)| \leq C(kh + n^{-1}), \text{ if } n = 2^k,$$

$$= C[(\log_2 n)h + n^{-1}].$$

If $1/2h \leq n < 1/h$, then we get

$$|f(x+h) - f(x)| \leq C(h \log \frac{1}{h} + 2h) .$$

Thus $E_n(f) = O(1/n)$ implies f has a modulus of continuity which is $O(h \log 1/h)$. This result is unimprovable, as examples show, but unsatisfying because the converse is not true; a modulus of continuity $O(h \log 1/h)$ does *not* guarantee $E_n(f) = O(1/n)$, but only $E_n(f) = O((\log n)/n)$. The following theorem of Zygmund remedies this defect.

Let Z denote the class of all functions f satisfying

$$|f(x+h) - 2f(x) + f(x-h)| \leq Ch$$

where C is a constant (depending on f, but not on x or h).

THEOREM 27 (Zygmund). *A necessary and sufficient condition that $E_n(f) = O(1/n)$ is $f \in Z$.*

Proof: The sufficiency is a consequence of Theorem 21, with $r = 2$.

Necessity. If $E_n(f) \leq A/n$, then

(2) $|f(x+h) - 2f(x) + f(x-h)| \leq |T_n(x+h) - 2T_n(x) + T_n(x-h)| + \frac{4A}{n}$

and if $|T_n''(x)| \leq B$, we get

(3) $$|T_n(x+h) - 2T_n(x) + T_n(x-h)| \leq Bh^2 \ .$$

Now

$$|T_n(x) - T_{2n}(x)| \leq \frac{C}{n} \ ,$$

and therefore by Bernstein's inequality,

$$|T_n''(x) - T_{2n}''(x)| \leq C_1 n \ .$$

We now choose $n = 2^k$ and proceed as in the previous theorem to obtain

$$|T_n''(x)| \leq C_2 n \ .$$

Therefore, we obtain from (2) and (3),

$$|f(x+h) - 2f(x) + f(x-h)| \leq C_3 (nh^2 + \frac{1}{n}) \ .$$

Choose n such that $1/h \leq n < 2/h$; the result follows.

Similarly we have

THEOREM 28. *If* $E_n(f) = O(n^{-(1+\alpha)})$, $0 < \alpha < 1$ *then* f *has a derivative* $f' \in$ Lip α. *Equivalently, if* $E_n(f) = O(n^{-(1+\alpha)})$, *then*

(4) $$|f(x+h) - 2f(x) + f(x-h)| \leq Ch^{1+\alpha} \ .$$

Proof: As in the previous proofs, one first shows that if $\|f - T_n\| \leq A/n^{1+\alpha}$, then $|T_n''(x)| \leq Cn^{1-\alpha}$ whenever n is a power of two.

Therefore, for $n = 2^k$ we have

$$|f(x+h) - 2f(x) + f(x-h)| \le 4An^{-1-\alpha} + Ch^2n^{1-\alpha},$$

which is $O(h^{1+\alpha})$ for a suitable choice of n. This gives (4).

To derive the conclusion in the other form we start from

$$(5) \qquad |T'_{2^k}(x) - T'_{2^{k+1}}(x)| \le B2^{-k\alpha}$$

(a consequence of the hypotheses and Bernstein's inequality). We obtain from (5) the estimate $|T'_{2^k}(x)| \le B_1 2^{-k\alpha}$, and this, together with (5), implies that the sequence $\{T'_{2^k}(x)\}$ converges uniformly to some continuous function g. Since also $\{T_{2^k}\}$ converges uniformly to f, it follows that g is the derivative of f. Moreover, (5) implies (just as in the proof of Theorem 26) that $g \in \mathrm{Lip}\, \alpha$. This completes the proof.

We have incidentally established the following proposition, which strictly speaking has nothing to do with polynomial approximation:

COROLLARY. *Let* $0 < \alpha < 1$. *The following two subclasses of* $C_{2\pi}$ *are identical:*

(i) *The set of absolutely continuous functions* f *with* $f' \in \mathrm{Lip}\, \alpha$,

(ii) *The set of* f *for which* $\omega_2(f; h) = O(h^{1+\alpha})$.

Indeed, if f is in class (i), then $E_n(f) = O(n^{-1-\alpha})$ by the corollary to Theorem 20, hence by Theorem 28 it is in class (ii); and, if f is in class (ii), then $E_n(f) = O(n^{-1-\alpha})$ by Theorem 21, hence by Theorem 28 it is in class (i).

Since a fair amount of machinery is involved in this proof, it is natural to seek a direct demonstration. In one direction, this is easy:

using the identity

$$f(x+h) - 2f(x) + f(x-h) = \int_x^{x+h} [f'(t) - f'(t-h)]dt$$

it is evident that (i) is a subset of (ii).

The deduction in the other direction seems harder; at least, no really easy proof is known to us. (Another proof, independent of polynomial approximations, will be given later as a corollary of a general Tauberian-type theorem.) Even to deduce from (4) that f is absolutely continuous takes a little work–this is quite easy though if one uses the Fourier expansion of f.

Exercise. Suppose $f \epsilon C_{2\pi}$ and $\int_0^1 t^{-3}\omega_2(f;\ t)^2 dt < \infty$. Show that f is absolutely continuous and $f' \epsilon L^2(0, 2\pi)$. (*Hint:* consider the Fourier expansion of f.)

Returning to Theorem 28, one can introduce the more general hypothesis $E_n(f) = O(\psi(n))$, where $\psi(n)$ may be $n^{-(r+a)}$ (r a positive integer) or some still more general function. The result one obtains is, of course, that the more rapidly $\psi(n)$ decreases, the more smooth f has to be; in the former case, for instance, f must have an r-th derivative of class Lip a. And exponential decay of $\psi(n)$ implies analyticity of f. The techniques for proving such results are similar to those we have used in this section, and we shall not give details here, especially as this section is mainly illustrative, and we shall develop a new and more general approach to inverse theorems later. Let us content ourselves with one final example, in which the hypothesis is *weaker* than in Theorem 28.

THEOREM 29. *If* $\sum_{n=1}^{\infty} E_n(f) < \infty$, *then* $f \epsilon$ Lip 1.

Proof: It is sufficient to prove that if $|f(x) - T_n(x)| \leq \delta_n$ where

$\delta_n \downarrow$ and $\Sigma\, \delta_n < \infty$, then $T'_{2^k}(x)$ are uniformly bounded, because then we should have, for any x_1, x_2

$$|T_{2^k}(x_1) - T_{2^k}(x_2)| \leq B\,|x_2 - x_1|\,,$$

and letting $k \to \infty$ gives the desired result. Using the triangle inequality, we get

$$\|T_n - T_{2n}\| \leq \delta_n + \delta_{2n} \leq 2\delta_n \;.$$

Using Bernstein's inequality, it follows that

$$\|T'_n - T'_{2n}\| \leq A_1 n\, \delta_n \;.$$

Put $n = 2^{k-1}$. Then

$$\|T'_{2^k} - T'_{2^{k-1}}\| \leq A_1\, 2^{k-1}\, \delta_{2^{k-1}} \;.$$

Changing k to $k-1, k-2, \ldots$ and adding, we obtain by means of the triangle inequality

$$|T'_{2^k}(x)| \leq A_1 \sum_{r=0}^{k-1} 2^r \delta_{2^r} \leq A_1 \sum_{r=0}^{\infty} 2^r \delta_{2^r} \leq A_2\,,$$

by the following lemma of Cauchy: *If* δ_n *are positive and decreasing and* $\Sigma\, \delta_n < \infty$, *then* $\Sigma\, 2^r \delta_{2^r} < \infty$. This completes the proof.

§5.2 *Comparison theorems.*

Before turning to inverse theorems in a more general setting, we wish to discuss a closely related topic, namely comparison of the error of approximation corresponding to two different kernels. We shall give a precise quantitative formulation of our earlier remarks to the effect that the approximation properties of a given kernel depend on the flatness at $x = 0$ of its Fourier transform. From here on, somewhat deeper aspects of Fourier theory come in, especially the notion of divisibility in the ring of Fourier-Stieltjes transforms (see §5.2.1). First of all, we wish to formulate the problem that shall occupy us in the most suitable fashion. For an L^1 kernel K, we have for the error of approximation

$$e(t;\lambda) = f(t)-f(t;\lambda) = f(t) - \int f(t-\frac{u}{\lambda})K(u)du .$$

This may be rewritten in the form

$$e(t;\frac{1}{a}) = \int f(t-au)\,d\sigma(u) \tag{6}$$

where $a = 1/\lambda$, and σ is the measure defined by

$$d\sigma(u) = d\delta(u) - K(u)du . \tag{7}$$

Here δ is the "Dirac measure" (unit point mass at $u = 0$). Thus, the error of approximation is expressible as the integral (6), where σ is a finite measure on the line and $\int d\sigma = 0$, and we are interested in the behavior of the integral as $a \to 0+$. Now, in posing this question it is no longer necessary to restrict attention to measures of the type (7). For example, suppose σ is the "dipole measure": unit positive mass at $u = +1$ and unit negative mass at $u = -1$. Then (1) becomes $f(t-a) - f(t+a)$. Thus, if we write

$$D_\sigma(f;a) = \sup_t \left| \int f(t-au)d\sigma(u)\right| \tag{8}$$

the statement $D_\sigma(f;a) = O(a^\alpha)$ expresses that $f \in \mathrm{Lip}\,\alpha$; whereas, if σ is defined by (7), the corresponding statement expresses that $\|f-(f*K_\lambda)\| = O(\lambda^{-\alpha})$. Thus, both "smoothness" and "degree of approximation" assertions are expressible in terms of expressions $D_\sigma(f;a)$ —we shall call such an expression the *σ-deviation* of f—

of the form (8), with suitable measures σ. From this point of view, there is no essential difference between a "direct" and an "inverse" theorem of approximation theory.[2] Both kinds of theorems are merely comparison theorems relating the behavior of $D_{\sigma_1}(f;\, a)$ to that of $D_{\sigma_2}(f;\, a)$ for a pair of measures σ_1, σ_2. This will be our point of view throughout the remainder of these notes. The reader acquainted with Wiener's work may notice that, in this formulation, the theorems of approximation theory look very like Tauberian theorems, and in fact we shall make use of "Tauberian" ideas.

Notice, too, that it is now natural to study the problem for any measure σ with $\int d\sigma = 0$. Thus, e.g. one formulation of the saturation problem is, to find a function $e(t)$ defined and increasing for $t > 0$ such that the set of f for which $D_\sigma(f;\, a) = o(e(a))$ $(a \to 0+)$ is a *finite dimensional* vector space, but the set of f for which $D_\sigma(f;\, a) = O(e(a))$ (the "saturation class" of σ) is not finite dimensional. For example, when σ is the "dipole" measure, we may take $e(t) = t$, and the saturation class is (by definition) Lip 1. (A more careful discussion should, of course, make the definition of the saturation class independent of possible ambiguity in the choice of $e(t)$.)

Exercise. If $f \in C$ and β is any finite measure, and $g(t) = \int f(t-u)\, d\beta(u)$, then $D_\sigma(g;\, a) \le V(\beta) D_\sigma(f;\, a)$, where V denotes total variation (this result partly justifies our earlier assertion, that convolution increases the smoothness of a function "as measured by nearly any criterion").

§5.2.1 Fourier-Stieltjes transforms.

We denote by W the (Wiener) ring of Fourier-Stieltjes transforms

$$S(x) = \hat{\sigma}(x) = \int e^{-itx} d\sigma(t)$$

where σ is a finite measure on the line. W is a commutative ring with unit element, with respect to ordinary multiplication. The norm in W is $\|\hat{\sigma}\| = \int |d\sigma|$. We assume known the basic properties of W, and notably the following fundamental theorem of N. Wiener.

LEMMA (Wiener). *Let $S_1, S_2 \in W$, and E a compact set of reals. If S_1 vanishes nowhere in E, and S_2 vanishes everywhere outside E, then S_1 divides S_2 i.e., $S_2 = S_1 S_3$ for some $S_3 \in W$.*

REMARK. In most of the applications we will make, the depth of this lemma is not really needed, since S_2/S_1 will be fairly smooth (say, Lip 1) and of compact support, therefore in W (even W_0) by elementary theorems. (By W_0 we denote the subring (actually an ideal) of W consisting of Fourier (-Lebesgue) transforms, i.e., those $\hat{\sigma}$ for which σ is absolutely continuous, hence $d\sigma(t) = K(t)dt$, with $K \in L^1$.)

§5.2.2 A comparison theorem based on global divisibility.

We begin with a very simple "comparison theorem," based on *global* divisibility of the Fourier transforms.

THEOREM 30. *If $\hat{\sigma}_1$ divides $\hat{\sigma}_2$, say $\hat{\sigma}_2 = \hat{\sigma}_1 S$ where $S \in W$, then for every bounded continuous f*

$$D_{\sigma_2}(f;\, a) \le \|S\|_W D_{\sigma_1}(f;\, a)\ .$$

Proof: The relation $\hat{\sigma}_2 = \hat{\sigma}_1 S$ corresponds, in the domain of measures, to the convolution relation $\sigma_2 = \sigma_1 * \sigma_3$ where $S = \hat{\sigma}_3$, in other words for every bounded continuous g we have

$$\int g\, d\sigma_2 = \int\int g(u+v)d\sigma_1(u)d\sigma_3(v) \ .$$

Hence,

$$\left| \int f(t-au)d\sigma_2(u) \right| = \left| \int\int f(t-au-av)d\sigma_1(u)d\sigma_3(v) \right|$$

$$\leq \int |d\sigma_3| \cdot \sup_v \left| \int f(t-au-av)d\sigma_1(u) \right|$$

$$= \|S\|_W \cdot D_{\sigma_1}(f;\, a) \ .$$

Taking the sup with respect to t gives the result.

REMARK. Clearly, the same argument proves the following more general assertion: If σ belongs to the ideal generated by $\sigma_1, \ldots, \sigma_n$, then

$$D_\sigma(f;\, a) \leq \sum_{i=1}^{n} A_i D_{\sigma_i}(f;\, a) \tag{9}$$

for suitable constants A_i (depending only on σ and the σ_i). As an evident consequence of Theorem 30, we have

THEOREM 31. *Let J, K be L^1 kernels, and suppose $1-\hat{J}$ divides $1-\hat{K}$ in W. Then for every bounded continuous f*

$$\|f-(f*K_\lambda)\| \leq A\,\|f-(f*J_\lambda)\| \tag{10}$$

where A is a constant depending only on the kernels J, K.

REMARK. As D. L. Ragozin has kindly pointed out to the author, the converse of Theorem 30 (and hence that of Theorem 31) is true†, i.e., an inequality of the type (10) implies that $1-\hat{J}$ divides $1-\hat{K}$ (and, more generally, (9) implies that $\hat{\sigma}$ belongs to the ideal in W generated by $\hat{\sigma}_1, \ldots, \hat{\sigma}_n$). Thus, divisibility relations on W belong quite *essentially* to the study of comparison theorems. However, under the weaker hypothesis of *local* divisibility (at 0) we can still prove useful results (see below).

DEFINITION. When (10) holds, we shall say K *is better than* J. Two kernels, each better than the other, are called equivalent. A test which is often applicable is given by

COROLLARY 1. *Let J, K be L^1 kernels such that $1-\hat{J}$ divides $\hat{J}-\hat{K}$ in W. Then, K is better than J.*

Proof: Suppose $\hat{J}-\hat{K} = (1-\hat{J})S$, where $S \in W$. Then $(1-\hat{K})/(1-\hat{J}) = 1+S$ is in W, and the result follows from Theorem 31.

COROLLARY 2. *The Fejér-de la Vallée Poussin and Cauchy kernels are equivalent.*

Proof: We have here

$$\hat{J}(x) = \begin{cases} 1-|x|, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases}$$

and $\hat{K}(x) = e^{-|x|}$.

Now,

$$G(x) = \frac{\hat{J}(x) - \hat{K}(x)}{1-\hat{J}(x)}$$

† See "Notes and Comments."

is an even function, and for $x > 0$ we have

$$G(x) = \begin{cases} \dfrac{1-x-e^{-x}}{x}, & 0 < x \leq 1 \\ -e^{-x}, & x \geq 1 . \end{cases}$$

Plainly $G \in W$, in fact $G \in W_0$ since it is absolutely continuous, and its derivative has finite variation, hence its inverse Fourier transform is $O(t^{-2})$ at infinity.

In like manner, writing

$$H(x) = \frac{\hat{K}(x) - \hat{J}(x)}{1-\hat{K}(x)}$$

we have, for $x > 0$

$$H(x) = \begin{cases} \dfrac{e^{-x}-1+x}{1-e^{-x}}, & 0 < x \leq 1 \\ \dfrac{e^{-x}}{1-e^{-x}}, & x \geq 1 \end{cases}$$

which is, for the same reason as G, an L^1 transform. (It is, moreover, not hard to estimate the norms of G, H.) This proves the corollary.

REMARK. It is instructive to compare via Fourier transforms one of the above kernels (say K) with the "dipole measure" σ referred to above. Neither the ratio $(1-\hat{K})/\hat{\sigma}$ nor its reciprocal is in W, in fact both are discontinuous at $x = 0$. Thus, in view of the cited converse to Theorem 30, neither a relation of the type $\|f - f * K_\lambda\| \leq \sup_t |f(t+\lambda^{-1}) - f(t)|$ nor a relation of the type $\omega(f; \lambda^{-1}) \leq A\|f - f * K_\lambda\|$, is possible. This strongly suggests (although it does not prove) that the common saturation class of the above two kernels neither includes, nor is included in, Lip 1. This is in fact true, and we shall return to this point later.

§5.2.3 A comparison theorem based on local divisibility.

The examination of certain examples suggests that a conclusion similar to that of Theorem 30 can be drawn in case $\hat{\sigma}_1$ only divides $\hat{\sigma}_2$ *locally* (at $x = 0$). More precisely, we say a function $G(x)$ belongs to W *at the point* x_0 if there is an $S \in W$ such that $S(x) = G(x)$ in a neighborhood of x_0. If now, $\hat{\sigma}_2/\hat{\sigma}_1$ belongs to W at $x = 0$, can we assert that

(11) $$D_{\sigma_2}(f; a) \leq AD_{\sigma_1}(f; a)$$

where A depends only on σ_1 and σ_2? As remarked above, in general we *cannot*, that is, the *global* nature of $\hat{\sigma}$, not merely its behavior at $x = 0$, influence the order of magnitude of $D_\sigma(f; a)$. Nevertheless, a somewhat weaker conclusion than (11), which implies (11) in many cases of interest, can be drawn.

THEOREM 32. *Let* σ_1 *be a real finite measure, and suppose* $\hat{\sigma}_1$ *divides* $\hat{\sigma}_2$ *at* $x = 0$. *Suppose for some* $x_0 > 0$ *and* $b < 1$, $\hat{\sigma}_1$ *does not vanish on the interval* $[b^2 x_0, x_0]$. *Then for every bounded continuous* f

(12) $$D_{\sigma_2}(f; a) \leq CD_{\sigma_1}(f; a) + A \sum_{i=0}^{\infty} D_{\sigma_1}(f; Bb^i a).$$

Here A, B, C *are positive constants depending only on* σ_1 *and* σ_2.

REMARK. If σ_1 is not the zero measure, then it must satisfy the hypothesis of the theorem, since the equation $\hat{\sigma}_1(x) = \hat{\sigma}_1(-x)$

together with the continuity of $\hat{\sigma}_1$ imply the existence of an interval on the positive axis where $\hat{\sigma}_1(x)$ does not vanish.

Proof: Let $P(x)$ be the function which equals one for $|x| \leq b^2 x_0$, zero for $|x| > bx_0$ and is piecewise linear ("trapezoid function"). Then, $P = \hat{p}$ for a certain $p \in L^1$. Let

$$Q(x) = P(x) - P(bx) .$$

Then, $Q = \hat{q}$, where

$$q(t) = p(t) - \frac{1}{b}\, p(\frac{t}{b}) .$$

Now, $Q(x) = 0$ except for x in the intervals $(b^2 x_0, x_0)$ and $(-x_0, -b^2 x_0)$, and on these intervals $\hat{\sigma}_1$ does not vanish by hypothesis and skew-symmetry. Therefore, by Wiener's Lemma, $\hat{\sigma}_1$ divides Q, say $\hat{\sigma}_1 S = Q$ for some $S \in W$, and therefore by Theorem 30, we have for all t

$$\left| \int f(t-au)\,[p(u) - \frac{1}{b}\, p(\frac{u}{b})]\, du \right| \leq A\, D_{\sigma_1}(f;\, a) \tag{13}$$

where A depends only on σ_1. Let us write $D_1(a)$ as an abbreviation for $D_{\sigma_1}(f;\, a)$. Replacing a by ba in (13) and making the change of variable $bu = v$ gives

$$\left| \int f(t-av)[\, \frac{1}{b} p(\frac{v}{b}) - \frac{1}{b^2}\, p(\frac{v}{b^2})]\, dv \right| \leq A\, D_1(ba) .$$

Repeating this process, adding and using the triangle inequality gives, if m is a positive integer

$$\left| \int f(t-au)[p(u) - \frac{u}{b^m}\, p(\frac{u}{b^m})]\, du \right| \leq A \sum_{i=0}^{m-1} D_1(b^i a)$$

$$\leq A \sum_{i=0}^{\infty} D_1(b^i a) .$$

Now, for fixed a, t

$$\lim_{\lambda\to\infty} \int f(t-au)\,\lambda p(\lambda u)\,du = \lim_{\lambda\to\infty} \int f(t-\frac{av}{\lambda})\,p(v)\,dv = f(t)$$

by dominated convergence; therefore, letting $m\to\infty$ in the last inequality,

$$\left|f(t)-\int f(t-au)\,p(u)\,du\right| \leq A \sum_{i=0}^{\infty} D_1(b^i a) \;. \tag{13}$$

Now, by hypothesis, there exists a positive number c, and $G \in W$ such that $\hat{\sigma}_1 G(x) = \hat{\sigma}_2(x)$ for $|x| \leq c$. Now, $P(x_0 x/c)$ vanishes for $|x| \geq c$, hence

$$\hat{\sigma}_1(x)G(x) - \hat{\sigma}_2(x) = [\hat{\sigma}_1(x)G(x) - \hat{\sigma}_2(x)][1 - P(\frac{x_0 x}{c})]$$

holds for all x. This shows that $\hat{\sigma}_2$ belongs to the ideal generated by $\hat{\sigma}_1$ and $1 - P(x_0 x/c)$. Therefore, by (9)

$$D_{\sigma_2}(f;\, a) \leq A_1 D_{\sigma_1}(f;\, a) + A_2 D_{\sigma}(f;\, a) \tag{14}$$

where $d\sigma(t) = d\delta(t) - (c/x_0)p(ct/x_0)dt$ (so that $\hat{\sigma}(x) = 1 - P(x_0 x/c)$). The constants $A_1 = \|G\|_W$ and $A_2 = \|\hat{\sigma}_1 G - \hat{\sigma}_2\|_W$ depend only on σ_1 and σ_2. Now,

$$\left|\int f(t-au)d\sigma(u)\right| = \left|f(t) - \int f(t-au)\cdot \frac{c}{x_0}\,p(\frac{cu}{x_0})\,du\right|$$

$$= \left|f(t) - \int f(t-Bav)\,p(v)\,dv\right|$$

$$\leq A \sum_{i=0}^{\infty} D_1(b^i B a)$$

by (13), where $B = x_0/c$ depends only on σ_1 and σ_2. Therefore, the last expression on the right is an upper bound for $D_\sigma(f;\ a)$ and so, from (14)

$$D_{\sigma_2}(f;\ a) \leq A_1 D_1(a) + A_2 A \sum_{i=0}^{\infty} D_1(b^i Ba)$$

where, we recall, $D_1(a)$ denotes $D_{\sigma_1}(f;\ a)$. This proves (12), except for a slight change of notation.

REMARKS. 1. In the important special case where $\hat{\sigma}_1 = 1 - \hat{K}$, with $K \in L^1$, $\hat{\sigma}_1$ does not vanish for large x, hence any b with $0 < b < 1$ may be chosen.

2. Observe that if $\hat{\sigma}_2$ vanishes in a whole neighborhood of 0, the divisibility hypothesis is satisfied for every choice of σ_1.

3. From (12) follows, in particular that $D_{\sigma_1}(f;\ a) = O(a^\alpha)$ implies $D_{\sigma_2}(f;\ a) = O(a^\alpha)$ $(a > 0)$.

4. It is also clear that the corresponding "little o" implication is valid. More generally, if $D_{\sigma_1}(f;\ a) \leq \psi(a)$, where ψ increases with a, $D_{\sigma_2}(f;\ a)$ can easily be shown to be less than $A \int_0^{Ba} u^{-1} \psi(u) du$ for suitable constants A, B. This remark also applies to later theorems where, to keep the discussion simple, we usually suppose bounds of the type $\psi(a) = Ca^\alpha$.

Theorem 32 has a number of applications, some of which shall be presented in the following paragraphs.

§5.3 *Applications of the comparison theorems.*

§5.3.1 Modulus of smoothness, and differentiability.

As a first application, let us consider the Corollary following Theorem 28. We restrict attention to the "hard" part. Suppose, then, for some continuous bounded f (note that there will be no need here to assume f periodic)

(15) $$|f(t+h) - 2f(t) + f(t-h)| \leq Ah^p, \quad h > 0$$

where $p = 1+\alpha > 1$, and A is independent of t and h. We wish to prove two things: (i) f is absolutely continuous, and (ii) $|f'(t+h) - f'(t)| \leq Bh^{p-1}$ for all $h > 0$, with B independent of t and h.

Before proceeding, it will be convenient to introduce the following notation. By the *binomial measure* β_r $(r \geq 1)$ we denote the measure having "mass" $(-1)^j \binom{r}{j}$ at the point j $(j = 0, 1, \ldots, r)$. Note that $\hat{\beta}_r(x) = (1-e^{-ix})^r$. ($\binom{r}{j}$ denote binomial coefficients.)

Let us first consider a function f of class C^1 which satisfies (15). We may rewrite (15) in the equivalent form

$$\left| \int f'(t+u)\, J\left(\frac{u}{h}\right) du \right| \leq Ah^p$$

where J is equal to -1 on $[-1, 0]$, $+1$ on $[0, 1]$, and 0 elsewhere. This may be further rewritten as

$$\left| \int f'(t-au)\, d\sigma(u) \right| \leq Aa^{p-1}$$

that is $D_\sigma(f'; a) \leq Aa^{p-1}$ where we replace h by a and

$$d\sigma = J(t)dt, \qquad \hat{\sigma}(x) = -2i\left(\frac{1-\cos x}{x}\right)$$

Now, $\hat{\beta}_1(x) = 1-e^{-ix}$, and it is easy to check that $\hat{\sigma}$ divides $\hat{\beta}_1$ at 0 (as well as the reciprocal relation), therefore, taking σ, β_1 as σ_1, σ_2 respectively in Theorem 32 we get

(16) $$D_{\beta_1}(f'; a) \leq MAa^{p-1}$$

where the constant M does not depend on a or A.

Given now an arbitrary $f \in C$ which satisfies (15), denote by K the Cauchy kernel, and consider $f^{\lambda} = f * K_{\lambda}$. It is readily verified that f^{λ} also satisfies (15), and therefore also (16), hence

$$(17) \qquad |Df^{\lambda}(t+a) - Df^{\lambda}(t)| \leq M A a^{p-1} \quad \left(D = \frac{d}{dt}\right).$$

This shows that the functions Df^{λ} are uniformly equicontinuous, and so there exists a sequence $\lambda_n \to \infty$ such that $\{Df^{\lambda_n}\}$ converges uniformly on compact sets to a continuous function g. This g satisfies, for all real u, v

$$\int_u^v g(t)dt = \lim_{n\to\infty} \int_u^v Df^{\lambda_n}(t)dt = \lim_{n\to\infty} [f^{\lambda_n}(v) - f^{\lambda_n}(u)]$$

$$= f(v) - f(u)$$

which shows that f is absolutely continuous, and $f' = g$. Now letting λ tend to infinity through the sequence λ_n in (17) gives $|f'(t+a) - f'(t)| \leq M A a^{p-1}$, i.e., f satisfies (ii) and we are done. (Note that in the present example $\hat{\sigma}$ does not divide $\hat{\beta}_1$ globally, so we really need Theorem 32.)

Clearly this argument can be carried out with h^p replaced by other functions of h in (15). We obtain the following result:

THEOREM 33. *Suppose* $f \in C$ *and*

$$(18) \qquad \int_0^1 u^{-2}\omega_2(f;\, u)du < \infty .$$

Then, f *is absolutely continuous, and* f' *is continuous with a modulus of continuity which is* $O(\int_0^h u^{-2}\omega_2(f;\, u)du)$.

The only point to check is the integral form of the condition involving ω_2. Let us write $\phi(u)$ for $(\omega_2(f;u))/u$, then the upper bound we get for $D_{\beta_1}(f';a)$ is a constant times $[\phi(a)+\sum_{i=0}^{\infty}\phi(Bb^i a)]$. Comparing the infinite sum with an integral (we skip some details here) we approximate $\sum_{i=0}^{\infty}\phi(Bb^i a)$ by

$$\int_0^{\infty}\phi(Bb^t a)dt = (-\log b)\int_0^1 \phi(Bau)\,\frac{du}{u}$$

$$= (-\log b)\int_0^{Ba} u^{-2}\,\omega_2(f;u)du$$

which implies the stated result.

REMARK. It can be shown by examples that Theorem 33 is "best possible," i.e., a condition essentially weaker than (18) cannot force the uniform continuity of f'. It is of interest to inquire just what "structural properties" of f are implied by conditions weaker than (18). We have already mentioned that for periodic f the condition $\int_0^1 u^{-3}(\omega_2(f;u))^2du < \infty$ implies that f is absolutely continuous, and $f' \in L^2$. This result, and the above, are very special cases of "embedding theorems" for so-called *Besov spaces*; see [64a, 66, 73, 75] (many-variable case), and more special results in [77] (related material also in [76]).

§ 5.3.2 New deduction of Jackson's theorem.

Let us show how to deduce Jackson's theorem (in the form of Theorem 21) from Theorem 32. (We do not hesitate to include a "direct" theorem in this chapter, since, as we have remarked, "direct" and "inverse" theorems, at least in the context of approximations of the form $f * K_\lambda$, are very much the same thing.)

Choose a function, call it $\hat{K}$, which vanishes for $|x| \geq 1$ and is of class C^2. Moreover, arrange that $1-\hat{K}$ is divisible by $\hat{\beta}_r = (1-e^{-ix})^r$ at 0 (a rather drastic way to ensure this would be to let $\hat{K}(x) = 1$ for $|x| \leq \frac{1}{2}$).[3] The kernel $K(t) = 1/2\pi \int \hat{K}(x)e^{ixt}dx$ is then in L^1, and $T = f * K_n \in \mathcal{T}_{n-1}$. Moreover, Theorem 32 implies, since $D_\beta(f;\, a) \leq \omega_r(f;\, a)$

$$\|f - T\| \leq C\omega_r(f;\, n^{-1}) + A \sum_{i=0}^{\infty} \omega_r(f;\, Bb^i n^{-1})$$

and the right side is easily shown not to exceed $M_r \int_0^{1/n} t^{-1} \omega_r(f;t)\,dt$, where M_r depends only on r. This ia slightly weaker than the result sought; however, if $\omega_r(f;\, t) \leq Nt^p$ for some $p > 0$, the upper bound obtained for $\|f - T\|$ is $p^{-1}NM_r n^{-p}$, which has the right order of magnitude. Theorem 21 in full generality can easily be deduced from this special form.

Exercise. Given that every $f \in C_{2\pi}$ satisfying $|f(x) - f(y)| \leq |x - y|$ admits approximation by $T \in \mathcal{T}_{n-1}$ with an error $\leq A_1 n^{-1}$ (A_1 absolute constant), deduce that every $f \in C_{2\pi}$ is approximable by $T \in \mathcal{T}_{n-1}$ with an error $\leq A_2 \omega(f;\, n^{-1})$. (*Hint:* Approximate f first by a suitable broken line function.)

§5.3.3 The "moving average," revisited.

Let's look again at the "moving average" process of §1.2. Denoting its kernel by J (i.e., $J(t)$ is ½ times the characteristic function of $[-1, 1]$), we have

$$\left| f(t) - \frac{1}{2a} \int_{t-a}^{t+a} f(u)\,du \right| \leq D_\sigma(f;\, a)$$

where $\hat{\sigma}(x) = 1 - \hat{J}(x) = 1 - (\sin x)/x$. Since $\hat{\sigma}$ divides $\hat{\beta}_2(x) = (1 - e^{-ix})^2$ *globally* in W, we obtain from Theorem 30 the result that

$$\omega_2(f;\, a) \leq A \sup_t \left| f(t) - \frac{1}{2a} \int_{t-a}^{t+a} f(u)du \right| \tag{19}$$

where A is a numerical constant. This result reconfirms our earlier findings on the saturation behavior of the moving average, but gives new information (an "inverse theorem") when the right side of (19) has a larger order of magnitude than a^2. Observe that here we could get by with the rather trivial Theorem 30, rather than Theorem 33, because of the fortuitous circumstance of *global* divisibility.

Exercises. 1. Prove that $\hat{\sigma}$ divides $\hat{\beta}_2$ in W, and obtain an estimate for A in (19).

2. Carry out a similar analysis for each of the five kernels listed in Remark 3 of §2.2.

§5.4 *Special comparison theorems with no divisibility hypothesis.*

It is a remarkable fact that for certain measures ρ an estimate for $D_\rho(f;\, a)$ in terms of $D_\sigma(f;\, a)$ is possible with a completely arbitrary choice of σ. Results of this kind yield as corollaries the "classical" inverse theorems surveyed in §5.1. We shall prove

THEOREM 34. *Let σ be a non-null real measure on R^1, $f \in C$ and $D_\sigma(f;\, a) \leq Ma^\alpha$, where M, α are positive numbers independent of a. Suppose, for some positive integer r, $x^{-r}\hat{\rho}(x)$ coincides in a neighborhood of 0 with an element of W. Then*

$$D_\rho(f;a) \leq \begin{cases} C_1 a^\alpha, & \text{if } \alpha < r \\ (C_2\,|\log a| + C_3)a^\alpha, & \text{if } \alpha = r \\ C_4 a^r, & \text{if } \alpha > r\,. \end{cases}$$

The C_i are constants depending on f, σ, and r but not on a.

For the proof we require a lemma.

LEMMA. *Let $\phi(x)$ be defined for $x \geq 0$, and suppose for all x, $0 \leq \phi(x) \leq M$ and*

$$\phi(x) \leq Ax^{\alpha} + C^{-\beta}\phi(Cx) \tag{20}$$

where α, β, A, C are positive constants, with $C > 1$. Then

$$\phi(x) \leq \begin{cases} \left(\frac{A}{1-C^{\alpha-\beta}}\right)x^{\alpha}, & \text{if } \alpha < \beta \\ (A_1|\log x| + A_2)x^{\alpha}, & \text{if } \alpha = \beta \\ A_3x^{\beta}, & \text{if } \alpha > \beta \end{cases}$$

where the A_i are independent of x.

Proof: Replacing x by Cx in (20) gives

$$\phi(Cx) \leq AC^{\alpha}x^{\alpha} + C^{-\beta}\phi(C^2x)$$

and substituting this into (20) we get

$$\phi(x) \leq (1 + C^{\alpha-\beta})Ax^{\alpha} + C^{-2\beta}\phi(C^2x) . \tag{21}$$

Now we can replace x by C^2x in (20), and substitute into (21). If we iterate this process, a simple induction shows, for every $n \geq 1$

$$\phi(x) \leq \left(\sum_{i=0}^{n-1} C^{(\alpha-\beta)i}\right)Ax^{\alpha} + C^{-n\beta}\phi(C^nx) . \tag{22}$$

We now consider cases.

(i) If $\alpha < \beta$, we may let $n \to \infty$ in (22) and get

$$\phi(x) \leq \left(\sum_{i=0}^{\infty} C^{(\alpha-\beta)i} \right) A x^{\alpha} ,$$

proving the first part of our assertion. In proving the remaining parts, it is clearly no loss of generality to assume that $x \leq 1$.

(ii) If $\alpha = \beta$ we get from (22)

$$\phi(x) \leq nAx^{\alpha} + MC^{-n\alpha}$$

and if we determine n so that $1/Cx \leq C^n < 1/x$, we get

$$\phi(x) \leq \left(\frac{A}{\log C} \right) x^{\alpha} \log \frac{1}{x} + MC^{\alpha} x^{\alpha} ,$$

(iii) If $\alpha > \beta$ we get from (22)

$$\phi(x) \leq \left(\frac{C^{(\alpha-\beta)n}}{C-1} \right) A x^{\alpha} + MC^{-n\beta} .$$

Choosing n as we did in (ii) gives

$$\phi(x) \leq \frac{A}{C-1} x^{\beta} + MC^{\beta} x^{\beta}$$

and the Lemma is proved.

Proof of Theorem 34. Let $P(x)$ be a twice differentiable function such that $P(x) = x^r$ for $|x| \leq 1$, and P vanishes for $|x| \geq 2$. Then $P = \hat{p}$, for a certain $p \in L^1$. Now, $P(2x) - 2^r P(x)$ vanishes for $|x| \leq ½$, and therefore by Theorem 32 the τ-deviation of f, where $d\tau = [½ p(t/2) - 2^r p(t)]dt$, does not exceed

$$C(Ma^{\alpha}) + A \sum_{i=0}^{\infty} M(Bb^i a)^{\alpha} = B_1 a^{\alpha}$$

where B_1 is a constant depending only on σ and r. That is,

$$\left| \int_{-\infty}^{\infty} f(t-au)[\tfrac{1}{2}p(u/2) - 2^r p(u)]du \right| \leq B_1 a^\alpha$$

$$\text{(23)} \qquad \left| \int_{-\infty}^{\infty} f(t-2av)\,p(v)dv - 2^r \int_{-\infty}^{\infty} f(t-au)\,p(u)du \right| \leq B_1 a^\alpha .$$

Now, for fixed t, let us write $\phi(a) = |\int f(t-au)\,p(u)du|$. Then, ϕ is bounded for $a > 0$, and from (23)

$$\phi(a) \leq 2^{-r} B_1 a^\alpha + 2^{-r}\phi(2a).$$

Applying the lemma (with $C = 2$ and $\beta = r$) shows us, for the *particular* measure ρ defined by $d\rho = pdt$, that D_p admits the estimates stated. For the general case we have only once more to apply Theorem 32, where now $d\sigma$ is pdt, *and* ρ is any measure such that x^r divides $\hat{\rho}$ at $x = 0$, and the proof is completed.

Since $\rho = \beta_r$ satisfies the hypothesis of Theorem 34, we have

COROLLARY. *if* $f \in C$ *and* σ *is a non-null measure on* R^1 *such that* $D_\sigma(f; a) = O(a^\alpha)$, *where* $\alpha > 0$, *then the r-th modulus of smoothness of* f *is large-O of* a^α, $a^\alpha \log 1/a$, or a^r *as* $a \to 0$ *according as* α *is less than, equal to, or greater than* r.

THEOREM 35. *Theorem 34 remains true if* r *is any positive real number, and the hypothesis concerning* ρ *is altered to* "$|x|^{-r}\rho(x)$ *coincides in a neighborhood of* 0 *with an element of* W."

The proof is the same. Note that Theorem 35 is applicable to the Fejér-de la Vallée Poussin and Cauchy kernels (here $r = 1$). Thus, for instance, we can read off the fact (taking $\sigma = \beta_1$ and $\alpha = 1$) that for $f \in$ Lip 1 its Fejér-de la Vallée Poussin approximant $f^\lambda = f * K_\lambda$ approximates it with an error that is $O(\lambda^{-1}\log\lambda)$.

Similarly, taking $\sigma = \beta_2$ shows that if $\omega_2(f;\, t) = O(t^{\alpha})$ for some $\alpha > 1$, the approximation error is $O(\lambda^{-1})$. By refining the above proofs (to handle orders of magnitude other than a power of a in the hypotheses) one can easily get a sharper result here, involving a logarithm factor.[4]

THEOREM 36. If $f \in C(R^1)$, $K \in L^1(R^1)$ *and for some* $\alpha > 0$

$$\left| f(t) - \int_{-\infty}^{\infty} f(t-u)\,\lambda K(\lambda u)du \right| = O(\lambda^{-\alpha}), \qquad \lambda \to +\infty$$

uniformly in t, *the r-th modulus of smoothness of* f *is large-O of* a^{α}, $a^{\alpha}\log 1/a$, *or* a^{r} *as* $a \to 0$ *according as* α *is less than, equal to, or greater than* r.

Proof: Making the change of variable $v = \lambda u$, and writing $a = 1/\lambda$, the hypothesis states that the σ-deviation of f is $O(a^{\alpha})$ as $a \to 0$, where $d\sigma = d\delta - K du$. The result now follows from the corollary to Theorem 34.

As we saw in §5.1 it is important to be able to estimate the derivatives of the approximant f^{λ}. A useful theorem for this purpose is

THEOREM 37. *If* $f \in C$, $K \in L^1$ *and* K *is not identically zero and has an integrable n-th derivative, and for some* s $(0 \le s < n)$

$$\frac{d^n}{dt^n}\int_{-\infty}^{\infty} f(t-u)\,\lambda K(\lambda u)du = O(\lambda^{s}), \qquad \lambda \to +\infty \tag{24}$$

uniformly in t, *the r-th modulus of smoothness of* f *is large-O of*

a^{n-s}, $a^r \log 1/a$, *or* a^r *as* $a \to 0$, *according as* r *is greater than, equal to, or less than* $n-s$.

Proof: The left side of (5) equals

$$\frac{d^n}{dt^n}\int_{-\infty}^{\infty} f(u)\,\lambda K(\lambda t-\lambda u)du = \lambda^n\int_{-\infty}^{\infty} f(u)\,\lambda K^{(n)}(\lambda t-\lambda u)du$$

$$= a^{-n}\int_{-\infty}^{\infty} f(t-au)K^{(n)}(u)du,$$

$$\text{where } a = \frac{1}{\lambda}\ .$$

Hence, taking $d\sigma = K^{(n)}(u)du$, the σ-deviation of f is $O(a^{n-s})$, and the conclusion follows from the corollary to Theorem 34.

In the following paragraphs we give some applications of the preceding theorems.

§5.4.1 *Relation between moduli of smoothness of different orders.*

Let us take $\sigma = \beta_2$, $a = 1$, $r = 1$ in the corollary to Theorem 34. We get: $\omega_2(a) = O(a)$ implies $\omega_1(a) = O(a \log 1/a)$. Examples may be constructed to show this estimate is unimprovable. Similar results are obtainable for other moduli of smoothness, and it is easy to obtain a refinement of Theorem 34 which gives the Marchaud inequalities (see Lorentz [38], p. 48) in full generality.

§5.4.2 New derivations of Theorems 26 and 27.

Suppose that $f \in C_{2\pi}$, and there exists for each n some $T_n \in \mathcal{T}_{n-1}$ with $\|f-T_n\| \le An^{-\alpha}$. Let σ be any non-null real measure such that $\hat{\sigma}(x)$ vanishes for $|x| \le 1$. Writing

$$T_n(t) = \sum_{|k|<n} c_k e^{ikt}\ ,$$

we get

$$\int f(t-au)d\sigma(u) = \int [f(t-au) - T_n(t-au)]d\sigma(u)$$

$$+ \sum_{|k|<n} c_k e^{ikt}\hat{\sigma}(ka) \ .$$

Now, given $a > 0$, choose n so that $a^{-1} \leq n < 1 + a^{-1}$. Then the second term on the right vanishes, giving

$$\left| \int f(t-au)d\sigma(u) \right| \leq \|f - T_n\| \cdot \int |d\sigma| \leq Bn^{-\alpha} \leq Ba^{\alpha}$$

where $B = A \int |d\sigma|$. Therefore, $D_\sigma(f;\, a) = O(a^\alpha)$. If now $0 < \alpha < 1$ we take, in the corollary to Theorem 34, $r = 1$ and Theorem 26 follows. If $\alpha = 1$, we take $r = 2$ and Theorem 27 follows. Evidently we can deduce more general theorems of this kind in the same way. Moreover, a certain branch of "classical" approximation theory is concerned with approximation on R^1 to bounded continuous (not necessarily periodic) functions by entire functions of exponential type (the "type" being the parameter which corresponds to the degree of a trigonometric polynomial) bounded on the real axis. (For an exposition of this theory, see Timan [54].)

A very similar argument to the above serves to establish the corresponding inverse theorems also for this mode of approximation. We may also take this occasion to point out that the proofs we have given of "direct" theorems below also extend to this context.

§5.4.3 Applications to harmonic functions.

Certain theorems for harmonic functions in a half-plane, connecting the smoothness of a function on the boundary and the growth in the interior of its partial derivatives, are easily deducible from the preceding theorems. This reveals that the results in question are not essentially function-theoretic in character, which is instructive in view of "standard" proofs based on complex function theory.

The next two theorems are due to Hardy and Littlewood, and to Zygmund, respectively. Strictly speaking, their results were formulated for a circle rather than a half-plane (it is, of course, fairly easy to deduce "circle" theorems of the kind under discussion from the analogous "half-plane" theorems, and vice versa.) From our viewpoint in these lectures the half-plane is nicer to work with because there the Poisson kernel takes the pleasant form $\lambda K(\lambda x)$.

THEOREM 38. *The class of functions* u *which admit a representation*

$$u(x, y) = \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{f(t)\,dt}{(x-t)^2 + y^2}, \qquad y > 0 \tag{25}$$

with f *bounded and of class* Lip a $(0 < a \leq 1)$ *is identical with the class of functions* u *harmonic and bounded for* $y > 0$ *for which* $u_x(x, y) = O(y^{a-1})$.

THEOREM 39. *The class of functions* u *which admit a representation* (25) *where* f *is bounded and* $\omega_2(f;\, a) = O(a)$ *is identical with the class of functions* u *harmonic and bounded for* $y > 0$ *for which* $u_{xx}(x, y) = O(y^{-1})$.

(In the above theorems, subscripts on u denote partial differentiation.)

Proof of Theorem 38: Let us first recall that the function u defined by (25) is the same as $f^{\lambda}(x) = (f * K_{\lambda})(x)$, where $\lambda = y^{-1}$ and $K(x)$ is the Cauchy kernel $[\pi(1+x^2)]^{-1}$.

Suppose first u satisfies (25), with f bounded and in Lip α. Clearly u is bounded for $y > 0$, and it is trivial to check that it satisfies the Laplace equation. We have now,

$$u_x(x, y) = \frac{d}{dx} \int f(t)K_{\lambda}(x-t)dt = \lambda \int f(t)J_{\lambda}(x-t)dt, \quad \lambda = y^{-1}$$

where $J = K'$, $J_{\lambda}(x) = \lambda J(\lambda x)$. Thus, we have to prove the "direct" estimate $f * J_{\lambda} = O(\lambda^{-\alpha})$ under the hypothesis $f \in$ Lip α. This is of course rather trivial to do by a straightforward estimation. It is also instructive to see how it follows from the general theory: according to Theorem 32 we have only to verify that $\hat{J}$ is divisible at 0 by $\hat{\beta}_1$. Since $\hat{J}(x) = ix\hat{K}(x)$, $\hat{\beta}_1(x) = 1 - e^{-ix}$, this is evident.

Conversely (the "hard" part) suppose u is harmonic and bounded for $y > 0$ and $u_x(x, y) = O(y^{\alpha-1})$. From the Herglotz theorem together with the second corollary following Theorem 4 it is easily deduced that u is representable in the form (25) with f measurable and bounded.

Let us first assume f continuous. Then

$$\frac{d}{dx} \int f(t)K_{\lambda}(x-t)dt = u_x(x, y) = O(\lambda^{1-\alpha}) .$$

Applying Theorem 37 with $n = 1$, $s = 1-\alpha$ and $r = 1$ now gives $f \in$ Lip α, in case $\alpha < 1$. For $\alpha = 1$ however, we get the imperfect result $\omega(f; a) = O(a|\log a|)$. So, rather than use Theorem 37 let us go back to Theorem 32. We have $f * J_{\lambda} = O(\lambda^{-\alpha})$ (where $J = K'$), and since $\hat{J}$ divides $\hat{\beta}_1$ at 0, we conclude $\omega(f; a) = O(a^{\alpha})$. If f

is not assumed continuous, we apply a trick which is by now very familiar: convolve f with a suitable smoothing kernel P_n, so that $f^n \to f$ a.e., where f^n denotes $f * P_n$, and let $u^n(x, y) = \int f^n(t)K_\lambda(x-t)dt$. It is readily checked, since (as a function of x), $u^n = u * P_n$, that $|u_x^n(x, y)| \leq Ay^{a-1}$ with A independent of x, y and n and this leads to $\omega(f^n; a) \leq Ba^a$ with B independent of a and n. From this it follows that the f^n are uniformly equicontinuous, and $\{f^n\}$ converges *uniformly* to a limit g in Lip a, which must therefore coincide with f a.e. Therefore, after correction on a set of measure zero (which does not affect (25)) f is in Lip a. This proves Theorem 38.

Proof of Theorem 39. As this proof is very similar to the preceding, we shall be rather concise. On the one hand, if u admits the representation (25) with f bounded and $\omega_2(f; a) = O(a)$, u is a bounded harmonic function for $y > 0$, and

$$u_{xx}(x, y) = \frac{d^2}{dx^2} \int f(t)K_\lambda(x-t)dt = \lambda^2 \int f(t)I_\lambda(x-t)dt, \quad \lambda = y^{-1}$$

where $I = K''$. Thus, we must deduce $f * I_\lambda = O(\lambda^{-1})$ from $\omega_2(f; a) = O(a)$. This follows by Theorem 32, since $\hat{\beta}_2(x) = (1-e^{-ix})^2$ divides $\hat{I}(x) = (ix)^2\hat{K}(x)$ at $x = 0$.

In the other direction we proceed as before, the crucial point now being that $\hat{I}$ divides $\hat{\beta}_2$ at 0.

From Theorem 38 we can deduce a version of a well-known theorem due to Privalov:

THEOREM 40. *Let $g = u + iv$ be holomorphic for $y > 0$ and bounded for $y \geq 1$. Suppose u satisfies* (25) *with f real, bounded* *and in* Lip α $(0 < \alpha < 1)$. *Then* v *is bounded for* $y > 0$ *and has a continuous extension to* $y \geq 0$, *and* $v(x, 0)$ *is in* Lip α.

Proof: From (25) we get, differentiating

$$u_y(x, y) = \frac{1}{\pi} \int \frac{f(t)[(x-t)^2 - y^2]dt}{[(x-t)^2 + y^2]^2} \quad . \tag{26}$$

Since $u_y = -v_x$, we have from (26), setting $t = ys$

$$v_x(x, y) = \frac{1}{\pi y} \int_{-\infty}^{\infty} \frac{f(x-ys)(1-s^2)ds}{(1+s^2)^2} \quad ,$$

hence

$$v_x(x, y) = \frac{1}{\pi y} \int_{-\infty}^{\infty} [f(x-ys) - f(x)]\, J(s) ds$$

$$|v_x(s, y)| \leq A y^{\alpha-1} \int_{-\infty}^{\infty} s^{\alpha} |J(s)| ds = B y^{\alpha-1}$$

where $J(s) = (1-s^2)(1+s^2)^{-2}$, and A and B are independent of y (note that the last integral is finite because $\alpha < 1$). Now, by Theorem 38, the proof would be complete, if we know v is bounded for $y > 0$. By hypothesis, it is bounded for $y \geq 1$. Now, by Theorem 38 we have $|u_x| \leq C y^{\alpha-1}$, hence the same estimate applies to v_y, and integrating, we deduce the boundedness of v for $0 < y \leq 1$. This completes the proof.

Let us illustrate how to deduce the "circle" results. We confine ourselves to the following case of the Hardy-Littlewood theorem: *Let* $U(re^{i\theta})$ *be harmonic for* $r < 1$, *and* $|U_\theta| \leq A(1-r)^{\alpha-1}$ *where* $0 < \alpha \leq 1$. *Then* U *admits a representation*

(27) $$U(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(1-r^2)f(t)dt}{1-2r\cos(\theta-t)+r^2}$$

where f has period 2π and is in Lip α.

(See Zygmund [62_I], p. 263; (27) is the so-called Poisson integral for the circle. Here boundedness of U need not be supposed, as it will follow from the other hypotheses.)

For the proof, suppose first that U is bounded for $r < 1$, and let $u(w) = U(e^{iw})$, where $\operatorname{Im} w > 0$. Writing $w = x + iy$, we see that u is harmonic and bounded for $y > 0$. Moreover, $u_x = ie^{ix}U_\theta(e^{-y}e^{ix})$, hence $|u_x| \le A(1-e^{-y})^{\alpha-1} < Ay^{\alpha-1}$ for small y, and by Theorem 38, u satisfies (25) for some $f \in$ Lip α, which also has period 2π since u does. Therefore

$$U(e^{-y}e^{ix}) = \frac{y}{\pi}\int_{-\infty}^{\infty} \frac{f(t)dt}{(x-t)^2+y^2}, \quad y > 0$$

$$= \frac{y}{\pi}\sum_{n=-\infty}^{\infty}\int_0^{2\pi} \frac{f(x-t)dt}{(t+2\pi n)^2+y^2} = \frac{1}{2\pi}\int_0^{2\pi} f(x-t)G(t,y)dt$$

where

$$G(t,y) = 2y\sum_{n=-\infty}^{\infty}[(t+2\pi n)^2+y^2]^{-1} = \frac{1-e^{-2y}}{1-2e^{-y}\cos t+e^{-2y}},$$

the series converging uniformly (the latter identity is a simple consequence of Poisson's summation formula; see Titchmarsh [53], p. 60). Writing r for e^{-y} we get (27), with a slight change of notation.

If U is not assumed bounded, we proceed as follows. For each p, $0 < p < 1$ form the harmonic function $U^p(re^{i\theta}) = U(pre^{i\theta})$.

Then $|U^p_\theta| \leq A(1-pr)^{a-1} < A(1-r)^{a-1}$. Applying the above result to U^p we see (since the $f(t)$ in (27) is then $U(pe^{i\theta})$) that the restrictions of U to the circles $r = p$ are *uniformly* in Lip a. If, therefore, $U(z_n)$ is bounded for any sequence z_n with $|z_n| \to 1$, U is bounded for $r < 1$. But this is certainly the case, otherwise U tends uniformly to either plus or minus infinity as $r \to 1$. This is impossible, for if $U \to -\infty$ (say) and V is the conjugate harmonic function, the non-vanishing analytic function $f = \exp(U + iV)$ should tend uniformly to zero as $r \to 1$.

REMARKS. From the foregoing proof we see that the harmonic (or holomorphic) functions in the unit disk are naturally embedded in the corresponding class in the upper half-plane, as those functions which have period 2π. This embedding (via the map $z = e^{iw}$) is sometimes more useful than that based on conformal transformation, since the circles $|z| = \text{const.}$ correspond in the former case to the lines Im w = const., but this is not so in the latter case.

Finally, we remark that "little o" versions of the preceding theorems are also true (the Lip classes being also replaced by the corresponding "little o" Lip classes).

§5.5 *Further remarks on saturation.*

Let us now, as we promised, complete the discussion of saturation for the Fejér-de la Vallée Poussin kernel. This will serve to acquaint us with a far reaching extension of the "Fourier transform technique" we have used thus far, and give new insight into a number of other phenomena as well. Let us start by finding the promised alternate proof of Theorem 14 (by a rather meandering path). We recall the hypotheses: f is of class C^2, f and f'' are bounded, and

(28) $$J(f;\, t) = -\frac{1}{\pi}\int_0^{\infty} \frac{f(t+u)-2f(t)+f(t-u)}{u^2}\, du$$

vanishes for all t; we wish to show f is constant. (The hypothesis of boundedness of f'' is really inessential, and we shall later eliminate it.)

What sort of expression is $J(f;\, t)$? The key insight is that it is a kind of *convolution*, albeit not of the usual type (e.g., convolution with an L^1 function, or with a finite measure) we have encountered thus far. Let us consider the functional α defined by

$$\alpha f = -\frac{1}{\pi}\int_0^{\infty} \frac{f(u)-2f(0)+f(-u)}{u^2}\, du$$

α is not defined for all $f \in C$, but it is defined, the integral converging absolutely, for all bounded $f \in C^2$. Then $J(f;\, t)$ is just α applied to the function $f(t-u)$ (considered as a function of u). Now, if α were a functional determined by an L^1 function, say $\alpha f = \int fK$, α applied to $f(t-u)$ would be just the convolution $K * f$. Let us, therefore, proceeding heuristically for the time being, write $J(f;\, t)$ in the suggestive form of a "convolution" $\alpha * f$. Now, let us be even more daring and compute the Fourier transform of α! (Of course, to one familiar with the Schwartz theory of distributions, there is nothing "daring" in this at all,[5] we are just doing routine maneuvers, but we do not wish to presuppose a knowledge of distribution theory here.) The Fourier transform $\hat{\alpha}$ could certainly be nothing other than

$$\hat{\alpha}(x) = \alpha_u e^{-ixu} = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{ixu}-2+e^{-ixu}}{u^2}\, du$$

$$= \frac{2}{\pi} \int_0^\infty \left(\frac{\sin xv}{v}\right)^2 dv = |x| \ .$$

Very interesting: $\hat{\alpha}(x)$ has the same behavior near $x = 0$, as $1-\hat{K}(x)$ (in fact, it is identical with it for $|x| \leq 1$) where $K(t)$ is the Fejér-de la Vallée Poussin kernel. This already suggests why the expression $J(f;\, t)$ should be intimately related to the approximation phenomena of this particular kernel.

But, how can this help us to prove Theorem 14? The idea is, if $\alpha * f = 0$, then formally $\beta * f$ should be zero for lots of functionals β, namely those for which $\hat{\alpha}$ divides $\hat{\beta}$. But, "divides" in what sense? A well-motivated guess would be: if $\hat{\beta}$ divided by $\hat{\alpha}$ *belongs to W*, then $\beta * f = 0$. Then, with so much freedom to choose β, we should be able to find more tractable functionals than α, hopefully even nice ones generated by L^1 functions. Thus, our strategy will be to first deduce that the convolution of f with each of a large family of integrable functions vanishes.

Now we are ready to begin a formal proof of Theorem 14. Of course, this *could* be presented without the preceding discussion, however it would be very hard to discover the steps which follow without having as a guide the heuristic argument we just gave. We need first a lemma.

LEMMA. *Let $f \in C^2$, and suppose f and f'' are bounded on $(-\infty, \infty)$, and define $J(f;\, t)$ by* (28). *Then $J(f; t)$ is bounded. Moreover, let $h \in L^1 \cap L^2$ and suppose further* (i) *$x\hat{h} \in L^2$, and* (ii) *$|x|\,\hat{h}(x)$ is the Fourier transform of some $g \in L^1$. Then*

$$\int_{-\infty}^{\infty} J(f;\, t)h(t)dt = \int_{-\infty}^{\infty} f(t)\, g(t)dt \ .$$

Proof: First of all, writing

$$\int_0^\infty \left|\frac{f(t+u) - 2f(t) + f(t-u)}{u^2}\right| du = \int_0^1 + \int_1^\infty$$

we see that the left-hand side is bounded by $\sup|f''(t)| + 4\sup|f(t)|$. Let now $b > 0$. The double integral

$$\int_{-\infty}^\infty \left(\int_b^\infty \frac{f(t+u) - 2f(t) + f(t-u)}{u^2} \, du \right) h(t)dt \tag{29}$$

converges absolutely, and therefore by Fubini's theorem equals

$$\int_b^\infty u^{-2} \left(\int_{-\infty}^\infty (f(t+u) - 2f(t) + f(t-u))\, h(t)dt \right) du$$

$$= \int_b^\infty u^{-2} \left(\int_{-\infty}^\infty (h(s+u) - 2h(s) + h(s-u))\, f(s)ds \right) du$$

$$= \int_{-\infty}^\infty f(s) \left(\int_b^\infty \frac{h(s+u) - 2h(s) + h(s-u)}{u^2} \, du \right) ds \; .$$

Now, writing H for $\hat{h}$, we have the inversion formula

$$h(s) = \frac{1}{2\pi} \int_{-\infty}^\infty H(x)\, e^{ixs}\, dx$$

which is valid since $H \in L^1$ (in turn a consequence of both H and xH being in L^2). Using this we get, for the inner (bracketed) integral in the last expression

$$\int_b^\infty u^{-2} \left(\frac{1}{2\pi} \int_{-\infty}^\infty H(x)(e^{ixu} - 2 + e^{-ixu})\, e^{ixs}\, dx \right) du$$

$$= \frac{1}{2\pi} \int_{-\infty}^\infty H(x)\, e^{ixs} \left(\int_b^\infty u^{-2}(e^{ixu} - 2 + e^{-ixu}) du \right) dx$$

$$= -\frac{1}{\pi} \int_{-\infty}^{\infty} H(x)e^{ixs}\left(\int_{b/2}^{\infty} \frac{\sin^2 tx}{t^2}\, dt\right) dx$$

$$= -\frac{1}{\pi} \int_{-\infty}^{\infty} |x|\, H(x)e^{ixs} p\left(\frac{bx}{2}\right) dx,$$

where

(30) $$p(y) = \int_{y}^{\infty} \frac{\sin^2 u}{u^2}\, du, \qquad p(-y) = p(y) \quad (y \geq 0)\,.$$

To sum up: we have shown that the expression (29) equals

(31) $$-\frac{1}{\pi} \int_{-\infty}^{\infty} f(s)\left[\int_{-\infty}^{\infty} |x|\, H(x)\, p\left(\frac{bx}{2}\right) e^{ixs}\, dx\right] ds$$

where p is defined by (30). Now, let $b \to 0$. Since the bracketed expression in (29) remains bounded, the dominated convergence theorem applies and (29) tends to $-\pi \int_{\infty}^{\infty} J(f;\, t)\, h(t) dt$.

The limiting behavior of (31) as $b \to 0$ is more delicate. Note first that $p(y) = O(|y|^{-1})$ at infinity, and so $p \in L^2$. Since p is even, and decreasing for $y > 0$, its L^2 transform P may be computed from the formula

$$P(x) = 2 \lim_{u \to \infty} \int_0^u p(y) \cos xy\, dy = \frac{2}{x} \int_0^{\infty} \frac{\sin^2 y \sin xy}{y^2}\, dy\,.$$

This last integral, being the Fourier transform of an L^2 function, is in L^2 (in fact, it is even $O(|x|^{-1})$), *therefore P is absolutely integrable.*

Now let us apply the Parseval formula to the bracketed expression in (31). We get (recalling that $|x| H(x) = \hat{g}(x)$ is in L^2)

$$(32)\qquad \int_{-\infty}^{\infty} |x|\, H(x) e^{isx} \cdot p\left(\frac{bx}{2}\right) dx = \int_{-\infty}^{\infty} g(t-s)\,\lambda\, P(\lambda t)\, dt$$

where $\lambda = 2b^{-1}$. Since $P \epsilon L^1$, and $\int P\, dt = \pi^2$, we see from Theorem 2 that the left side of (32), as a function of s, converges in L^1 norm to $\pi^2 g(s)$ as $b \to 0$. Therefore, the limit as $b \to 0$ of the expression (31) is $-\pi \int_{-\infty}^{\infty} f(s)\, g(s)$. This completes the proof of the Lemma.

REMARK. The hypothesis that f'' is bounded may be eliminated from the Lemma, *providing h is of compact support.* In fact, the only place where the boundedness of f'' was used was in showing that the expression (29) tends, as $b \to 0$, to the expression obtained by setting $b = 0$ in (29). If now h vanishes outside the interval $[t_1, t_2]$, it is sufficient, in order to justify the passage to the limit, to verify that the bracketed expression in (29) remains less than a fixed limit as $b \to 0$, uniformly for $t_1 \leq t \leq t_2$. Now,

$$\int_0^1 \left|\frac{f(t+u) - 2f(t) + f(t-u)}{u^2}\right| du \leq \text{maximum of } |f''(t)| \text{ on } [t_1 - 1, t_2 + 1] \qquad (t_1 \leq t \leq t_2)$$

and $\int_1^{\infty} \leq 4 \sup |f|$, as before.

Proof of Theorem 14. We suppose $f \epsilon C^2$ and bounded, and $J(f; t)$ vanishes. Let $h(t) = 1 - |t|$ on $[-1, 1]$ and 0 elsewhere, so $H(x) = \sin^2 x/x^2$. This h satisfies the hypotheses of the Lemma, and is of compact support. (As to (ii), observe that $|x|\, H(x)$ is absolutely continuous and has an L^2 derivative.) Let $|x|\, H(x) = \hat{g}(x)$, $g \epsilon L^1$. For every $\lambda > 0$ and y real the function $H(x/\lambda) e^{iyx}$ is the Fourier transform of $h_{\lambda, y}(t) = \lambda h(\lambda y - \lambda t)$ which, just as h, satisfies (i) and (ii) and is of compact support. Now, $|x|H(x/\lambda)e^{iyx}$ is the Fourier transform of $\lambda^2 g(\lambda y-\lambda t)$. Applying the Lemma (with $h_{\lambda,y}$ in place of h) we get $\int_{-\infty}^{\infty} f(t)\lambda g(\lambda y-\lambda t)dt = 0$, i.e., $f*g_\lambda$ for all $\lambda>0$. Now, $\hat{g}$ divides $\hat{\beta}_2$ at $x=0$, therefore by Theorem 32, $D_{\beta_2}(f;a)$ vanishes for all $a>0$. Hence f is a linear polynomial and, being bounded, reduces to a constant. This completes the proof.

Exercise. If $J(f;t) = 0$ for all $t \in [t_1, t_2]$, f is linear on $[t_1, t_2]$.

It is now a simple matter to fill the gap in Chapter III and "identify" (i.e., give an alternate description of) the saturation class for the Fejér-de la Vallée Poussin kernel K. We shall prove: *If f is the uniform limit of a sequence of functions $\{f_n\}$ such that $J(f_n;t)$ remains uniformly bounded, then* $\|f-f*K_\lambda\| = O(\lambda^{-1})$.

Indeed, applying the Lemma to f_n and $h_{\lambda,y}$ we get

$$\int f_n(t)\,\lambda^2 g(\lambda y-\lambda t)dt = \int J(f_n;t)\,h_{\lambda,y}(t)\,dt\ .$$

The right side does not exceed $A\|h_{\lambda,y}\|_1 = A\|h\|_1$, with A independent of n, therefore $\|f_n*g_\lambda\| \le A\lambda^{-1}$. Letting $n\to\infty$ we get $\|f*g_\lambda\| \le A\lambda^{-1}$. Finally, since $\hat{g}(x) = |x|^{-1}\sin^2 x$ divides $1-\hat{K}(x)$ at 0, we get $\|f-f*K_\lambda\| \le A_1\lambda^{-1}$, proving the assertion.

Exercise. The saturation class for the Fejér-de la Vallée Poussin kernel neither includes, nor is included in Lip 1.

Another way of describing the saturation class is as follows. We can write

$$J(f;t) = -\frac{1}{\pi a}\int_0^\infty \frac{f(t+au)-2f(t)+f(t-au)}{u^2}\,du$$

and now the integral on the right, in terms of the above defined functional α, may be written $\alpha_u f(t - au)$ (the subscript indicates we consider $f(t-au)$ as a function of u). If, therefore, we define, for all bounded C^2 functions f the *α-deviation* by

$$D_\alpha(f;\, a) = \sup_t |\alpha_u f(t-au)|$$

we can state: *the saturation class of the Fejér-de la Vallée Poussin kernel is identical with the class of uniform limits of sequences* $\{f_n\}$ *where* $f_n \in C^2 \cap L^\infty$ *and* $D_\alpha(f_n;\, a) \leq Aa$ *(A independent of n and a).*

This formulation (we may concisely describe the saturation class as the bounded "weak" solutions of $D_\alpha(f;\, a) = O(a)$) is suggestive, and poses the question to what extent the general theory developed in this chapter may be extended to β-deviations, where β denotes a linear functional not representable by a finite measure. We shall not attempt, however, to take up this question here, but shall consider another example in §5.6

Yet another way of describing the saturation class of K is in terms of *conjugate functions;* this is the form in which Zamansky [60], who discovered the result (in the case of periodic f) described it. The *conjugate function,* or Hilbert transform, may be defined with various degrees of generality. For our purposes we make the following definition (Ahiezer [2], p. 171). Let f be measurable on $(-\infty, \infty)$ and $(1+|t|)^{-1}f(t) \in L^2$. Let T denote the Fourier-Plancherel isometry on L^2, and let $g = T(t-i)^{-1}f$, $h(u) = i \operatorname{sgn} u\, g(u)$, and finally $\tilde{f}(t) = (t-i)T^{-1}h$. This $\tilde{f}$ is called the *conjugate of f*, and is readily seen to satisfy

$$\int_{-\infty}^{\infty} \frac{|\tilde{f}(t)|^2}{1+t^2}\, dt = \int_{-\infty}^{\infty} \frac{|f(t)|}{1+t^2}\, dt \quad .$$

In particular, each bounded function has a conjugate determined a.e. If f has period 2π, this conjugate coincides with the usual Fourier series definition, and if $f \in L^2(-\infty, \infty)$, $\tilde{f}$ is the Hilbert transform.

Exercise. For $f \in C^2$, $\tilde{f}$ coincides a.e. with an absolutely continuous function whose derivative is a constant times $J(f; t)$.

THEOREM 41. *Let $K(t)$ be either of the kernels $e^{-|t|}$ or $(1/\pi)(\sin^2 t/t^2)$, and f bounded. Then the following two assertions are equivalent:*

(i) $\|f - f * K_\lambda\| = O(\lambda^{-1})$

(ii) *The conjugate function of f coincides a.e. with some function in* Lip 1.

We omit the proof of this theorem, which follows easily once the Exercise is established. Finally, for other recent literature on saturation, we refer the reader to [67, 68, 69, 71, 72].

§5.6 *A final remark concerning Jackson's Theorem.*

The consideration, as in §5.5, of more general functionals than those defined by finite measures, is fruitful also in connection with Jackson's theorem, when the information postulated about f is in terms of derivatives, rather than differences. Suppose, for instance, for some positive integer r, $f \in C^r$ and f and its derivatives up to order r are bounded, in particular $|f^{(r)}(t)| \leq 1$. Introducing the symbol δ_r for the functional $f^{(r)}(0)$, the δ_r-deviation is

$$D_{\delta_r}(f; a) = \sup_t |\delta_r f(t - au)| = a^r \sup |f^{(r)}| .$$

The Fourier transform of δ_r is $(-ix)^r$. If, therefore, we choose a kernel K such that x^r divides $\hat{K}$ in W, we expect, by analogy with

Theorem 30, that $f * K_\lambda$ will approximate f with an error $O(\lambda^{-r})$. This is, in essence, just what we found in Theorem 20, only there the conditions on K were in the form that certain moments vanished, i.e., the derivatives of $\hat{K}$ of order up to and including r vanish at 0. The Fourier argument we have just outlined can be carried out rigorously. In addition to having a very clear motivation, it is also very convenient to work with when proving various versions of "Jackson's theorem" for functions of several variables. It is noteworthy that $\hat{\delta}_r(x)$ vanishes only at 0 and very cooperatively even tends to infinity for large x. This allows a fairly simple argument (based on *global* divisibility) to work, whereas in the "modulus of smoothness" variant of Jackson's theorem we encounter the Fourier transform $\hat{\beta}_r(x) = (1-e^{-ix})^r$ which has "extraneous" zeroes at $x = 2\pi n$ $(n = \pm 1, \pm 2, \ldots)$, and so necessitates the use of the deeper Theorem 32, based on *local* divisibility.

Exercise. Construct a proof of Jackson's theorem along the lines of the argument sketched above. (A thorough discussion of Jackson's theorem and its extensions, from the point of view of this paragraph may be found in the author's paper [74].)

Chapter V: Footnotes added in proof

Page 68. [1] The Fourier transform of g (in the distribution sense) has its support in $[-A, A]$ —that is one way to formulate the relevant property of g.

Page 77. [2] Of course, this is true only insofar as approximations of the type $f * K_2$ are the object of discussion.

Page 89. [3] In applying Theorem 32, we must have a criterion for (local) membership in W. The following simple criterion suffices for all applications in this book: If $S(x)$ is of class Lip 1 in a neighborhood U of 0, there is a function in W which coincides with S in U.

Page 94. [4] It is of interest that when the analogous theory is carried out for L^p norms with $1 < p < \infty$ these logarithmic factors disappear. This is because the role of the class W is taken over by the larger class M_p of "Fourier multipliers," and $(1-\hat{K})/\hat{\beta}_1$ and its reciprocal are (locally) of class M_p. In particular, the saturation class of the Fejér-de la V.P. kernel is precisely Lip $(1, p)$ (see [65]).

Page 103. [5] α is a "tempered distribution," and as such possesses a Fourier transform in the L. Schwartz sense, namely $\hat{\alpha}(x) = |x|$.

PREFACE. Earlier work in the spirit of the present study may be found in de la Vallée Poussin [20], sections of Bochner's 1937 planographed lecture notes from Princeton on "Harmonic Analysis," and Ahieser [1] (the original edition of which came out in 1940). In [1, 41, 54] one may find discussion of many relevant papers of S. Bernstein. Among Butzer's many papers, we mention especially [10, 11]. Also, much of Timan's treatise [54] and M. Golomb's notes [26] are germane to our study.

CHAPTER I 1.1. There are now a number of books available covering various aspects of approximation theory, and we shall not venture to list those not bearing strongly on the topics treated in these lectures. Lorentz [38] and Timan [54] are a concise, and an encyclopedic book respectively with broad coverage.

1.2. The averaging trick employed here seems difficult to trace to any one source; it is found in many books (e.g., Littlewood [37]), and Timan [54], p. 177, refers to work of V. A. Steklov from 1916.

The general insight that approximations are generated by convolving with a peaking kernel is probably due to Weierstrass, whose proof [56] of his celebrated approximation theorem we have reproduced here. He may well have been motivated by mathematical physics, notably the solution as a convolution integral of the initial value problem for the heat equation. Weierstrass also employed a convolution integral with a Jacobi theta function as kernel in order to prove the "trigonometric" version of his approximation theorem. Landau [36] gave a more direct proof of Weierstrass' theorem by employing an *ad hoc* polynomial kernel. See also Graves [27], where further references are given.

1.4. N. Wiener seems first to have appreciated the vast range and unifying power of the convolution concept (see [59], and his book [48]). Curiously, Wiener did little work in approximation theory, although he forged many valuable tools for its study.

CHAPTER II. 2.1. Remark 1 following Theorem 3, Corollary 1: The theorem on entire approximation of continuous functions is due to Carleman [18]. See also Kaplan [34]. A considerable literature has arisen extending Carleman's result; for a definitive generalization, see Araklian [2]. Approximation by entire functions of exponential type was pioneered by S. Bernstein; see Ahiezer [1], Timan [54], Golomb [26].

2.2. Theorem 4 was suggested by Theorem 6 of Bochner and Chandrasekharan [7]. So far as we are aware, necessary and sufficient conditions on a kernel K for $f*K_\lambda$ to converge almost everywhere (when $f \in L^\infty$, say) are not known. See also Butzer [10], §2 where error bounds are given for local approximation.

Corollary 2. The simpler analytic dependence of the Poisson kernel for the half plane on its "parameter" than that of the circle may serve to illuminate our remark in the second paragraph of the Preface.

CHAPTER III. 3.1. The notion of saturation was introduced by Favard [22, 23] and for that reason the saturation class is sometimes called "Favard class" in the literature. Important pioneering studies were made by Zamansky [60, 61] in connection with summability methods applied to a Fourier series. Despite a large literature (see for instance Butzer [9], Sunouchi [50], Butzer and Görlich [16], and further references there) it seems to us there is work to be done here, not only to find the saturation classes of subtle operators, but to clarify the very notion of saturation order and class from a really general point of view, in the framework of functional analysis.

3.2. Theorems 6- 10 are the obvious fruit of applying a very familiar elementary technique, and known, in principle at least, to many investigators (compare e.g., Lorentz [39]). Semigroup arguments sometimes employed to establish special cases of these theorems, although elegant, seem not the most appropriate tool here.

3.3. The treatment here (continued in §5.5) may have some points of novelty. For a closely related investigation bearing on periodic functions see Zamansky [60, 61]. The non-periodic case seems rather harder.

3.5. Theorem 16 is from Titchmarsh [51], Theorem 17 from Hardy and Littlewood [28]. For other proofs see the original papers, and Butzer [12].

The "mollifier" technique which we employ repeatedly is especially widespread in the study of partial differential equations. Among many papers we mention Friedrichs [24].[1] The alternate technique of "dualizing" a condition in terms of smooth "test functions" is also widespread in this field (compare "Weyl's lemma," Exercise 5) and underlies various notions of "generalized functions" or "distributions" (L. Schwartz [46]). Recently, mollifiers and distributions have been playing an increasing role in general analysis, including approximation theory. As samples we from a vast field we may cite Mergelyan's theorem (see the proof in Collingwood and Lohwater [19]), and the study of "mean-periodic functions," e.g., Kahane [33], p. 26. There are also good grounds for formulating and proving Theorems 11–14 in the language of distributions (see §5.5).

3.5.1. Exercises 8, 9 are results of Hardy and Littlewood [29]. See also Ilyin [32]. In content (although not in intent) these are forerunners of a class of inequalities (usually for functions of *several* variables) known in the Soviet literature by the general name of "embedding theorems" (see Kantorovich and Akilov [35], Chapter X), relating various norms of partial derivatives and differences of functions, and their restrictions to subdomains. Such inequalities are important in establishing the smoothness of "weak" solutions of partial differential equations, as well as that of the minimizing functions (in a "weak" sense) of certain variational problems (Morrey [40]). In the Soviet literature on "embedding theorems" especially, methods from approximation theory are often used, and the tie-in with the material in §3.5 is clear (on this point, see especially Ilyin [32], Nikolski [41a], Kudryavtsev [35a]).[2] Of course, these studies are germane chiefly to approximation theory in several variables, and some of the results (e.g., concerning the restrictions of functions in certain classes to lower dimensional varieties) have no interesting counterpart in one dimension.

CHAPTER IV. 4.1. A very thorough and readable guide to "classical" material on approximation theory throughout the next two chapters is Natanson [41]. For Fourier transform theory, besides sources already cited, Goldberg [45] is concise and clear; see also Bochner [5], Titchmarsh [53], Zygmund [62], Vol. II (Chapter XVI). There are also several treatises dealing with Fourier analysis on groups, where one can, with a little effort, dig out the "classical" versions of the theorems needed here.[3]

Theorem 19, although fairly obvious, seems not to have been used in this context elsewhere. There is considerable convenience in not having to exhibit a kernel explicitly, especially since (as we shall learn later in these lectures) the approximation properties are read off from the *Fourier transform* of the kernel, anyway. Theorem 19 can be generalized to several variables and, together with the "divisibility" method (see Chapter V, and especially 5.5), gives a streamlined formalism for proving analogues of Jackson's theorem in higher dimensions [74].

4.3. The inequalities estimating ω_r in terms of ω_s when $r < s$ are due to Marchaud (see Timan [54], Boman [81]). Boman also gives far reaching extensions of these inequalities involving mappings between Euclidean spaces, employing Fourier transform techniques not unlike those we shall use in Chapter V.

4.4. The idea of performing (single or iterated) integration by convolving with a suitable kernel is very well known, and underlies various definitions of fractional integration (see Zygmund [62], Vol. II, p. 133 for the case of periodic functions). In the periodic case the only condition required is that the function have mean value zero. In more general situations (e.g., the infinite line) the situation is more subtle and it must be required that the "spectrum" (in some sense) of the function omit a neighborhood of zero.

This situation is closely related to an inequality of Bohr and Favard for almost periodic functions (references in Hörmander [31], where the inequality is substantially generalized using concepts and methods of distribution theory and generalized harmonic analysis).

4.4.2. At a recent conference in Oberwolfach, Jean-Pierre Kahane, in response to a question of the author, gave a very simple demonstration of the following proposition:

For every $s \geq 1$, *there exists a function* $p_s \in L^1$ *such that* $\hat{p}_s(x) = |x|^{-s}$ *for* $|x| \geq 1$ *and* $\|p_s\|_1 \leq A$, *where* A *is an absolute constant.*

With the permission of Professor Kahane, we reproduce here his argument. Define a function $P = P_s$ as follows:

$$P(x) = \begin{cases} |x|^{-s}, & |x| \geq 1 \\ (2-|x|)^{-s}, & |x| \leq 1. \end{cases}$$

The point to note is that, on $[0, 2]$, P is symmetric about $x = 1$ (and similarly on $[-2, 0]$ about $x = -1$). We claim that this $P = \hat{p}_s$ fulfills the requirements. Indeed, we have $P = Q + R_1 + R_2$, where

$$Q(x) = \begin{cases} 2^{-s}, & |x| \leq 2 \\ x^{-s}, & |x| \geq 2 \end{cases}$$

$$R(x) = \begin{cases} (1+|x|)^{-s} - 2^{-s}, & |x| \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

and $R_1(x) = R(x-1)$, $R_2(x) = R(x+1)$. It is easy to check that Q, R_1, R_2 are L^1 Fourier transforms, and we need only show that their W-norms are bounded by constants independent of s. Now, R_1 and R_2 are translates of R which is even and convex; for such a function, as we remarked in the proof of Theorem 22, the W-norm equals the sup norm. Hence $\|R_1\|_W = \|R_2\|_W = 1-2^{-s}$. For Q we may use the estimate

$$\|Q\|_W^2 \leq \frac{1}{2} \int_{-\infty}^{\infty} (Q(x)^2 + Q'(x)^2)\, dx$$

$$= \frac{1}{2} \left[4 \cdot 2^{-2s} + \frac{2^{-2s+2}}{2s-1} + \frac{s^2\, 2^{-2s}}{2s+1} \right]$$

Clearly, this is bounded by a constant independent of s for $s \geq 1$, and these estimates, together with $\|P\|_W \leq \|Q\|_W + \|R_1\|_W + \|R_2\|_W$, prove the assertion. It seems that a similar (but slightly more complicated) procedure would establish the analogous result for the function defined as x^{-s} for $|x| \geq 1$, where s is an *odd integer*.

4.4.3. Kolmogorov has introduced several important general concepts into the study of the approximation of classes of functions (especially the determination of lower bounds), notably "entropy" and "widths" (see Lorentz [38], Chapters 9, 10).

4.5. Theorem 24, although obvious, does not seem to be formulated elsewhere. Combined with Theorem 19, it gives an alternative approach to a problem studied by Bochner [6] since a non-null integrable function K, with $K(t) \geq 0$ can be found which falls off "nearly exponentially" at infinity,[4] and whose Fourier transform vanishes for $|x| \geq 1$. Such a kernel K generates trigonometric polynomial approximations to a given periodic function f which give an "almost exponentially small" error within each interval of constancy of f, and moreover satisfy the remaining requirements of Bochner's Theorem 1.

CHAPTER V. 5.2, 5.3. For the approach here see the author's papers [48, 49]. In [49] extensions of some of the results to several variables and L^p norms are given. Once again, division in W of certain Fourier transforms plays a decisive role. This role is in essence the same as that played by Fourier-Stieltjes transforms as "multipliers" on various function classes, familiar in Fourier analysis. However, our L^p versions with $1 < p < \infty$ do not generally give sharp results. For instance, Zygmund [63], using rather deep Fourier methods, proved for periodic functions f that $\|f(x+h) - 2f(x) + f(x-h)\|_p = O(h)$ implies

$$\|f(x+h)-f(x)\|_p \text{ in large 0 of } \begin{cases} h|\log h|^{1/p}, & 1 \leq p \leq 2 \\ h|\log h|^{1/2}, & 2 \leq p < \infty \\ h|\log h|, & p = \infty \end{cases}$$

and (surprisingly) these results are sharp. (The case $p = 2$ had been done also by A. F. Timan and M. F. Timan (see [54], p. 121).) Our general theory gives $h\,|\log h|$ in all cases, i.e., an unsharp result for $1 < p < \infty$. It would be of obvious interest to derive Zygmund's results within the framework of a general theory, along the lines developed here.† Another interesting area for study is how far *pointwise* (rather than norm) approximation theorems can be inferred from the Fourier transform of the kernel.

5.2.2. The converse theorem referred to is a consequence of the following proposition (a variant of which seems first to have been formulated by de Leeuw and Mirkil [70].

Let $\sigma, \sigma_1, \ldots, \sigma_n$ be finite measures on the line, and suppose there exist constants $A_1, \ldots, A_n$ such that for all continuous f

† Afterword: this has since been done; see [65].

vanishing at infinity

(1) $$\left| \int f\, d\sigma \right| \le \sum_{1}^{n} A_i \|f * \sigma_i\|_\infty .$$

Then, σ belongs to the ideal in W generated by the σ_i.

(Here, $f * \sigma_i$ denotes $\int f(t+u)\, d\sigma_i(u)$. We may remark that it would be enough to assume that (1) holds for $f \in C^\infty$ of compact support.)

Outline of proof. Let $C_{0,n}$ denote the Banach space of all n-tuples $F = (f_1, \ldots, f_n)$ of functions of class C_0 (i.e., continuous, vanishing at infinity) normed by $F = \max_i \|f_i\|_\infty$. The most general bounded linear functional on $C_{0,n}$ clearly has the form

$$F \to \sum_{i=1}^{n} \int f_i\, d\rho_i$$

where the ρ_i are finite measures on the line. Now, denote the range of the map $T\colon f \to (f * \sigma_1, \ldots, f * \sigma_n)$, as f varies over C_0, by S. Note that, because of (1), $Tf = Tg$ implies $\int f\, d\sigma = \int g\, d\sigma$ hence the map $\phi\colon F \to \int (T^{-1} F) d\sigma$ from S to the reals is single-valued (even though T^{-1} needn't be). Moreover ϕ is additive, and bounded in view of (1). Therefore, ϕ is extendible to a bounded linear functional on $C_{0,n}$, hence there exist finite measures ρ_i such that

$$\phi F = \sum_{1}^{n} \int f_i\, d\rho_i \quad \text{for } F = (f_1, \ldots, f_n) .$$

In particular, taking $F = Tf \in S$ we get

(2) $$\int f\,d\sigma = \sum_{i=1}^{n} \int\left(\int f(t+u)\,d\sigma_i(u)\right) d\rho_i(t) \ .$$

Now, (2) is valid for all $f \epsilon C_0$. Since every $f \epsilon C$ (bounded continuous functions) is the pointwise limit of a uniformly bounded sequence of functions in C_0, (2) holds for all $f \epsilon C$, by a passage to the limit. Now, take $f(t) = e^{ixt}$ in (2), and we get

$$\hat{\sigma}(x) = \sum_{i=1}^{n} \hat{\rho}_i(x)\hat{\sigma}_i(x) \ ,$$

and the proposition is proved.

5.4. Theorem 37 was motivated by the technique used in §5.1 in proving inverse theorems, based on estimates of the derivatives of the approximating functions. Aside from a few scattered results in the literature, a systematic study of the derivatives of approximating functions (in the context of trigonometric polynomial approximation to periodic functions) appeared only quite recently (Butzer and Pawelke [17]).

5.4.3. For Theorems 38, 39 (in the circle) see Zygmund [62], Vol. I, p. 263; for Theorem 40 and a generalization of it (likewise obtainable by our methods) *ibid.*, p. 121.

5.5, 5.6. The "β deviations", i.e., their comparison in terms of divisibility of Fourier transforms, for functionals β of various classes, and also the notion of saturation classes of functionals, seem to us interesting topics for future investigation. More specifically: if β denotes a linear functional on some class F of functions, and $f(t-au)$ (as a function u) is in F for all t real, $a > 0$ whenever f is, the β-*deviation* (relative to a suitable norm) is

$\|\beta_u f(t-au)\|$. Of course, β should be such that a Fourier transform can be reasonably defined for it; this is certainly the case if F contains the exponential functions e^{-ixt}, but there are other possibilities too (e.g., F is the class of C^∞ functions with compact support, and β a so-called "tempered distribution"). We have been able to prove a version† of Theorem 30 where σ_1 and σ_2 are functionals on C^n. On the other hand, the obvious generalization of Theorem 32 is not true, as the following example shows. Let a denote the linear functional on C^1 (with the usual Banach norm) defined by $af = f(2) - 2f(0) + f(-2)$, and let $\beta f = f'(1) - f'(-1)$. Then $\hat{a}(x) = (e^{ix} - e^{-ix})^2$, $\hat{\beta}(x) = ix(e^{ix} - e^{-ix})$. Moreover,

$$a_u f(t-au) = f(t+2a) - 2f(t) + f(t-2a)$$

$$\beta_u f(t-au) = a[f'(t+a) - f'(t-a)] .$$

Here, $\hat{a}$ divides $\hat{\beta}$ at 0 (in the ring W). Suppose now $D_a(f; a) \leq Aa$. If Theorem 32 were applicable in this situation, we should have $D_\beta(f; a) \leq BAa$, with B independent of a and A, and this inequality is untrue in general, since it would imply for arbitrary $f \in C$ (by our standard argument) that $\omega_2(f; a) = O(a)$ implies $f \in$ Lip 1.

More on Theorem 14: Even after our lengthy discussion, there remains much to say†† about Theorem 14. Purely *formally,* it is an analogue of the following slight variant of Wiener's Tauberian theorem:

† See §2 of our paper [74].

†† Afterword: If I had it to do over again I would avoid the *tour de force* of Fourier transform theory in §5.4. The same results could be obtained much more easily, with the aid of the L. Schwartz theory, or else using the material in §2 of [74].

If $K * f = 0$, $K \in L^1$, $f \in L^\infty$ *and* $\hat{K}(x) \neq 0$ *for* $x \neq 0$, *then* $f =$ *constant.*

Probably various known methods of proving the Wiener Tauberian theorem could be adapted to prove Theorem 14 (and theorems of the same type, with α replaced by other functionals). A proof "in the style of Beurling," might run as follows: if $\alpha * f = 0$ then $\alpha * g = 0$ where g is any finite linear combination of translates of f, and passing to the limit in a suitable topology $\alpha * e_\lambda = 0$, where $e_\lambda(t) = e^{-i\lambda t}$, whenever λ belongs to the "spectrum" S of f. Hence, $\hat{\alpha}(\lambda) = 0$ for $\lambda \in S$, showing that S consists of, at most, the number 0, and now, if "spectral synthesis" holds, f is spanned by $\{e_\lambda\}$, $\lambda \in S$ and so is constant a.e. For a proof of Wiener's theorem along these lines, see Beurling [3a]. The general principle is: the solutions of $\beta * f = 0$ (β some functional) are approximable by the exponentials $\{e^{-i\lambda t}\}$ where λ runs through the zeroes of $\hat{\beta}$. Making the argument rigorous in a concrete case (starting with the introduction of a suitable topology, etc.) may involve formidable difficulties, but the principle has been validated in a great many instances.

Notes and Comments

Footnotes added in proof

Page 115. [1] The so-called Friedrichs mollifiers have compact support, but this isn't necessary (or even possible to arrange) in certain applications.

Page 116. [2] Also [66, 73, 75]. The volume of published material on this subject is truly colossal. We found Peetre [73] particularly helpful, and hope this valuable survey article will become available in English.

Page 117: [3] Katznelson's *Introduction to Harmonic Analysis* (Wiley, 1968), which has just appeared, treats both "classical" and "modern" harmonic analysis, and should provide all background needed to read this book (and a great deal to spare!).

Page 119. [4] The limitation being that $\hat{K}$ must not be forced into a quasi-analytic class.

BIBLIOGRAPHY

[1]. N. I. Ahiezer, *Lectures on the Theory of Approximation,* second ed., Moscow, 1965 (Russian; English translation of first ed. available).

[2]. N. V. Araklian, Uniform approximation on closed sets by entire functions, *Izv. Akad. Nauk SSSR Ser Mat. 28* (1964) 1187-1206 (Russian).

[3]. H. Berens and P. L. Butzer, Approximation and Laplace transforms, in *On Approximation Theory,* ed. by Butzer and Korevaar, ISNM Vol. 5, Birkhäuser, Basel, 1964 pp. 24-42.

[3a]. A. Beurling, Un théorème sur les fonctions bornées et uniformement continues sur l'axe réel, *Acta. Math., 77* (1945) 127-136.

[3b]. ______, Sur les intégrales de Fourier absolument convergentes, *Ninth Scandinavian Congres of Mathematicians, Proceedings,* Helsingfors 1939, pp. 345-366.

[5]. S. Bochner, *Lectures on Fourier Integrals,* Princeton, 1949.

[6]. ______, Localization of best approximation, in *Contributions to Fourier Analysis,* Princeton, 1950.

[7]. S. Bochner and K. Chandrasekharan, *Fourier Transforms,* Princeton, 1949.

[8]. J. Boman, Partial regularity of mappings between Euclidean spaces, *Acta Math., 119* (1967) 1-25.

[9]. P. L. Butzer, Zur Frage der Saturationsklassen singulärer Integraloperatoren, *Math. Z., 70* (1958) 93-112.

[10]. ______, Representation and approximation of functions by general singular integrals, *Indagationes Math., 22* (1960) 1-24.

[11]. ______, Fourier transform methods in the theory of approximation, *Arch. Rat. Mech. Anal., 5* (1960) 390-415.

[12]. ______, On some theorems of Hardy, Littlewood and Titchmarsh, *Math. Ann., 142* (1961) 259-269.

[13]. ______, On Dirichlet's problem for the half-space and the behavior of its solution on the boundary, *J. Math. Anal. Appl., 2* (1961) 86-96.

[14]. ______, Beziehungen zwischen den Riemannschen, Taylorschen und gewöhnlichen Ableitungen reellwertiger Funktionen, *Math. Ann., 144* (1961) 275-298.

[15]. P. L. Butzer and H. Berens, Semi-groups of operators and approximation, *Grundlehren der Math. Wiss., vol. 145,* Springer, New York, 1967.

[16]. P. L. Butzer and E. Görlich, Saturationsklassen und asympotische Eigenschaften Trigonometrischer singulärer Integrale, *Weierstrass Festband,* Westdeutscher Verlag, Opladen, 1966, 55 pp.

[17]. P. L. Butzer and S. Pawelke, Ableitungen von trigonometrischen Approximationsprozessen, *Acta Szeged, 28* (1967) 173-183.

[18]. T. Carleman, Sur un théorème de Weierstrass, *Arkiv för Mat., 20* (1927) 1-5.

[19]. E. F. Collingwood and A. J. Lohwater, *The Theory of Cluster Sets,* Cambridge tract no. 56, Cambridge, 1966.

[20]. C. de la Vallée Poussin, *Lecons sur l'approximation des Fonctions d'une Variable Réelle*, Gauther-Villars, Paris, 1919.

[21]. K. de Leeuw, On the adjoint semigroup and some problems in the theory of approximation, *Math. Zeitschr.*, *73* (1960) 219-234.

[22]. J. Favard, Sur l'approximation dės fonctions d'une variable réelle, in *Colloque d'Analyse Harmonique* (Publ. CNRS) *15*, Paris, 1949, pp. 97-100.

[23]. ____, Sur l'approximation dans les espaces vectoriels, *Annali di Mat. Pura Appl.*, *29* (1949) 259-261.

[24]. K. Friedrichs, On the differentiability of the solutions of linear elliptic differential equations, *Comm. Pure Appl. Math.*, *6* (1953) 1-72.

[25]. R. Goldberg, *Fourier Transforms*, Cambridge tract no. 52, Cambridge, 1961.

[26]. M. Golomb, *Lectures on Theory of Approximation*, Argonne National Laboratory, mimeographed notes, 1962.

[27]. L. M. Graves, Some general approximation theorems, *Annals of Math.*, *42* (1941) 281-292.

[28]. G. H. Hardy and J. E. Littlewood, Some properties of fractional integrals, *Math. Zeitschr.*, *27* (1928) 565-606.

[29]. _____, A convergence criterion for Fourier series, *Math. Zeitschr.*, *28* (1928) 612-634.

[30]. I. I. Hirschman and D. V. Widder, *The Convolution Transform*, Princeton, 1955.

[31]. L. Hörmander, A new proof and a generalization of an inequality of Bohr, *Math. Scand.*, *2* (1954) 33-45.

[32]. V. P. Ilyin, On a theorem of Hardy and Littlewood, *Trudy Mat. Inst. Steklova, 53* (1959) 128-144 (Russian).

[33]. J. P. Kahane, *Lectures on Mean Periodic Functions* (mimeographed), Bombay, 1959.

[34]. W. Kaplan, Approximation by entire functions, *Michigan Math. J., 3* (1955) 43-52.

[35]. L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, Moscow, 1959 (Russian; English translation available).

[35a]. L. D. Kudryavtsev, Direct and inverse theorems of embedding; applications to the solutions of elliptic equations by variational methods, *Trudy Mat. Inst. Steklova, 55* (1959) 3-182.

[36]. E. Landau, Über die Approximation einer stetigen Funktion durch eine ganze rationale Funktion, *Rend. Palermo, 25* (1908) 337-345.

[37]. J. E. Littlewood, *Lectures on the Theory of Functions,* Oxford University Press, 1944.

[38]. G. G. Lorentz, *Approximation of Functions,* Holt, Rinehart and Winston, 1966.

[39]. G. G. Lorentz, Inequalities and the saturation class of Bernstein polynomials, in *On Approximation Theory,* ed by Butzer and Korevaar, ISNM, Vol. 5, Birkhäuser, Basel, 1964, pp. 200-207.

[40]. C. J. Morrey, Jr., Multiple Integrals in the Calculus of Variations, *Grundlehren der Math. Wiss., 130,* Springer, 1966.

[41]. I. P. Natanson, *Konstruktive Funktionentheorie,* DVW, Berlin, 1955.

[42a]. S. M. Nikolski, Inequalities for entire functions of finite type and their applications in the theory of differentiable functions of several variables, *Trudy Mat. Inst. Steklova, 38* (1951) 244-278.

[42]. H. R. Pitt, *Tauberian Theorems,* Oxford University Press, 1958.

[43]. W. Rudin, *Real and Complex Analysis,* McGraw-Hill, New York, 1966.

[44]. I. J. Schoenberg, On smoothing operations and their generating functions, *Bull. A.M.S., 59* (1953) 199-230.

[45]. ______, On Variation diminishing approximation methods, in *On Numerical Approximation,* ed. by Langer, Madison, Wisconsin, 1959.

[46]. L. Schwartz, *Théorie des Distributions* (two volumes), Paris, 1950-51.

[47]. H. S. Shapiro, Smoothing and Approximation of Functions, *Matscience Report 55* (mimeographed), Madras, 1967.

[48]. ______, Some Tauberian theorems with applications to approximation theory, *Bull. A.M.S., 74* (1968) 500-504.

[49]. ______, A Tauberian theorem related to approximation theory, *Acta. Math.*

[50]. G. Sunouchi, Saturation in the theory of approximation, in *On Approximation Theory,* ed. by Butzer and Korevaar, ISNM, Vol. 5, Birkhäuser, Basel, 1964, pp. 72-87.

[51]. E. C. Titchmarsh, A theorem on Lebesgue integrals, *J. Lond. Math. Soc., 2* (1927) 36-37.

[52]. ______, *The Theory of Functions,* second edition, Oxford University Press, 1939.

[53]. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals,* second edition, Oxford University Press, 1948.

[54]. A. F. Timan, *Theory of Approximation of Functions of a Real Variable,* Moscow, 1960 (Russian; English translation available).

[55]. M. Tsuji, *Potential Theory in Modern Function Theory,* Maruzen, Tokyo, 1959.

[56]. K. Weierstrass, *Mathematische Werke, Bd. III,* Berlin, 1903, pp. 1-17.

[57]. D. V. Widder, *The Laplace Transform,* Princeton, 1941.

[58]. N. Wiener, *The Fourier Integral,* Cambridge, 1932.

[59]. _____, Tauberian Theorems, *Annals of Math., 33* (1932) 1-100.

[60]. M. Zamansky, Classes de saturation de certaines procédés d'approximation des séries de Fourier des fonctions continues, *Ann. Sci. Ecole Norm. Sup., 66* (1949) 19-93.

[61]. _____, Classes de saturation des procédés de sommation des séries de Fourier et applications aux séries trigonométriques, *Ann. Sci. Ecole Norm. Sup., 67* (1950) 161-198.

[62]. A. Zygmund, *Trigonometric Series,* second ed., Cambridge, 1959.

[63]. _____, A remark on the integral modulus of continuity, *Univ. Nac. Tucumán, Revista A.7* (1950) 259-269.

[64]. _____, The approximation of functions by typical means of their Fourier series, *Duke Math. J., 12* (1945) 695-704.

Additional References

[65]. Jan Boman and H. S. Shapiro, Inverse theorems of approximation theory in L^p, in preparation.

[66]. V. I. Burenkov, Imbedding and continuation for classes of differentiable functions of several variables defined in the whole space, translated from Russian in *Progress in Math.*" *Vol. 2,* Plenum Press, New York, 1968, 73-161.

[67]. P. L. Butzer and E. Görlich, Characterizations of Favard classes for functions of several variables, *Bull. A.M.S., 74* (1968) 149-152.

[68]. P. L. Butzer and R. J. Nessel, Favard classes for n-dimensional singular integrals, *Bull. A.M.S., 72* (1966) 493-498.

[69]. _____, Contributions to the theory of saturation for singular integrals in several variables I, *Indag. Math., 28* (1966) 515-531.

[70]. K. de Leeuw and H. Mirkil, A priori estimates for differential operators in L_∞ norm, *Illinois J. Math., 8* (1964) 112-124.

[71]. E. Görlich, Saturation theorems and distributional methods, *Proceedings of Conference at Oberwalfach,* ISNM, Birkhäuser, Basel, 1969 (to appear).

[71a]. E. Görlich and R. Nessel, Peano und Riemann-Ableitungen in der Norm, *Arch. Math.* (Basel), *12* (1967) 399-410.

[72]. R. J. Nessel, Contributions to the theory of saturation for singular integrals in several variables II, III, *Indag. Math., 29* (1967) 52-73.

[73]. Jaak Peetre, *Reflections on Besov Spaces,* mimeographed, Lund, 1966 (Swedish).

[74]. H. S. Shapiro, Approximation by trigonometric polynomials to periodic functions of several variables, *Proceedings of Conference at OLerwolfach,* ISNM, Birkhäuser, Basel 1969 (to appear).

[75]. M. H. Taibleson, On the theory of Lipschitz spaces of distributions in Euclidean *n*-space, I, *J. Math. Mech., 13* (1964) 407-419 and II, *ibid., 14* (1965) 821-840.

[76]. P. Ulyanov, Embedding of certain function classes H_p^ω, *Izv. Akad. Nauk SSSR, Ser. mat., 32* (1968) 649-686 (Russian).

[77]. M. Weiss and A. Zygmund, A note on smooth functions, *Indag. Math., 21* (1959) 52-58.

A VAN NOSTRAND REINHOLD MATHEMATICAL STUDY

under the general editorship of

PAUL R. HALMOS
Indiana University

FREDERICK W. GEHRING
The University of Michigan

•

About this book:

The present book is a revised and expanded edition of notes based on lectures given at the Institute of Mathematical Sciences, Madras, in January and February of 1967. The unifying theme is the relation between a given function and its "smoothed" version, obtained by forming its convolution with $K\lambda(x) = \lambda K(\lambda x)$ where the "kernel" K is an integrable function satisfying $\int_{-\infty}^{\infty} K(x)dx = 1$ and λ is a "large" parameter. Special as it is, this problem nevertheless subsumes much attractive material, e.g., the limiting behavior of harmonic functions in a half plane and of solutions of the heat equation, as well as the classical theory (Jackson, S. Bernstein) of the approximation of periodic functions by trigonometric polynomials.

The material should be accessible to a student who has had a course in Lebesgue integration, except that in Chapter V a fair amount of harmonic analysis is needed. The relevant chapters of Rudin's "Real and Complex Analysis" should adequately prepare a prospective reader, and for the first four chapters, familiarity with Chapters X, XI, and XII [illegible] tions" is more than sufficient. [illegible] ld prove rewarding to a student [illegible] ad some basic courses and wants [illegible] he standard tools.

The Author:

HAROLD S. SHAPIRO (Ph.D., Massachusetts Institute of Technology) is currently Professor of Mathematics at The University of Michigan. Professor Shapiro's writing and research is largely devoted to the area of Complex Analysis.

VAN NOSTRAND REINHOLD CO.