MATSCIENCE REPORT 55

SMOOTHING AND APPROXIMATION OF FUNCTIONS

HAROLD S. SHAPIRO
Visiting Professor, MATSCIENCE

THE INSTITUTE OF MATHEMATICAL SCIENCES, MADRAS-20, INDIA.

MATSCIENCE REPORT No.55

THE INSTITUTE OF MATHEMATICAL SCIENCES
Madras-20, (INDIA)

SMOOTHING AND APPROXIMATION OF FUNCTIONS

Harold S. Shapiro+,
Visiting Professor, MATSCIENCE

Permanent Address:-
Department of Mathematics, University of Michigan, Michigan, U.S.A.

P R E F A C E

The present notes are based on lectures given at the Institute of Mathematical Sciences, Madras, in January and February of 1967. In essence they constitute an introductory course in the theory of approximation (however, we discuss only questions of "good" approximation, or degree of approximation, rather than "best" approximation, which requires quite different concepts and methods). The discussion is limited to functions on the real line, and (except for a few introductory Theorems) 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.

The treatment given here is not altogether orthodox, and has some theoretical pretensions. We have treated trigonometric approximation on the line group, rather than the circle group. This is not an altogether trivial change, since the unrolling of the circle onto the line gives rise to different kernels and integral formulas, often simpler ones (this device was noted by la Vallée Poussin, as well as later authors, especially Bochner, Akhieser, and Butzer, yet not, so far as we know, fully exploited).

Most of the material in Chapter 5 is new, and may be thought of as carrying out a program begun by P.L.Butzer in a series of papers (see the References at the end). To our knowledge, the theorems given here are the first known "inverse theorems"

for general kernels, in fact even for quite special kernels we could find no inverse theorems in the literature where the order of magnitude of the approximation is larger than the saturation order (these terms shall be duly explained in the following pages). The "Tauberian" formulation of "inverse problems" we have given is so general that it reveals both "direct" and "inverse" problems as part of a single general problem, which we have in large measure solved.

We have striven for thematic unity in the presentation, and taken as our main theme the generation of approximations by convolution integrals. Not unexpectedly, where there are convolutions we haven't long to wait before Fourier transforms make their appearance. The only specialized knowledge required for reading these notes (that is, knowledge beyond the elements of real analysis) is the elements of Fourier transforms, as expounded in the early chapters of Titchmarsh's treatise, plus the deeper Wiener theory needed in chapter 5. During the lectures this material was reviewed, but has not been reproduced in these notes.

The reader interested in further pursuing the topics treated here may refer to the literature cited at the end. There are moreover a number of other excellent books on approximation theory devoted predominantly to different topics, to name a few, those of P.J.Davis, P.Korovkin, G.Meinardus and the mimeographed notes of S.Golomb.

The author wishes to thank Dr.K.R.Unni, Mr.G.N.Keshavamurthy and Mr.M.R.Subrahmanya for their kind assistance in the preparation of these notes.

H.S.S.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introductory remarks

A fundamental problem in analysis is to find functions which approximate a given function f in some sense and which are better behaved than f, in the sense of smoothness (or regularity). Roughly speaking, we are interested in approximating ''bad'' functions by ''good'' functions. It is natural to expect that we must construct ''good'' functions from the given function f by some smoothing operations on f itself.

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 ?

The answer to this question is in the affirmative; here we approximate in the L^1-metric and obtain the following theorem.

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 by the formula

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

where

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \quad (2)$$

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

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

so that

$$f_a(x) = \int_{-\infty}^{\infty} f(x+t)\,G_a(t)\,dt \quad (3)$$

where

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

Using

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

it follows that

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

$$= \int_{-\infty}^{\infty} \left[f(x+t) - f(x) \right] G_a(t)\,dt \qquad (4)$$

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)

$$= \int_{-\infty}^{\infty} G_a(t)\,\omega(t)\,dt \qquad (5)$$

where

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

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

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

a is small. This completes the proof of Theorem 1.

1.3 Let us now generalize from this example. It is convenient to make a change of variable and write (3) in the form

$$f_a(x) = \int_{-\infty}^{\infty} f(x-t) G_a(t)\,dt. \qquad (3')$$

This suggests the following definition

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

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

+) 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.

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

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

and writing, for $\lambda > 0$,

$$^{+)}K_\lambda(x) = \lambda K(\lambda x)$$

we have

$$G_a(x) = \lambda K(\lambda x) \quad \text{if} \quad \lambda = \frac{1}{a}.$$

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

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

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

$$= \int_{-\infty}^{\infty} f(x-\tfrac{t}{\lambda})\, K(t)\,dt. \quad (6)$$

In the remainder of these lectures we shall examine the behaviour 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 <u>supremum norm</u>, but the techniques developed below apply <u>mutatis mutandis</u> 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 (<u>direct theorems</u> of approximation theory); on intrinsic limits to the goodness of the approximation imposed by the peculiarities of the kernel K (<u>saturation theorems</u>); and on inference of smoothness properties of f from the smallness of the deviation

---

+) In these notes we denote $(f*K_\lambda)(x)$ either by $f^\lambda(x)$ or $f(x;\lambda)$.

$f(x;\lambda)-f(x)$ (<u>inverse theorems</u>). It will turn out that these questions are decisively influenced by the behaviour of the Fourier transform $\hat{K}(x)$ of $K(t)$, notably its <u>flatness</u> 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 behaviour of a convolution identity or "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 a function $K(x)$ of <u>one</u> variable.) Dually, observe that the Fourier transform of K_λ is $\hat{K}(\frac{x}{\lambda})$, 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 The main purpose of these lectures 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 success 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, bounds on f and its derivatives, 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 presense 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 transforms puts powerful techniques from harmonic analysis at our disposal. Furthermore, the asymptotic behaviour of convolutions is usually easy to estimate.

In addition, it turns out (although this is far from obvious <u>a priori</u>) 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 further enhance the importance of convolutions, and play an essential role in the theory which follows.

+) For precise formulation, see Chapter 4.

CHAPTER 2

SOME ELEMENTARY THEOREMS ON APPORXIMATION

2.1 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 1.

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

$$^{+)}\text{(i)} \quad f^\lambda = f * K_\lambda \in L^1$$

<u>and</u>

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

PROOF. (i) follows from property (i) of convolutions (§ 1.3).

Next, as in the proof of Theorem 1,

$$\|f^\lambda - f\|_1 \leq \int_{-\infty}^{\infty} \omega(t)\,|K_\lambda(t)|\,dt \tag{1}$$

+) 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

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

Given $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon)$ such that

$$\omega(t) < \varepsilon \quad \text{for} \quad |t| < \delta.$$

Moreover ω is bounded, say $\omega(t) \leq C$. Then (1) gives

$$\|f^{\lambda}-f\|_1 \leq \int_{|t|<\delta} \omega(t)\,|K_{\lambda}(t)|\,dt + \int_{|t|\geq\delta} \omega(t)\,|K_{\lambda}(t)|\,dt$$

$$\leq \varepsilon \int_{|t|<\delta} |K_{\lambda}(t)|\,dt + C \int_{|t|\geq\delta} |K_{\lambda}(t)|\,dt \qquad (2)$$

Now

$$\int_{-\delta}^{\delta} |K_{\lambda}(t)|\,dt = \int_{-\delta}^{\delta} \lambda\,|K(\lambda t)|\,dt \leq \int_{-\infty}^{\infty} |K(t)|\,dt =$$

$$= \|K\|_1 \qquad (3)$$

and

$$\int_{|t|\geq\delta} |K_{\lambda}(t)|\,dt = \int_{|u|\geq\lambda\delta} |K(u)|\,du < \varepsilon \text{ if } \lambda > \lambda_0(\varepsilon). \qquad (4)$$

The result now follows from (2), (3) and (4).

We can now greatly strengthen Theorem 1:

<u>Corollary.</u> 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 $K(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$. 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

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

we see that F_n is entire, and moreover for every $R > 0$ $\max_{|z| \leq R} |G_n(z)|$ 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.

<u>Exercise.</u> The reader should extend Theorem 2 and the Corollary to L^p norm ($p > 1$).

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

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

PROOF. As before, we have

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

so that

$$|f(x)-f(x;\lambda)| \leq \int_{-\infty}^{\infty} \omega(t)\,|K_\lambda(t)|\,dt \tag{5}$$

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.

<u>Corollary 1</u>. 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. 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 we shall later give a still stronger result of this kind, with approximating functions of exponential type. Such results are not valid for arbitrary continuous functions.

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 We now prove a theorem concerning point-wise convergence.

THEOREM 4. Let σ be a function of finite variation on $(-\infty,\infty)$. Let K be a function such that $K \in L^1$ $L^1(-\infty,\infty)$, $K(-x) = K(x)$ and $\int_{-\infty}^{\infty} K(x)\,dx = 1$. Moreover, suppose K satisfies at least one of the following two conditions

(i) K is continuous and nonincreasing on $(0,\infty)$

(ii) K is absolutely continuous, $xK'(x) \in L^1(-\infty,\infty)$ and $xK(x) \to 0$ as $x \to \pm\infty$.

Let

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

Then

A) 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

$$\lim_{\lambda \to \infty} f^{\lambda}(x) = \sigma'(x). \qquad (*)$$

B) If, moreover, K satisfies (i), then (*) holds also whenever $\sigma'(x) = \pm\infty$.

Remark 1. Wherevr σ possesses a finite or infinite 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. Important examples of kernels K satisfying form (i) of the hypotheses are:

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

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

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

On the other hand,

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

satisfies neither form of the hypotheses,[+)] and

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

satisfies form (ii).

We shall prove the theorem under hypothesis (i). The proof when K satisfies (ii) is left to the reader.

[+)] None the less, part A) of the conclusion of Theorem 4 applies also to this kernel, as a slight modification of the following proof shows.

PROOF. By definition

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

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

$$= K_{\lambda}(t)\left[\sigma(x+t) - \sigma(x-t)\right]\Big|_{0}^{\infty} - \int_{0}^{\infty} \left[\sigma(x+t) - \sigma(x-t)\right] dk_{\lambda}(t)$$

$$= - \int_{0}^{\infty} \left[\sigma(x+t) - \sigma(x-t)\right] dK_{\lambda}(t) \qquad (6)$$

if x is a point of continuity of σ, which we shall suppose. Suppose x is a point at which σ possesses a symmetric derivative. Then

$$\frac{\sigma(x+t) - \sigma(x-t)}{2t} = \sigma'(x) + \mathcal{E}_x(t) \qquad (7)$$

where $\mathcal{E}_x(t) \to 0$ as $t \to 0$, and $|\mathcal{E}_x(t)| \leq C$, where C is a constant.

Using (7), (6) becomes

$$f^{\lambda}(x) = -\int_0^{\infty} [\sigma'(x) + \mathcal{E}_x(t)]\ 2t dK_{\lambda}(t)$$

$$= -2\sigma'(x)\int_0^{\infty} t dK_{\lambda}(t) - 2\int_0^{\infty} \mathcal{E}_x(t)\, t dK_{\lambda}(t)$$

$$= -2\sigma'(x)[tK_{\lambda}(t)]_0^{\infty} + 2\sigma'(x)\int_0^{\infty} K_{\lambda}(t)dt$$

$$- 2\int_0^{\infty} \mathcal{E}_x(t)\, t dK_{\lambda}(t).$$

+) Since $tK_{\lambda}(t) \to 0$ as $t \to \infty$ and $\int_0^{\infty} K(t)dt = \frac{1}{2}$, we have

$$f^{\lambda}(x) = \sigma'(x) - 2\int_0^{\infty} \mathcal{E}_x(t)\, t dK_{\lambda}(t) \qquad (8)$$

To complete the proof, it is enough to show that

$$I = \int_0^{\infty} |\mathcal{E}_x(t)|\, t\, |dK_{\lambda}(t)| < \mathcal{E} \text{ for } \lambda \text{ large enough.}$$

We choose δ such that $|\mathcal{E}_x(t)| < \mathcal{E}$ for $0 \leq t \leq \delta$.

+) This is a consequence of the monotonicity and integrability of K_{λ}.

Then

$$I \leq \varepsilon \int_0^\delta t|dK_\lambda(t)| + C \int_\delta^\infty t|dK_\lambda(t)| . \tag{9}$$

Since K_λ is nonincreasing, we have

$$|dK_\lambda(t)| = -dK_\lambda(t).$$

Therefore

$$\int_0^\delta t|dK_\lambda(t)| = -\int_0^\delta t dK_\lambda(t) = -tK_\lambda(t) \Big|_0^\delta + \int_0^\delta K_\lambda(t)\,dt$$

$$= -\delta K_\lambda(\delta) + \lambda \int_0^\delta K(\lambda t)\,dt \leq \int_0^{\lambda\delta} K(x)\,dx \leq \frac{1}{2} \tag{10}$$

Similarly

$$\int_\delta^\infty t|dK_\lambda(t)| = \lambda\delta K(\lambda\delta) + \int_{\lambda\delta}^\infty K(x)\,dx \tag{11}$$

From (9), (10) and (11) we get

$$I \leq \frac{\varepsilon}{2} + C\left[\lambda\delta K(\lambda\delta) + \int_{\lambda\delta}^\infty K(x)\,dx\right]$$

and since the bracketed expression tends to zero for large λ, we have $I \leq \varepsilon$ for $\lambda \geq \lambda_0$, proving A).

We now proceed to prove (B). Suppose that $\sigma'(x) = +\infty$. Then, if $M > 0$ is given, we can choose δ such that $|t| < \delta$ implies

$$\frac{\sigma(x+t) - \sigma(x-t)}{2t} \geq M \tag{12}$$

and by (6),

$$f^{\lambda}(x) = \int_0^{\infty} \frac{\sigma(x+t) - \sigma(x-t)}{2t} \, 2t \, (-dK_{\lambda}(t))$$

$$= \left(\int_0^{\delta} + \int_{\delta}^{\infty} \right) \frac{\sigma(x+t) - \sigma(x-t)}{2t} \, 2t \, (-dK_{\lambda}(t)). \tag{13}$$

Now, by (12), we obtain

$$\int_0^{\delta} \frac{\sigma(x+t) - \sigma(x-t)}{2t} \, 2t \, (-dK_{\lambda}(t)) \geq 2M \int_0^{\infty} t(-dK_{\lambda}(t))$$

$$= 2M \left[\int_0^{\lambda\delta} K(t)\,dt - \lambda\delta K(\lambda\delta) \right]$$

$$\geq \frac{M}{2} \text{ if } \lambda \geq \lambda_1(M).$$

Finally we show that

$$\int_{\delta}^{\infty} \frac{\sigma(x+t) - \sigma(x-t)}{2t} \, 2t \, (-dK_{\lambda}(t))$$

remains bounded. This follows from

$$\left| \int_{\delta}^{\infty} \frac{\sigma(x+t) - \sigma(x-t)}{2t} \, 2t \, (-dK_{\lambda}(t)) \right| \leq \frac{C}{\delta} \int_{\delta}^{\infty} (-dK_{\lambda}(t)) =$$

$= \frac{C}{\delta} K_{\lambda}(\delta) = \frac{C}{\delta^2} \cdot \lambda\delta K(\lambda\delta) < 1$, for $\lambda > \lambda_2(M)$. Hence, for λ large enough, $f^{\lambda}(x) > \frac{M}{2} - 1$, which means that $f^{\lambda}(x) \to \infty$ as $\lambda \to \infty$.

This completes the proof of Theorem 4.

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

<u>Proof</u>. Apply Theorem 4 to $\sigma(x) = \int_{-\infty}^{x} f(t)\,dt$.

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

COROLLARY 2. <u>Let</u> $u(x,y)$ <u>be a harmonic function in</u> $y > 0$ <u>and nonnegative. Then</u>

$$\lim_{y \to 0+} u(x,y) \quad \underline{\text{exists and is finite}} \quad \text{a.e.}$$

PROOF. By a theorem of Herglotz there exists a bounded non-decreasing function σ, and a real number c, such that

$$u(x,y) = cy + \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{d\sigma\,(t+x)}{t^2 + y^2} \tag{13}$$

where $c \geq 0$. To complete the proof, it is enough to express the integral in (13) as the convolution product of a suitable kernel K and σ; then Theorem 4 will give the result. To this end, we take

$$K(x) = \frac{1}{\pi(1+x^2)} \quad \text{and} \quad \lambda = \frac{1}{y} \quad \text{(Cauchy kernel)}$$

Then $$K_\lambda(x) = \lambda K(\lambda x) = \frac{y}{\pi(x^2+y^2)} .$$

<u>Remark.</u> The integral defining $u(x,y)$ in (13) is the <u>Poisson-Stieltjes integral</u> (relative to the upper half plane) of σ. The theory of boundary behaviour of harmonic functions is in large measure just the behaviour of this integral for small y; it is of interest that many classical theorems can be carried over to the behaviour of the integral $f*K_\lambda$ for a wide class of kernels K. This programme shall not be carried out here, however.

COROLLARY 3. (<u>Uniqueness theorem for Fourier-Stieltjes transform</u>). <u>Let</u>

$$f(x) = \int_{-\infty}^{\infty} e^{itx} d\sigma(t) \tag{14}$$

<u>where</u> $d\sigma$ <u>is any function of finite variation on</u> $(-\infty, \infty)$. <u>If</u> $f(x) = 0$ <u>for all</u> x, <u>then</u> σ <u>is a constant</u>.

PROOF. Multiply (14) by $e^{-a|x|-ixu}dx$ where $a > 0$, u real and integrate. Then

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

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

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

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

We may write this as

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

where $\lambda = \frac{1}{a}$, $K(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

From Theorem 4 it follows, first, that $\sigma'(x)$ is never equal to $\pm\infty$; therefore by classical results in differentiation theory, σ is absolutely continuous. Now $\sigma'(x) = 0$ a.e. by Theorem 4 and hence σ is a constant.

2.3 <u>Degree of approximation.</u> Having now illustrated various kinds of approximation (L^1, uniform, pointwise almost everywhere) we shall, in the remainder of these lectures, confine ourselves to <u>uniform</u> approximation of continuous functions.

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

DEFINITION. We say that $f \in \mathrm{Lip}\,\alpha$, $0 < \alpha \leq 1$, 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 $f \in \mathrm{Lip}\,\alpha$, then

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

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

$$= C\lambda^{-\alpha} \int_{-\infty}^{\infty} |x^{\alpha} K(x)|\,dx.$$

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

Thus we have the following theorem.

THEOREM 5. <u>If</u>

$$M = \int_{-\infty}^{\infty} |x^{\alpha} K(x)|\, dx < \infty, \quad 0 < \alpha \leq 1 \tag{15}$$

<u>then</u> $f \in \mathrm{Lip}\,\alpha$ <u>implies</u>

$$\| f^{\lambda} - f \| \leq \frac{C\,M}{\lambda^{\alpha}} .$$

If we don't wish to assume K satisfies (15), we can proceed as follows:

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

$$\leq C \int_{-T}^{T} |t|^{\alpha}\, |K_{\lambda}(t)|\, dt + C_1 \int_{|t| > T} |K_{\lambda}(t)|\, dt$$

$$= C\lambda^{-\alpha} \int_{-T\lambda}^{T\lambda} |uK(u)|\, du + C_1 \int_{|u| > T\lambda} |K(u)|\, du \tag{16}$$

where T is a positive number which we are free to choose optimally. For example, if $K(t) = \mathcal{O}(t^{-2})$ we see (choosing $T = 1$ in (16)) that $f \in \mathrm{Lip}\,\alpha$ implies $\|f^{\lambda} - f\| = \mathcal{O}\left(\frac{\log \lambda}{\lambda}\right)$.

This is a case of some importance, since it covers the Cauchy and Fejér-la Vallée Poussin 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 $\mathcal{O}\left(\frac{1}{\lambda}\right)$, in view of Theorem 5.

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

CHAPTER 3

SATURATION THEOREMS

3.1 One of the most important phenomena of approximation theory 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. Let

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

Then

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

so that

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

hence

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

Now $f(x)$ is as smooth a function as we could wish, and yet we only get the degree of approximation a^2.

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

<u>Suppose</u> $f''(x)$ <u>exists</u>. <u>Then</u>

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

PROOF. By hypothesis, we have

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

where

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

Now

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

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

(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 $\| f_a - f \| = \mathcal{O}(a^2)$, then f is a linear polynomial

(ii) if $\| f_a - f \| = o(a^2)$, then

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

where A is a constant.

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, We shall (without troubling to give a <u>formal</u> definition of "saturation") 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 <u>saturation class</u> is the set of f satisfying (ii) (which, as we'll see later, is the same as the class of absolutely continuous f with $f' \in \mathrm{Lip}\ 1$)".

<u>Exercise</u>. Show that for the <u>asymmetric</u> (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

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

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

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 correct (and a rigorous version will be stated and proved later). The qualitative formulation is the principle: <u>the flatter $\hat{K}$ is at $x = 0$, the smaller the order of magnitude at which saturation occurs.</u>

In particular, if $\hat{K}$ has many vanishing derivatives at $x = 0$ (equivalently, if many <u>moments</u> 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 neighbourhood of $x = 0$.

3.2 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 even, $\int_{-\infty}^{\infty} K(x)\,dx = 1$ and such that

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

Let

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

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

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

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 behaviour of a quadratic polynomial; the coefficient of $\frac{t^2}{2!}$ 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 conincides with the usual in case $f \in C^2$ in the neighbourhood of x.

PROOF OF THEOREM 6. As before

$$f^{\lambda}(x)-f(x) = \int_{-\infty}^{\infty} \left[f(x+t)-f(x) \right] K_{\lambda}(t)\,dt.$$

From (1), we get (denoting by $\langle M \rangle$, a quantity bounded in absolute value by M (Lindelöf symbol)).

$$f^{\lambda}(x)-f(x) = \int_{-\delta}^{\delta} \left[tf'(x)+\frac{t^2}{2} f''(x)+\varepsilon_x(t)t^2 \right] K_{\lambda}(t)\,dt$$

$$+ \left\langle 2\|f\| \int_{|t|>\delta} |K_{\lambda}(t)|\,dt \right\rangle$$

$$= f''(x) \int_0^{\delta} t^2 K_{\lambda}(t)\,dt + \left\langle 2\varepsilon \int_0^{\delta} t^2 |K_{\lambda}(t)|\,dt \right\rangle$$

$$+ O\left(\int_{|t|>\delta\lambda} |K(t)|\,dt \right)$$

Now

$$\int_0^{\delta} t^2 K_{\lambda}(t)\,dt = \lambda^{-2} \int_0^{\lambda\delta} x^2 K(x)\,dx = \lambda^{-2}(A+o(1)), \qquad (3)$$

and

$$\int_0^{\delta} t^2 |K_\lambda(t)|\, dt \leq \lambda^{-2} \int_0^{\infty} t^2 |K(t)|\, dt = B\lambda^{-2}. \quad (4)$$

Moreover, we have

$$\lim_{\lambda \to \infty} \int_{|t| > \delta\lambda} |K(t)|\, dt = 0 \quad (5)$$

because

$$\lambda^2 \int_{|t| \geq \delta\lambda} |K(t)|\, dt \leq \frac{1}{\delta^2} \int_{|t| \geq \delta\lambda} t^2 |K(t)|\, dt \to 0 \quad \text{as } \lambda \to \infty.$$

Hence, using (3), (4), (5)

$$\overline{\lim_{\lambda \to \infty}} \left| \lambda^2 \left[f^\lambda(x) - f(x) \right] - Af''(x) \right| \leq 2B\varepsilon.$$

This completes the proof, since ε is arbitrary.

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

$f(x+h) - 2f(x) + f(x-h) = \mathcal{O}(h^2)$, uniformly with respect to x.

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 $\|f-f^{\lambda}\| = o(\frac{1}{\lambda^2})$ then f is constant.

(ii) If $\|f-f^{\lambda}\| = O(\frac{1}{\lambda^2})$ 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 $\sup_{a<x<b} |f(x)-f^{\lambda}(x)| = o(\frac{1}{\lambda^2})$ then f is linear in (a,b).

(ii) If $\sup_{a<x<b} |f(x)-f^{\lambda}(x)| = O(\frac{1}{\lambda^2})$ then $f \in S$ in (a,b)

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-la Vallée Poussin) we find the identical saturation behaviour as in the case of the 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 get

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

Then from (i) is follows that

$$|g^{\lambda}(x)| \leq \frac{\varepsilon(\lambda)}{\lambda^2} \quad \text{for} \quad a < x < b \tag{6}$$

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 don't 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 (so-called ''Friedrichs method of mollifiers'').

We define (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 $\mu > \frac{1}{\delta}$

$$\left| (g^\lambda * J_\mu)(x) \right| = \left| \int_{-\delta}^{\delta} g^\lambda(x+t)\mu J(\mu t)\,dt \right| \leq \frac{\mathcal{E}(\lambda)}{\lambda^2} \ .$$

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

This implies

$$\left| f_\mu^\lambda(x) - f_\mu \right| \leq \frac{\mathcal{E}(\lambda)}{\lambda^2} \ , \quad a+\delta \leq x \leq b-\delta$$

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 (6) is replaced by

$$\left| g^\lambda(x) \right| \leq \frac{B}{\lambda^2} \quad \text{for} \quad a < x < b$$

and we conclude, as before,

$$\left| f_\mu^\lambda(x) - f_\mu(x) \right| \leq \frac{B}{\lambda^2}, \qquad a+\delta \leq x \leq b-\delta .$$

Therefore, by Theorem 6, $\left|f''_\mu(x)\right| \leq \frac{B}{|A|} = B_1$, hence by the mean value theorem, if $x-h$ and $x+h$ lie in $[a+\delta, b-\delta]$,

$$\left| f_\mu(x+h) - 2f_\mu(x) + f_\mu(x-h) \right| \leq B_1 h^2 \tag{7}$$

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

<u>Remark</u>. We have already pointed out earlier that S is identical with the class of f having Lip 1 derivatives. 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 (7), 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

$$\int_{x_1}^{x_2} \varphi(t)\,dt = f(x_2) - f(x_1).$$ The details are left to the reader.

<u>Exercise.</u> Prove that S is identical with the class of absolutely continuous f satisfying $f' \in \mathrm{Lip}\ 1$. (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).

3.3. For kernels which don't fall off fast enough at infinity for the preceding theorems to be applicable (i.e. when x^2K isn't integrable), saturation problems may be fairly delicate. (This may also be the case when the second moment exists and is zero; however, few kernels of practical interest have this property). One rather trivial result is:

THEOREM 9. <u>Let</u> K <u>and</u> xK <u>be integrable, then at any point</u> x <u>where</u> $f'(x)$ <u>exists,</u>

$$\lim_{\lambda \to \infty} \lambda\left[f^{\lambda}(x)-f(x)\right] = -\left(\int_{-\infty}^{\infty} tK(t)\,dt\right) f'(x)$$

PROOF.

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

and letting $\lambda \to \infty$, the result follows by dominated convergence.

From Theorem 9, one easily deduces (as in the proof of Theorem 8)

THEOREM 10. *For* K *as in Theorem 9, the saturation order is* λ^{-1} *and the saturation class is Lip 1.*

For kernels with xK not integrable (or, xK integrable and having integral equal to zero) we have still no result. In particular, determination of the saturation order and class for the Cauchy and Fejér-la Vallée Poussin kernels is unsettled. For the Cauchy kernel we have the following:

THEOREM 11. *Let* $K(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$, *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*

$$\lim_{\lambda \to \infty} \lambda \left[f(x;\lambda) - f(x) \right] = J(f;x) \tag{8}$$

PROOF.

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

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

$$\lambda\left[f(x;\lambda)-f(x)\right] = \frac{1}{\pi}\int_0^\infty \left[\frac{f(x+t)-2f(x)+f(x-t)}{t^2} \cdot \frac{\lambda^2 t^2}{1+\lambda^2 t^2}\right] dt \qquad (9)$$

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. <u>Theorem 11 holds with</u> K <u>replaced by</u> $\frac{\sin^2 x}{\pi x^2}$, <u>except that the right side of</u> (8) <u>must be replaced by</u> $\frac{1}{2} J(f;x)$.

PROOF. The proof proceeds as before, but now in (9) 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 $\sin^2\lambda t = \frac{1}{2}(1 - \cos 2\lambda t)$, 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 - la Vallée Poussin kernel, and let* f *be bounded and continuous on* $(-\infty,\infty)$.

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

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

PROOF. By Theorem 12, and the "mollifier technique" the proof is reduced to demonstrating

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

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

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. We consider the Poisson integral

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

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

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

$$= \frac{1}{\pi} \int_{0}^{\infty} \left[f(x+t) + f(x-t) \right] M(t,y)\, dt$$

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

$$u_y(x,y) = \frac{1}{\pi} \int_{0}^{\infty} \left[f(x+t) - 2f(x) + f(x-t) \right] M(t,y^2)\, dt$$

$$= \frac{1}{\pi} \int_{0}^{\infty} \frac{f(x+t)-2f(x)+f(x-t)}{t^2} \cdot \frac{t^2(t^2 - y^2)}{(t^2+y^2)^2}\, dt. \tag{10}$$

For the remainder of the proof we suppose that f'' is bounded on $(-\infty,\infty)$ (the general case is easily reduced to this by the ''mollifier'' technique). Let $C = \sup |f''(x)|$, then from (10)

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

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

showing that u_y is bounded in the half-plane $y > 0$. Now, from (10), 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, and so finally

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

is constant, and the Theorem is proved.

<u>Remark</u>. The saturation class, as described in part (ii) of Theorem 13, admits an alternate description in terms of conjugate harmonic functions. We leave the details to the reader.

3.4 If the function f happens to be <u>periodic</u> (say of period 2π) one can use a Fourier method to find the saturation order and (often) the saturation class. (such a device is however not available for f merely continuous and bounded on $(-\infty,\infty)$, since it is extremely difficult to introduce any useful definition of Fourier transform for such functions.) The basis of the Fourier method is

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

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

<u>Let</u> $K \in L^1$, <u>then</u>

$$|c_n| \cdot |1 - \hat{K}(\tfrac{n}{\lambda})| \leq \|f - (f*K_\lambda)\| \tag{11}$$

PROOF. Denote the right side of (11) by M. Then

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

Hence

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

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 - la Vallée Poussin kernel, $1 - \hat{K}(\frac{n}{\lambda}) = \frac{|n|}{\lambda}$ for $|\lambda| \geq n$. If therefore $\lambda.\|f-(f*K_\lambda)\| \to 0$ as $\lambda \to \infty$, we deduce from (11) that nc_n vanishes for each n, hence f is constant. (It is easily seen 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 $O(\frac{1}{\lambda})$ implies, using (11), that the Fourier coefficients of f are $O(\frac{1}{n})$. 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 5.

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. (Tatchmarsh). Let $f \in L^1(-\infty,\infty)$ and $\int_{-\infty}^{\infty} |f(x+t) - f(x)|\, dx = o(t)$, $t \to 0+$. Then f is (almost everywhere) a constant.

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

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

Therefore $f(x+t) - f(x) \geq tf'(x) - \frac{Mt^2}{2}$

Now, for any numbers a,b with $a < b$, by hypothesis

$$\int_a^b |f(x+t) - f(x)|\, dx \leq \varphi(t)$$

where $\frac{\varphi(t)}{t} \to 0$ as $t \to 0+$. Hence

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

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

Dividing by t, and letting $t \to 0$ we conclude $f(b) \leq f(a)$. Since the same arguement 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,

$$\int |f^\lambda(x+t) - f^\lambda(x)|dx = \int \left| \int (f(x+t-u) - f(x-u))K_\lambda(u)du \right| dx$$

$$\leq \int K_\lambda(u) \left[\int |f(x+t-u) - f(x-u)| dx \right] du$$

$$\leq \int |f(x+t) - f(x)| dx$$

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.

THEOREM 17. <u>(Hardy)</u>. <u>Let</u> $f \in L^1(-\infty,\infty)$ <u>and</u>

$$\int_{-\infty}^{\infty} |f(x+t) - f(x)| dx = \mathcal{O}(t), \ t \to 0+.$$ <u>Then</u> f <u>concides almost everywhere with a function of bounded variation.</u>

PROOF. 1) Suppose $f \in C^2$ 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 intervals. Then (reasoning just as 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)

$$\int_E |f'(x)|\, dx \leq A .$$

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

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

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

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.

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

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

CHAPTER 4

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 $\tilde{C}$ denote the class of continuous functions of period 2π. If $f \in \tilde{C}$ and $f \in \mathrm{Lip}\,\alpha (0 < \alpha \leq 1)$ then, as is well known, f can be approximated by an element of $\mathcal{T}_n$ with an error $\mathcal{O}(n^{-\alpha})$. The usual proof is to convolve f with a suitable peaking kernel K (here 'convolution' is meant relative to the circle group: $(f*g)(x) = \frac{1}{2\pi} \int_0^{2\pi} f(x-t)\, g(t)dt$), 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 $B_n\left(\frac{\sin nx}{\sin x}\right)^2$ is not sufficient - it works for each $\alpha < 1$, but fails at $\alpha = 1$. This reflects the fact that the Fejér kernel is saturated with order $\frac{1}{n}$, and the saturation class does not contain all Lip 1 functions. The Jackson kernel, on the other hand, is saturated with order $\frac{1}{n^2}$.

It is important to understand what structural property of the Jackson kernel makes it 'less saturated' than the Fejér kernel. This 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 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 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 $f \in \tilde{C}$. We have

THEOREM 18. <u>Let</u> $f \in \tilde{C}$ (<u>i.e. continuous and of period</u> 2π), <u>and let</u> $K \in L^1$, $\hat{K}(x) = 0$ <u>for</u> $|x| \geq 1$. <u>Then</u> $f*K_n \in \mathcal{T}_{n-1}$.

Note. From here on, Fourier transforms will occur quite often in the analysis. We understand by $\hat{K}$, the 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.

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

We give two proofs of Theorem 18.

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

$$f*K_\lambda = \hat{K}(\tfrac{m}{\lambda})\ e^{imt}$$

If λ is a positive integer n, this belongs to $\mathcal{T}_{n-1}$ since $\hat{K}(\frac{m}{n}) = 0$ 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 $f \in \tilde{C}$, 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 $\mathcal{J}_{n-1}$, hence itself belongs to $\mathcal{J}_{n-1}$ since this set is closed.

SECOND PROOF. If $f \in \tilde{C}$, the same is true of $g = f*K_n$. Therefore, to prove $g \in \mathcal{J}_{n-1}$ 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.

<u>Remarks</u>. If we want $f*K_n \in \mathcal{J}_{n-1}$ for <u>all</u> 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 <u>fixed</u> value of n, the weaker requirement that $\hat{K}(\frac{m}{n}) = 0$ for $|m| \geq n$ suffices. We can also express the conclusion in this form, which will be useful later: if $J \in L^1$, the necessary and sufficient condition that $f \in \tilde{C}$ implies $f*J \in \mathcal{J}_{n-1}$ is that $\hat{J}(x) = 0$ for all <u>integral</u> x such that $|x| \geq n$. We remark also

that the conclusion of Theorem 18 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 $\sum_{r=-\infty}^{\infty} M_r < \infty$, 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.

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

THEOREM 19.

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

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

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

<u>Conclusion</u>: $\|f*K_\lambda - f\| \leq A\lambda^{-r}\omega(\frac{1}{\lambda})$, 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 hypotheses of Theorem 12, $\exists\ T \in \mathcal{T}_{n-1}$ such that

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

where A_r is a constant depending only on r.

<u>Deduction of Corollary from Theorem 19</u>. 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 12. To this end, let $H \in C^{r+3}$ and satisfy

(i) $H(0) = 1,\ H'(0) = \ldots = H^{(r)}(0) = 0$

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

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

<u>Remark</u>. We would 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 $\omega(t) = t$ (see paragraph 4.4 below).

For the proof of Theorem 19 we will require a slightly non-standard estimate for the remainder in Taylor's formula.

LEMMA. Under the hypotheses of Theorem 19, 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

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

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 19

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

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

the lemma $|R(t)| \leq \frac{1}{r!} \left(\frac{|t|}{\lambda}\right)^r \omega\left(\frac{|t|}{\lambda}\right)$. 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 $\left(\frac{|t|}{\lambda}\right) \leq (|t|+1)\omega\left(\frac{1}{\lambda}\right)$ the Theorem is proved.

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 may turn have a certain modulus of continuity.

Another way 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

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

$$\omega_r(t) = \omega_r(f;t) = \sup_{|u| \leq t} \varphi_r(f;u)$$

ω_r is called the rth 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) $\omega_r(t) \leq 2\,\omega_{r-1}(t)$

(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

$\omega_r(tu) \leq (u+1)^r \omega_r(t)$, 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 5.

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 $f \in C(-\infty,\infty)$ and bounded, we define

$$\psi_m(u) = \psi_m(f;u)$$

$$= \sup_t \left| f(t) - \int f(t-uy)dm(y) \right|$$

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

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

$$K(t) = \int p\left(\frac{t}{y}\right) \frac{dm(y)}{y} \qquad (1)$$

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

PROOF. That $K \in L^1$ follows from

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

Write $g = f*K$. Now,

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

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

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

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

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

proving the Lemma.

THEOREM 20. <u>Let</u> $f \in \tilde{C}$, <u>then for every positive integer</u> r <u>one can find</u> $T \in \mathcal{T}_{n-1}$ <u>such that</u>

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

<u>where</u> B_r <u>is a constant depending only on</u> r.

<u>Remark</u> Our proof would not show this, but B_r can even be taken independent of r. See the next section.

<u>PROOF</u>. Let $p_o \in L^1$, and suppose further that $t^r p_o \in L^1$, $\hat{p}_o(x) = 0$ for $|x| \geq 1$, such p_o obviously exist. Write $p(t) = np_o(nt)$, and define K by (1). Finally, let $T = f_* K$. We claim this T fulfills the requirements. 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 n, therefore $T \in \mathcal{J}_{n,1}$ by Theorem 18. Now, by suitable choice of the measure m, $\omega_m(t)$ becomes the modulus of smoothness $\omega_r(t)$, and finally, using the Lemma

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

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

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

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

Remark. Theorem 20 implies Jackson's theorem in the formulation of paragraph 4.2 because of the inequality

$$\omega_{r+1}(f;\tfrac{1}{n}) \leq n^{-r}\,\omega(f^{(r)};\tfrac{1}{n})$$

4.4. Fundamental direct theorem - Third variant.

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

THEOREM 21. Let $f \in C^r$ have period 2π, and $|f^{(r)}(t)| \leq 1$. Then $T \in \mathcal{T}_{n-1}$ can be found such that

$$\|f-T\| \leq A_r n^{-r} \qquad (2)$$

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

Remarks. In 4.4.1 this result shall be further improved upon; It is worth nothing that in Jackson's original formulation, A_r grew exponentially with r. The proof is based on two lemmas.

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 $\sum_{-\infty}^{\infty} c_n e^{int}$, 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

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

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

where A_r depends only on r, and is $\mathcal{O}(r)$. For <u>even</u> 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, for example with a line segment, and taking this as the graph of $\hat{g}$.)

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

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

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

In case r is even, we can complete the graph by drawing tangent lines to the curve $y = x^{-r}$ at $x = \pm 1$. With this choice, $\hat{g}(x)$ is even, and <u>convex</u> on $(0,\infty)$ hence by a known theorem on Fourier integrals (<u>Reference</u>: 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 <u>odd</u> is left to the reader.

<u>Remark</u>. Lemma 2 deals with a special case of a fundamental and important 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 $\hat{p}$, and if so to obtain an estimate for $\|p\|_1$.

PROOF OF THEOREM 21. 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

$$\text{(i)}\ \hat{J}_r(x) = i^{-r}\hat{p}(x),\ |x| \geq n$$

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

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 $|x| \geq n$, $T \in \mathcal{J}_{n-1}$. Moreover,

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

and the theorem is proved.

4.4.2 Despite the satisfactory appearance of the above analysis, the result obtained is far from definitive, and we now wish to outline a procedure for obtaining best possible results. The beautiful theorem of Favard-Akhieser-Krein is

THEOREM 22.

<u>Hypotheses.</u> <u>Same as in Theorem 21.</u>

Conclusion. (2) holds, with $A_r = B_r$, where

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

In particular, $B_1 = \frac{\pi}{2}$, and $B_r \leq \frac{\pi}{2}$, $r = 1,2,\ldots$ The conclusion becomes false if A_r is taken to be any number less than B_r.

Remark. One usually formulates the result under the slightly weaker hypothesis that $f \in C^{r-1}$ and $|f^{(r-1)}(t_2)-f^{(r-1)}(t_1)| \leq |t_2-t_1|$. There is no difference in the proof, the formula $f = f^{(r)} * J_r$ is still valid, where now $f^{(r)}$ is measurable and $|f^{(r)}(t)| \leq 1$ a.e. (alternatively, the use of mollifiers shows there is no essential loss of generality if such theorems are proved even using the additional assumption $f \in C^{\infty}$)

For the proof we shall require several lemmas. The most important is

LEMMA 3. Let us define, for $0 < t < 2\pi$

$$D_r(t) = \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\cos(kt - \frac{r\pi}{2})}{k^r}, \quad r = 1,2,\ldots$$

(Note that we have

$$D_1(t) = \frac{\pi - t}{2\pi}, \quad D_r'(t) = D_{r-1}(t), \quad r = 1,2,\ldots)$$

If r is even [odd], then for any even [odd] $T \in \mathcal{T}_{n-1}$, $D_r - T$ has at most n [at most n-1] zeros, counting multiplicity, on the open interval $(0,\pi)$.

PROOF. The proof, based on the differential identity $D_r' = D_{r-1}$ and an inductive argument, may be supplied by the reader. (Or, see Lorentz, p.118).

LEMMA 4. The following Fourier expansions are valid.

$$\text{sign} \sin t = \frac{4}{\pi} \sum_{m=0}^{\infty} \frac{\sin(2m+1)t}{2m+1}$$

$$\text{sign} \cos t = \frac{4}{\pi} \sum_{m=0}^{\infty} (-1)^m \frac{\cos(2m+1)t}{2m+1}$$

PROOF. Straightforward computation.

LEMMA. Given n [given n-1] distinct points of $(0,\pi)$ we may construct a cosine [sine] polynomial taking prescribed values at those points.

PROOF. Left to the reader (or, look up 'Trigonometric Interpolation' in any standard book, e.g. that of Zygmund).

LEMMA 6. Let r, n be positive integers. There exists a function $p = p_{r,n} \in L^1(-\infty,\infty)$ such that

(i) $\hat{p}(x) = x^{-r}$, x <u>integral</u>, $|x| \geq n$

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

where B_r is the number defined in Theorem 22.

<u>Remark</u>. The similarity with Lemma 2 should be noted. Here, however, we can make p smaller since the interpolation conditions on $\hat{p}$ are only at <u>integral</u> values of x. We remark also, without proof, that B_r is the best possible constant in Lemma 6.

PROOF. We shall look for a function p satisfying the requirements, and vanishing outside $(0,2\pi)$. For such p, $\hat{p}(m)$ is just 2π times the m-th Fourier coefficient, and so we should have, for some $T \in \mathcal{T}_{n-1}$

$$p(t) = T(t) + \frac{1}{2\pi} \sum_{|m| \geq n} m^{-r} e^{imt} \tag{3}$$

That is, any function of the form (3) (and defined to be zero outside $(0,2\pi)$) satisfies (i), and now we wish to choose $T \in \mathcal{T}_{n-1}$ suitably so that $\|p\|_1$ becomes as small as possible. This is a problem of best approximation in the L^1 norm, and

the motivation for the elegant trick which follows lies in the theory of best L^1 approximation (see Lorentz, Chapter 8). We shall complete the proof only for the case r <u>even</u>, the reader may supply the needed modifications for the odd case. For r even, the sum on the right side of (3) is

$$\frac{1}{\pi}\sum_{k=n}^{\infty} k^{-r}\cos kt = (-1)^{\frac{r}{2}} D_r(t) + U$$

for a suitable $U \in \mathcal{T}_{n-1}$, where D_r is as defined in Lemma 3. It is therefore sufficient to produce $T_o \in \mathcal{T}_{n-1}$ such that $\|D_r - T\|_1 \leq B_r n^{-r}$. Now, by Lemma 5, we can construct a cosine polynomial $T_o \in \mathcal{T}_{n-1}$ which takes at the points $(m+\frac{1}{2})\frac{\pi}{n}$, $m = 0,1,\ldots n-1$ the same values as D_r. Then, by Lemma 3, $D_r - T_o$ vanishes nowhere on $(0,\pi)$ except in those points, where it must have <u>simple</u> zeros, consequently it changes sign at those points. Therefore $(D_r - T_o)\cos nt$ does not change sign on $(0,\pi)$ and since a cosine polynomial is symmetric about $t = \pi$ it does not change sign on $(0,2\pi)$. Therefore $\|D_r - T_o\|_1$ is the absolute value of

$$\int_0^{2\pi} [D_r(t) - T_o(t)] \operatorname{sign} \cos nt \, dt$$

and since, by Lemma 4

$$\operatorname{sign} \cos nt = \frac{4}{\pi}\sum_{m=0}^{\infty} (-1)^m \frac{\cos(2m+1)nt}{2m+1}$$

the last integral is simply

$$\int_0^{2\pi} D_r(t)\,\mathrm{sign}\cos nt\,dt$$

which we can evaluate by Parseval's formula, obtaining the value

$$(-1)^{\frac{r}{2}}\frac{4}{\pi}\left[n^{-r} - \frac{1}{3}(3n)^{-r} + \frac{1}{5}(5n)^{-r} - \dots\right]$$

$$= (-1)^{\frac{r}{2}}\frac{4}{\pi}n^{-r}\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m+1)^{r+1}} = (-1)^{\frac{r}{2}}B_r$$

and the Lemma is proved.

PROOF OF THEOREM 22. First let us show that (2) holds with $A_r = B_r$. We can repeat the proof of Theorem 21 almost word for word, except that we use for p the $p_{r,n}$ constructed in Lemma 6. J_r is taken to be any L^1 function such that $\hat{J}_r(x) = i^{-r}\hat{p}(x)$ for <u>integral</u> x, $|x| \geq n$. Then $J - i^{-r}p$ has a Fourier transform which vanishes for integral $x, |x| \geq n$ and this guarantees that its convolution with $f^{(r)}$ is in $\mathfrak{T}_{n-1}$. The rest of the proof is unchanged.

It remains to show that the constant B_r is best possible. We will confine ourselves to the case r even. Let

$$f(t) = \frac{4}{\pi}\sum_{m=0}^{\infty}(-1)^m\frac{\cos(2m+1)nt}{(2m+1)^{r+1}n^r}$$

then $f(t) = (-1)^{\frac{r}{2}}$ sign cos nt, which is bounded by one. Moreover, if k is an integer

$$f\left(\frac{k\pi}{n}\right) = (-1)^k B_r n^{-r}$$

Thus, f takes alternately the values $\pm B_r n^{-r}$ at the points $\frac{k\pi}{n}$ $(k=0,1,2,\ldots)$. Invoking a classical theorem of Chebyshev (we must refer to Lorentz for details) it follows that $\|f-T\| \geq B_r n^{-r}$ for all $T \in \mathcal{T}_{n-1}$.

Remark. The last paragraph constitutes one kind of converse to Jackson's theorem. Another kind of converse will be taken up in Chapter 5.

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*

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

where B_r is as in Theorem 22.

PROOF. Preserving the notation of the preceding proof, we have

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

Now,

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

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

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

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

4.5 Localisation 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

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

where

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

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

<u>Example</u>. Let K = Cauchy kernel, and $f \in \mathrm{Lip}\,\alpha$ $(\alpha < 1)$ on $(-1,1)$. Then f can be modified outside of $(-1,1)$, so that the modified function f^* is in $\mathrm{Lip}\,\alpha$ on $(-\infty,\infty)$, and so $|f^*(t;\lambda) - f^*(t)| \leq C\lambda^{-\alpha}$. Moreover, since here $\varphi(u) = \mathcal{O}\left(\frac{1}{u}\right)$, we have for $|t| < 1-\delta$

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

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

CHAPTER 5

INVERSE THEOREMS

5.1 Let us, in the first instance, talk about trigonometric polynomial approximation. Already in the last Chapter (section 4.4.3) we showed that Jackson's theorem is, in a sense, best possible. Another kind of result in the reverse direction is the famous theorem of S.Bernstein, that if $f \in \tilde{C}$ is approximable to the order $n^{-\alpha}$ $(0<\alpha<1)$ then $f \in \mathrm{Lip}\,\alpha$. 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 comprises such properties as Lipschitz smoothness, absolute continuity, etc. In the present chapter we discuss some of these generalisations, 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; the various generalisations are proved by a similar technique.

We require first

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

A <u>weak form</u> 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. Let $K \in L^1$ such that $\hat{K}(m)=m$ for $m=0, \pm 1, \pm 2, \ldots \pm n$.

Let $T \in \mathcal{T}_n$, so that $T(x)=\sum_{k=-n}^{n} c_k e^{ikx}$, and $\|T\|=1$.

Then

$$i \int_{-\infty}^{\infty} T(x)K(x)dx = i \sum_{m=-n}^{n} c_m \int_{-\infty}^{\infty} K(x)e^{imx}dx$$

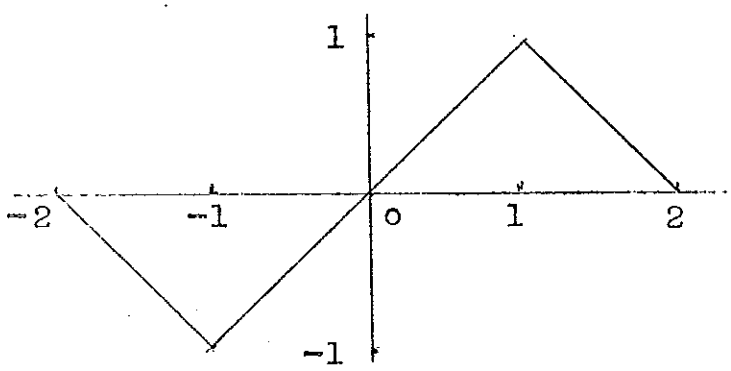
$$= \sum_{m=-n}^{n} -(im)c_m = -T'(0).$$

It is sufficient to show that $|T'(0)| \leq An$. (For, the class of trigonometric polynomials of norm one and degree not exceeding n is invariant under translation).

Now

$$|T'(0)| \leq \int_{-\infty}^{\infty} |K(x)|dx.$$

Concretely, we choose $\hat{K}(x)=nJ\left(\frac{x}{n}\right)$ where $J(x)$ is given by the graph:



$J(x)$ is zero for $|x|\geq 2$. J is an L^1 Fourier transform, say $J=\hat{j}$. Note that $\hat{K}(x)=x$ for $|x|\leq n$. Moreover

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

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} nJ\left(\frac{x}{n}\right) e^{itx}dx$$

$$= n^2 j(nt), \text{ therefore}$$

$$|T'(0)|\leq n^2\int_{-\infty}^{\infty}|j(nt)|dt=An$$

with $A=\|j\|_1$. This proves the Theorem.

<u>Remark 1</u>. In the above proof, A is easily seen to be $\frac{4}{\pi}$. To try to get the correct constant unity, one is led to another case of the general problem remarked in the last chapter, namely minimize $\|f\|$ over all $f\in L^1$ with $\hat{f}(x)=x$ for $|x|\leq 1$.

<u>Remark 2</u>. A (perhaps simpler) variant of the above proof is as follows. Let $K\in L^1$ and $\hat{K}(x)=1$ for $|x|\leq 1$. Moreover, suppose $\hat{K}$ is smooth enough so that $\frac{dK}{dt}$ is integrable. Then, one verifies easily the identity $T=T*K_n$, i.e.

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

whenever $T\in\mathcal{J}_n$. Hence, for $T\in\mathcal{J}_n$

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

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

NOTATION. For any $f \in \tilde{C}$, we define (the "degree of approximation"):

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

THEOREM 26. (Bernstein). <u>Let</u> $0<\alpha<1$. <u>If</u> $E_n(f)= \mathcal{O}(n^{-\alpha})$, <u>then</u> $f \in \mathrm{Lip}\,\alpha$.

<u>Remark 1</u>. The Theorem is not true for $\alpha=1$.

<u>Remark 2</u>. To appreciate Bernstein's argument, it is worthwhile to see first what a straightforward approach yields. If $\|f-T_n\|\leq An^{-\alpha}$, the T_n are uniformly bounded, say $\|T_n\|\leq B$. Then, by Bernstein's inequality (weak form), $\|T'_n\|\leq Cn$.

Therefore

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

(if we choose n around $h^{-\frac{1}{1+\alpha}}$). This shows $f \in \mathrm{Lip}\,\frac{\alpha}{1+\alpha}$ rather than $\mathrm{Lip}\,\alpha$. The reason why we obtain an imperfect result is that the estimate $T'_n(x)= \mathcal{O}(n)$ is too crude. The extra information that $\{T_n\}$ <u>converges with a certain speed</u> must be exploited to produce a better estimate for T'_n. If we

can show that in fact $|T_n'(x)| \leq Cn^{1-\alpha}$, then the sharp result will follow.

PROOF OF BERNSTEIN'S THEOREM. By hypotheses

$$|T_n(x)-T_{2n}(x)| \leq |T_n(x)-f(x)|+|f(x)-T_{2n}(x)|$$

$$\leq 2An^{-\alpha} .$$

By Bernstein's inequality, this gives

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

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

$$|T'_{2^k}(x) - T'_{2^{k-1}}(x)| \leq B(2^k)^{\beta}.$$

Now changing k to k-1, k-2,... we obtain a system of inequalities

$$|T'_{2^{k-1}}(x) - T'_{2^{k-2}}(x)| \leq B(2^{k-1})^{\beta}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$|T'_2(x) - T'_1(x)| \leq B2^{\beta}.$$

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

$$|T'_{2^k}(x) - T'_1(x)| \leq Bn^{\beta} \left[1+ \frac{1}{2^{\beta}}+ \ldots +\right]$$

$$\leq Cn^{\beta} .$$

Therefore

$$|T_n'(x)| \leq C_1 n^{1-\alpha} \text{ if } n=2^k,\ k=1,2,\ldots \tag{1}$$

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

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

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

$$|f(x+h)-f(x)| \leq C_2\left(\left(\frac{2}{h}\right)^{1-\alpha}h+h^{\alpha}\right) = C_3h^{\alpha}.$$

This proves the theorem.

<u>Remark 3</u>. Case $\alpha=1$. 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 $\frac{1}{2h} \leq n < \frac{1}{h}$, then we get

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

Thus $E_n(f) = \mathcal{O}(\frac{1}{n})$ implies f has a modulus of continuity which is $\mathcal{O}(h \log \frac{1}{h})$. This result is unimprovable, as examples show, but unsatisfying because the converse is not true; a modulus of continuity $\mathcal{O}(h \log \frac{1}{h})$ does <u>not</u> guarantee

$E_n(f)=\mathcal{O}(\frac{1}{n})$, but only $E_n(f) = \mathcal{O}\left(\frac{\log n}{n}\right)$.

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). <u>A necessary and sufficient condition that</u> $E_n(f) = \mathcal{O}(\frac{1}{n})$ <u>is</u> $f \in Z$.

PROOF. The sufficiency is a consequence of Theorem 20, with r=2.

<u>Necessity</u>. If $E_n(f)\leq \frac{A}{n}$, then

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

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

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

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),

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

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

Similarly we have

THEOREM 28. <u>If</u> $E_n(f) = \mathcal{O}\left(n^{-(1+\alpha)}\right)$, $0<\alpha<1$ <u>then</u> f <u>has a derivative</u> $f' \in \mathrm{Lip}\,\alpha$. <u>Equivalently</u>, <u>if</u> $E_n(f) = \mathcal{O}\left(n^{-(1+\alpha)}\right)$, <u>then</u>

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

PROOF. As in the proof of Zygmund's theorem one shows, if $\|f-T_n\| \leq \frac{A}{n^{1+\alpha}}$, that

$$|T_n'(x+h)-T_n'(x)| \leq Ch^{\alpha}$$

and the result follows easily from this. The equivalence of the formulation in terms of the second difference is left as an exercise for the reader.

We remark that an analogous result holds if $E_n(f) = \mathcal{O}\left(n^{-(r+\alpha)}\right)$ where r is a positive integer and $0<\alpha<1$. For details, see the books of Lorentz, Natanson or

Timan. This result also follows from the general theorem proved in the next Section.

THEOREM 29. <u>If</u> $\sum_{n=1}^{\infty} E_n(f) < \infty$, <u>then</u> $f \in \mathrm{Lip}\ 1$.

PROOF. It is sufficient to prove that if $|f(x)-T_n(x)| \leq \delta_n$ where $\delta_n \searrow$ and $\sum \delta_n < \infty$, 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

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

Put $n=2^k$. 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,... etc and then adding, by means of the triangle inequality we obtain

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

by the following lemma of Cauchy:

LEMMA. <u>If</u> $a_n \searrow (a_n>0)$ <u>and</u> $\sum a_n < \infty$, <u>then</u> $\sum 2^n a_{2^n} < \infty$.

As our final illustration of an "inverse theorem", we have

THEOREM 30. (Weiss-Zygmund, 1959). <u>If</u> $E_n(f) = \mathcal{O}\left(\frac{1}{n(\log n)^{\frac{1}{2}+\varepsilon}}\right)$ <u>for some</u> $\varepsilon > 0$ <u>then</u> f <u>is absolutely continuous and</u> $f' \in L^p$ <u>for all</u> $p < \infty$.

We shall not prove this theorem; the case $p=2$, easily handled by Fourier series methods, should be a pleasant exercise for the reader.

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

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

where $a=\frac{1}{\lambda}$, and σ is the measure defined by

$$d\sigma(u) = d\,\delta(u)-K(u)du. \tag{2}$$

Here δ is the "Diracimeasure" (unit point mass at $u=0$). Thus, the error of approximation is expressible as the integral (1), where σ is a finite measure on the line and $\int d\sigma = 0$, and we are interested in the behaviour 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 (2). 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

$$\varphi_\sigma(f;a) = \sup_t\left|\int f(t-au)d\sigma(u)\right| \tag{3}$$

the statement $\varphi_\sigma(f;a) = \mathcal{O}(a^\alpha)$ expresses that $f \in \mathrm{Lip}\,\alpha$; whereas, if σ is defined by (2), the corresponding statement expresses that $\|f-(f*K_\lambda)\| = \mathcal{O}(\lambda^{-\alpha})$. Thus, both "smoothness" and "degree of approximation" assertions are expressible

in the form (3) with suitable measures σ . From this point of view, there is no essential difference between a "direct" and an "inverse" theorem of approximation theory. Both kinds of theorems are merely comparison theorems relating the behaviour of $\varphi_{\sigma_1}(f;\alpha)$ to that of $\varphi_{\sigma_2}(f;\alpha)$ for a pair of measures σ_1, σ_2 . This will be our point of view throughout the remainder of these notes. The sophisticated reader has perhaps noticed that, in this formulation, the theorems of approximation theory look very like Tauberian theorems. This insight is quite correct, and in fact we shall make use of "Tauberian" ideas.

5.2.1 <u>Fourier-Stieltjes transforms</u>. 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:

THEOREM 31. <u>Let</u> $S_1, S_2 \in W$, <u>and</u> E <u>a compact set of reals</u>. <u>If</u> S_1 <u>vanishes nowhere in</u> E, <u>and</u> S_2 <u>vanishes everwhere outside</u> E, <u>then</u> S_2 <u>divides</u> S_1 <u>i.e.</u> $S_2 = S_1S_3$ <u>for some</u> $S_3 \in W$.

The reader may refer, e.g. to the book "Tauberian Theorems" by H.R.Pitt.

By W_o we denote the subring (actually an ideal) of W consisting of Fourier (-Lebesgue) transforms, i.e. those σ for which σ is absolutely continuous, hence $d\sigma(t)=K(t)dt$, with $K \in L^1$.

5.2.2 We begin with a very simple "comparison theorem", based on <u>global</u> divisibility of the Fourier transforms.

THEOREM 32. <u>If</u> $\hat{\sigma}_1$ <u>divides</u> $\hat{\sigma}_2$, <u>say</u> $\hat{\sigma}_2 = \hat{\sigma}_1 S$ <u>then for every bounded</u> $f \in C(-\infty, \infty)$

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

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

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

Hence,

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

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

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

Since the right side is independent of t, the result follows.

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

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

for suitable constants A_i (depending only on σ and the σ_i). As an evident consequence of Theorem 32, we have

THEOREM 33. Let J, K be L^1 kernels, and suppose $1-\hat{J}$ divides $1-\hat{K}$ in W. Then for every bounded continuous f

$$\|f-(f*K_\lambda)\| \leq A\|f-(f*J_\lambda)\| \tag{4}$$

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

DEFINITION. When(4) 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

THEOREM 34. 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 $\dfrac{1-\hat{K}}{1-\hat{J}} = 1+S$ is in W, and the result follows from Theorem 33.

COROLLARY. The Fejér 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,

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

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 its inverse Fourier transform is $\mathcal{O}(t^{-2})$ 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 again an L^1 transform. (It is, moreover, not hard to estimate the norms of G,H).

Thus, the Fejér and Cauchy kernels are equivalent from the standpoint of degree of approximation. The fact, already noted, that they have the same saturation order does not imply a degree of approximation λ^{-1} by the Fejér kernel, nor vice versa. For, the Fourier transform of the dipole measure is $2i \sin x$, and neither of the ratios $\frac{\sin x}{|x|}$, $\frac{|x|}{\sin x}$ is continuous at $x=0$ hence <u>a fortiori</u> neither can coincide in any neighbourhood of zero with an element of W.

In like manner, it is not surprising that $f \in \mathrm{Lip}\ 1$ and class, is a reflection of this.

5.2.3 It is natural to ask whether a conclusion similar to that of Theorem 32 can be drawn in case $\hat{\sigma}_1$, only divides $\hat{\sigma}_2$ <u>locally</u> (at $x=0$). More precisely, we say a function $G(x)$ belongs to W <u>at the point</u> x_0 if there is an $S \in W$ such that $S(x)=G(x)$ in a neighbourhood of x_0. If now, $\frac{\hat{\sigma}_2}{\hat{\sigma}_1}$ belongs to W at $x=0$, can we assert that (writing φ_1 for φ_{σ_1})

$$\varphi_2(f;a) \leq A\, \varphi_1(f;a) \tag{5}$$

where A depends only on σ_1 and σ_2 ? In general we <u>cannot</u>, that is, the <u>global</u> nature of σ, not merely its behaviour at $x=0$, influence the order of magnitude of $\varphi_\sigma(f;a)$. Nevertheless, a somewhat weaker conclusion than (5), which implies (5) in many cases of interest, can be drawn.

THEOREM 35. <u>Let</u> σ_1 <u>be a real finite measure, and suppose</u> $\hat{\sigma}_1$ <u>divides</u> $\hat{\sigma}_2$ <u>at</u> $x=0$. <u>Suppose for some</u> $x_0>0$ <u>and</u> $b<1$ $\hat{\sigma}_1$ <u>does not vanish on the interval</u> $[b^2x_0, x_0]$. <u>Then for every bounded</u> $f \in C(-\infty,\infty)$

$$\varphi_{\sigma_2}(f;a) \leq C \sum_{i=0}^{\infty} \varphi_{\sigma_1}(f;Bb^i a) \tag{6}$$

<u>where</u> B,C <u>are constants depending only on</u> σ_1 <u>and</u> σ_2.

PROOF. Let $P(x)$ be the function which equals one for $|x|\leq b^2x_0$, 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

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

Now, $Q(x)=0$ except for x in the intervals (b^2x_0, x_0) and $(-x_0, -b^2x_0)$, and on these intervals $\hat{\sigma}_1$ does not vanish by hypothesis (note that σ_1 real implies $\hat{\sigma}_1(-x)=\hat{\sigma}_1(x)$). Therefore, by Theorem 31, $\hat{\sigma}_1$ divides Q, say $\hat{\sigma}_1 S=Q$ for some $S \in W$, and therefore by Theorem 32, we have for all t

$$\left|\int f(t-au)\left[p(u)-\frac{1}{b}\,p\left(\frac{u}{b}\right)\right]du\right| \leq A\,\varphi_{\sigma_1}(f;a) \tag{7}$$

where A depends only on σ_1. Let us write $\varphi_1(a)$ as an abbreviation for $\varphi_{\sigma_1}(f;a)$. Replacing a by ba in (7) and

making the change of variable bu=v gives

$$\left| \int f(t-av) \left[\frac{1}{b} p\left(\frac{v}{b}\right) - \frac{1}{b^2} p\left(\frac{v}{b^2}\right) \right] dv \right| \leq A \varphi_1 (ba)$$

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

$$\left| \int f(t-au) \left[p(u) - \frac{1}{b^m} p\left(\frac{u}{b^m}\right) \right] du \right| \leq A \sum_{i=0}^{m-1} \varphi_1 (b^i a)$$

$$\leq A \sum_{i=0}^{\infty} \varphi_1 (b^i a).$$

Now, for fixed a,t

$$\lim_{\lambda \to \infty} \int f(t-au) \lambda p(\lambda u) du$$

$$= \lim_{\lambda \to \infty} \int f(t-v) \frac{\lambda}{a} p\left(\frac{\lambda v}{a}\right) dv = f(t)$$

therefore, letting $m \to \infty$ in the last inequality,

$$|f(t) - \int f(t-au)p(u)du| \leq A \sum_{i=0}^{\infty} \varphi_1 (b^i a) \tag{8}$$

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, $P\left(\frac{x_0 x}{c}\right)$ vanishes for $|x| \geq c$, hence

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

holds for all x. This shows that $\hat{\sigma}_2$ belongs to the ideal

generated by $\hat{\sigma}_1$ and $1-P\left(\frac{x_o x}{c}\right)$. Therefore, by the Remark following Theorem 32,

$$\varphi_{\sigma_2}(f;a) \leq A_1 \varphi_{\sigma_1}(f;a) + A_2 \varphi_{\sigma}(f;a) \tag{9}$$

where $d\sigma(t) = d\delta(t) - \frac{c}{x_o} p\left(\frac{ct}{x_o}\right) dt$ (so that $\hat{\sigma}(x) = 1 - P\left(\frac{x_o x}{c}\right)$).

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,

$$\begin{aligned} \left|\int f(t-au)d\sigma(u)\right| &= \left|f(t) - \int f(t-au) \cdot \frac{c}{x_o} p\left(\frac{cu}{x_o}\right) du\right| \\ &= \left|f(t) - \int f(t-Bav)p(v)dv\right| \\ &\leq A \sum_{i=0}^{\infty} \varphi_1(b^i Ba) \end{aligned}$$

by (8), where $B = \frac{x_o}{c}$ depends only on σ_1 and σ_2. Therefore, the last expression on the right is an upper bound for $\varphi_\sigma(f;a)$ and so, from (9)

$$\varphi_{\sigma_2}(f;a) \leq A_1 \varphi_1(a) + A_2 A \sum_{i=0}^{\infty} \varphi_1(b^i Ba)$$

where, we recall, $\varphi_1(a)$ denotes $\varphi_{\sigma_1}(f;a)$. This proves the theorem.

<u>Remark</u>. 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.

COROLLARY. If σ_1 is any non-null measure and $\hat{\sigma}_1$ divides $\hat{\sigma}_2$ at $x \neq 0$, then for every $\theta > 0$, if $\varphi_\sigma(f;a) = \mathcal{O}(a^\theta)$, then $\varphi_\sigma(f;a) = \mathcal{O}(a^\theta)$ (as $a \to 0+$).

Theorem 35, in conjunction with the remarks at the beginning of Chapter 4, can easily be adapted to give yet another proof of Jackson's theorem - we leave the details to the interested reader.

5.2.4 We now turn attention to a deeper and harder question: what assertion can we make about φ_{σ_2} if the asymptotic behaviour of φ_{σ_1} is known, <u>when $\hat{\sigma}_1$ does not divide $\hat{\sigma}_2$; even locally</u>? Surprisingly, fairly strong assertions can be made in many such cases. This question comprises, from a general point of view, the "inverse problem" of approximation theory, as well as other problems of a familiar nature, for instance to bound the modulus of continuity of a function in terms of its (second order) modulus of smoothness.

The technique we shall use in this section is a direct adaptation of Bernstein's idea used in the proof of Theorem 26. We need first a generalisation of Bernstein's inequality to "low frequency functions" (by this we mean here the convolution of a bounded function with an L^1 function whose Fourier transform has compact support).

THEOREM 36. (Bernstein inequality for the derivative of a low frequency function). <u>Let</u> f <u>be bounded and measurable on</u> $(-\infty, \infty)$, $K \in L^1$, K <u>absolutely continuous with an</u> L^1 <u>derivative, and</u> $\hat{K}(x)=0$ <u>for</u> $|x| \geq \Lambda$. <u>Let</u> $g=f*K$, <u>then</u>

$\|g'\| \leq B\Lambda\|g\|$, <u>where</u> B <u>is an absolute constant</u>. (the norms are sup norms).

PROOF. The proof is modeled on our second proof of Theorem 25. Let $p \in L^1$ be so chosen that $\hat{p}(x)=1$ for $|x| \leq 1$. Then, for all x,

$$\hat{K}(x)\hat{p}\left(\frac{x}{\Lambda}\right) = \hat{K}(x) ,$$

hence

$$K(t) = \int K(v)\Lambda p\left(\Lambda(t-v)\right) dv$$

therefore, differentiating,

$$K'(t) = \int K(v)\Lambda^2 p'(\Lambda(t-v))dv$$

or,

$$K'(t) = \int K(t-v)\Lambda^2 p'(\Lambda v)dv \tag{10}$$

Now,

$$g(t) = \int f(u)K(t-u)du$$

$$g'(t) = \int f(u)K'(t-u)du$$

$$= \int f(u)\left[\int K(t-u-v)\Lambda^2 p'(\Lambda v)dv\right]du$$

(using (10))

$$= \int g(t-v)\Lambda^2 p'(\Lambda v)dv$$

$$|g'(t)| \leq \|g\| \cdot A \int |p'(u)|\, du$$

which proves the theorem (with $B=\|p'\|_1$; as we have already remarked, we could take $B=\frac{4}{\pi}$).

LEMMA. Let K be integrable and r times differentiable, with $K^{(r-1)}$ absolutely continuous and $K^{(r)}$ integrable. Suppose $\hat{K}(x)=0$ for $|x|\geq 1$. Write $f^{\lambda}=f*K_{\lambda}$ and $\psi(\lambda)=\|f-f^{\lambda}\|$. Then, for $\lambda=2^m$, $m=1,2,\ldots$

$$\left\| \frac{d^r f^{\lambda}}{dt^r} \right\| \leq C_1 \sum_{j=0}^{m} 2^{(j+1)r} \psi(2^j) + \|K^{(r)}\|_1 \cdot \|f\|$$

where C_1 is an absolute constant.

PROOF. For all λ, by the triangle inequality

$$\|f^{\lambda}-f^{2\lambda}\| \leq \psi(\lambda)+\psi(2\lambda).$$

Now, $f^{\lambda}-f^{2\lambda} = f*J$ where

$$J(t) = K_{\lambda}(t)-K_{2\lambda}(t)$$

$$\hat{J}(x) = \hat{K}\left(\frac{x}{\lambda}\right) - \hat{K}\left(\frac{x}{2\lambda}\right)$$

which vanishes for $|x|\geq 2\lambda$, therefore by Theorem 36 (applied r times)

$$\left\| \frac{d^r}{dt^r} (f^{\lambda}-f^{2\lambda}) \right\| \leq B(2\lambda)^r \left[\psi(\lambda)+\psi(2\lambda)\right]$$

writing this inequality for $\lambda = 1, 2, \ldots 2^{m-1}$ and adding gives

$$\left\| \frac{d^r f^{2^m}}{dt^r} - \frac{d^r f^1}{dt^r} \right\| \leq B2^r \sum_{j=0}^{m-1} 2^{jr} \left[\psi(2^j) + \psi(2^{j+1}) \right]$$

$$\leq B2^r \sum_{j=0}^{m} 2^{jr+1} \psi(2^j)$$

which implies the stated result, with $C_1 = 2B$, since

$$\left\| \frac{d^r f^1}{dt^r} \right\| \leq \|K^{(r)}\|_1 . \|f\|.$$

<u>Remark</u>. Clearly, by the same reasoning we can prove, for $\lambda = A^m$, where $A > 1$

$$\left\| \frac{d^r f^\lambda}{dt^r} \right\| \leq C_1(A) \sum_{j=0}^{m} A^{(j+1)r} \psi(A^j) + C_2 \|K^{(r)}\|_1 . \|f\|.$$

Now it is easy to prove the important

LEMMA. Let P be a C^∞ function such that $P(x) = 1$ for $|x| \leq \frac{1}{2}$ and $P(x) = 0$ for $|x| \geq 1$, and write $P = \hat{p}$ (where p has integrable derivatives of all orders). For any finite measure σ on the line and any positive integers r, m

$$\left\| \frac{d^r (f * p_\lambda)}{dt^r} \right\| \leq C + A_1 \sum_{j=0}^{m} \sum_{j=0}^{\infty} b^{-(j+1)r} \varphi_\sigma(f; Bb^{i+j}).$$

Here $b (0 < b < 1)$ depends only on σ and $\lambda = b^{-1}$. A_1, B, C are independent of λ and r.

PROOF. Let $d\sigma_2 = d\delta - p\,dt$, then $\hat{\sigma}_2(x) = 1 - P(x)$ vanishes for $|x| \leq \frac{1}{2}$, hence $\hat{\hat{\sigma}}(x)$ divides $\hat{\sigma}_2(x)$ at $x=0$, and we can apply Theorem 35, getting (since in the present case $\varphi_{\sigma_2}(f;a) = \|f-(f*p_\lambda)\|$ with $\lambda = \frac{1}{a}$:

$$\|f-(f*p_\lambda)\| \leq C \sum_{i=0}^{\infty} \varphi_\sigma(f; \frac{Bh^i}{\lambda}) . \tag{11}$$

Applying the preceding Lemma (in the form enunciated in the Remark) with $h = \frac{1}{b}$, and taking for ψ the right side of (11), now yields the result.

5.3 <u>Inverse theorems (Bernstein type).</u> We are now in a position to deduce far-reaching generalisations of the classical inverse theorems surveyed in 5.1. First, however, let us derive from the last lemma a useful estimate. We have

LEMMA. Under the hypotheses of the preceding Lemma, if $\varphi_\sigma(f;u) \leq B_1 \varphi(u)$, where φ is some function such that $\varphi(b) < 1$ and $\varphi(uv) \leq \varphi(u)\varphi(v)$ for all u,v, then

$$\left\| \frac{d^r(f*p_\lambda)}{dt^r} \right\| \leq C + A_2 \sum_{j=0}^{m} b^{-jr} \varphi(b^j) \tag{12}$$

where $\lambda = b^{-m}$. Here C, A_2 are independent of λ .

PROOF. We have, by hypothesis

$$\varphi_\sigma(f; Bb^{i+j}) \leq \varphi(b^j)\varphi(B)\varphi(b^i)$$

which implies the stated conclusion.

THEOREM 37. <u>Let</u> f <u>be continuous and bounded on</u> $(-\infty, \infty)$ <u>and suppose for some non-null measure</u> $\int f(t-au)d\sigma(u) = \mathcal{O}(a^{\alpha})$ <u>as</u> $a \to 0+$. <u>Then</u>

$$\omega(f;u) = \mathcal{O}(u^{\alpha}) \quad \text{if } 0<\alpha<1$$

$$\omega_2(f;u) = \mathcal{O}(u) \quad \text{if } \alpha=1.$$

COROLLARY. Let f be continuous and of period 2π and approximable by trigonometric polynomials to the order $n^{-\alpha}$. Then the conclusion of Theorem 37 holds.

First, let us see that one can deduce the Corollary from Theorem 37. This is not altogether obvious since the Corollary is an inverse theorem in its classic form, in which it is not given that the approximating polynomials are generated by some fixed kernel. Suppose then,

$$\|f-T_n\| \leq An^{-\alpha}, \quad T_n \in \mathcal{T}_n$$

Let σ be a non-null measure such that $\hat{\sigma}(x)=0$ for $|x|\leq 1$. Then, if $a>0$

$$\left|\int f(t-au)d\sigma(u)\right| \leq An^{-\alpha}\int |d\sigma| + \left|\int T_n(t-au)d\sigma(u)\right|.$$

Now, since $T_n \in \mathcal{T}_n$, the last integral vanishes for $n \leq \frac{1}{a}$.

Therefore, choosing n as the greatest integer not exceeding $\frac{1}{a}$, the right side is bounded by $A\left(\frac{a}{1-a}\right)^{\alpha}\int |d\sigma| = \mathcal{O}(a^{\alpha})$,

and the result follows by Theorem 37.

PROOF OF THEOREM 37. We treat first the case $\alpha<1$. By (12), with $r=1$ and $\varphi(u)=u^{\alpha}$, we get for $\lambda=b^{-m}$

$$\left\| \frac{d(f*p_\lambda)}{dt} \right\| \leq C+A_2 \sum_{j=0}^{m} b^{-jr} b^{\alpha j} \leq A_3 \lambda^{1-\alpha} .$$

Moreover, by Theorem 35 with $\sigma_1 = \sigma$, $d\sigma_2 = d\delta - pdt$ and $a=\frac{1}{\lambda}$ we have

$$\|f-(f*p_\lambda)\| \leq A_4 \sum_{i=0}^{\infty} \left(\frac{Bb^i}{\lambda}\right)^{\alpha} = A_5 \lambda^{-\alpha} \quad , \text{ for all real } \lambda .$$

Hence, writing $f^\lambda = f*p^\lambda$ we have, for any t real and $h>0$, and any positive integer m

$$|f(t+h)-f(t)| \leq 2\|f-f^\lambda\| + |f^\lambda(t+h)-f^\lambda(t)|$$

$$\leq 2A_5 \lambda^{-\alpha} + A_3 \lambda^{1-\alpha} h$$

where $\lambda = b^{-m}$. Choosing now m so that

$$b^{-m-1} \leq h < b^{-m}$$

the right side is $\leq 2A_5(bh)^{\alpha} + A_3 h^{\alpha-1}.h = A_6 h$ and we are done. To handle the case $\alpha=1$, we apply the estimate (12) with $r=2$, $\varphi(u)=u^{-1}$ and get

$$\left\| \frac{d^2 f}{dt^2} \right\| \leq C+A_2 \sum_{j=0}^{m} b^{-j} \leq A_7 \lambda$$

for $\lambda = b^{-m}$. Now,

$$|f(t+h)-2f(t)+f(t-h)| \leq 4\|f-f^\lambda\| + |f^\lambda(t+h)-2f^\lambda(t)+f^\lambda(t-h)|$$

The last term on the right does not exceed

$$h^2\left\|\frac{d^2f^\lambda}{dt^2}\right\| \leq A_7\lambda h^2.$$

Therefore, for $\lambda = b^{-m}$

$$|f(t+h)-2f(t)+f(t-h)| \leq A_8\left(\frac{1}{\lambda}+\lambda h^2\right) \leq A_9 h$$

for a suitable choice of m.

Of course, all we have done here is to imitate the proofs of more special theorems given in 5.1. In a similar manner we can show (where as usual ω_r denotes the modulus of smoothness of order r):

THEOREM 38. <u>Let</u> f <u>be continuous and bounded</u>, <u>and suppose for some non-null measure</u> σ,

$$\left|\int f(t-au)\,d\sigma(u)\right| \leq C\varphi(a)$$

<u>where</u> $\varphi(u)$ <u>is increasing for</u> $u>0$, $\varphi(1)=1$ <u>and</u> $\varphi(uv)\leq\varphi(u)\varphi(v)$. <u>Then, for all positive integers</u> r,m <u>and every</u> $h>0$ <u>we have</u>

$$\omega_r(h) \leq C_1\left[\varphi(b^m) + h^r\sum_{j=0}^{m} b^{-jr}\varphi(h^j)\right] \tag{13}$$

<u>Here</u> b <u>is a positive constant less than one which depends only on</u> σ.

In applying this theorem, we of course seek to minimize the right side by suitable choice of m.

Example. Let σ be the "dipole measure", and $\varphi(u)=u$. Then, the hypothesis is equivalent to saying $\omega_2(h)=\mathcal{O}(h)$. Taking r=1 in (13), we get

$$\omega(h)\leq C_1\left(b^m+h\sum_{j=0}^{m}1\right)=C_1(b^m+mh)$$

and this is $\leq C_2\, h \log\frac{1}{h}$ if m is chosen suitably.

5.4 Inverse Theorems. (Zygmund type). We have observed, in proving inverse theorems of the Bernstein type (that is, "such a degree of approximation implies such smoothness") that the crucial role was played by estimates for the derivatives of the approximating functions $f*K_\lambda$. It is a remarkable fact that such estimates on the derivatives, even without a priori information on the degree of approximation, implies smoothness. To our knowledge, the only published theorem of this kind is one due to A. Zygmund (Duke Math. J. 12 (1945) pp. 47-76). He showed:

THEOREM 39. Let $u(re^{it})$ be harmonic in the disk $r<1$ and satisfy

$$\left|\frac{\partial^2 u(re^{it})}{\partial t^2}\right|\leq\frac{C}{1-r} \tag{14}$$

Then, u is uniformly continuous in $r<1$ and its boundary function $f(t)=u(e^{it})$ satisfies

$$\omega_2(f;h)=\mathcal{O}(h). \tag{15}$$

The proof that u is extendable continuously to the closed disk is function-theoretic, and rather easy. The main point is that the boundary function satisfies (15). This fact is a consequence of

THEOREM 40. *Let* $U(x+iy)$ *be continuous for* $y \geq 0$ *and harmonic for* $y>0$, *and satisfy*

$$\left| \frac{\partial^2 U(x+iy)}{\partial x^2} \right| \leq \frac{C}{y} \tag{16}$$

Then $f(x)=U(x)$ *satisfies* (15).

Indeed, if u is harmonic in the disk, $U(z)=u(e^{iz})$ is harmonic in the upper half-plane, and if u satisfies (14), U satisfies (16). Theorem 40 is, in turn, the case $K(t)=\frac{1}{\pi}\cdot\frac{1}{1+t^2}$ of the following proposition:

THEOREM 41. *Let* f *be bounded and continuous on* $(-\infty, \infty)$ *and* K *an* L^1 *function, not identically zero, with* $K \in C^2$ *and* $K'' * L^1$. *If*

$$\left\| \frac{d^2(f * K_\lambda)}{dt^2} \right\| \leq C\lambda ,$$

then

$$\omega_2(f;h) = \mathcal{O}(h). \tag{17}$$

It happens, remarkably enough, that Theorem 41 is true with no further assumptions about K. This is an easy

consequence of Theorem 37, indeed

$$(f * K_\lambda)(t) = \int f(u)\lambda K\left(\lambda(t-u)\right) du$$

$$\frac{d^2(f * K_\lambda)(t)}{dt^2} = \lambda^2 \int f(u)\lambda K''\left(\lambda(t-u)\right) du$$

$$= \lambda^2 \int f(t - \frac{u}{\lambda}) K''(u) du$$

The hypothesis is thus (writing $a = \frac{1}{\lambda}$)

$$\left| \int f(t-au) K''(u) du \right| \leq Ca$$

and this implies (17) by Theorem 37, since K'' is not identically zero.

A logically prior theorem to Theorem 41 is

THEOREM 42. <u>Let f be bounded and continuous on $(-\infty,\infty)$, and K a non-null L^1 function which is piecewise continuous with an L^1 derivative. If</u>

$$\left\| \frac{d(f * K_\lambda)}{dt} \right\| \leq C\lambda^{1-\alpha}, \quad 0<\alpha \leq 1$$

<u>then</u> $f \in \mathrm{Lip}\,\alpha$.

PROOF. Just as in the previous proof, we get

$$\left| \int f(t - \frac{u}{\lambda}) K'(u) du \right| \leq C\lambda^{-\alpha}$$

and by Theorem 37, the result follows when $\alpha < 1$. When $\alpha = 1$,

Theorem 37 does not give the full result, but it is obvious by direct inspection, since (writing $f^{\lambda} = f*K_{\lambda}$)

$$|f^{\lambda}(t+h) - f^{\lambda}(t)| \leq Ch$$

and letting $\lambda \to \infty$ gives the result.

More general theorems of this kind could be formulated, similar to Theorem 38.

CHAPTER 6

APPROXIMATION BY ENTIRE FUNCTIONS OF EXPONENTIAL TYPE

We shall conclude these lectures with some very hasty remarks about an important and interesting kind of approximation. Here some knowledge of complex function theory will be assumed.

An entire analytic function $F(z)$ is of <u>exponential type</u> if it satisfies an inequality of the form

$$|F(re^{i\theta})| \leq Ae^{Br} \tag{1}$$

and the g.l.b. of numbers B such that an inequality of type (1) holds is called the <u>type</u> $\tau(F)$ of F. By E_λ $(\lambda > 0)$ we denote the class of entire functions of type not exceeding a, which are bounded on the real axis. One of the interesting applications of the preceding general theory is to the approximation of bounded uniformly continuous functions on $(-\infty, \infty)$ by functions in E_λ. It is a remarkable fact that E_λ is (for each fixed λ) closed with respect to the sup norm on $(-\infty, \infty)$. (For this, and other useful background, see Boas, Entire Functions). Let us write $E = \bigcup E_\lambda$.

THEOREM 43. <u>Let</u> f <u>be bounded and uniformly continuous on</u> $(-\infty, \infty)$, <u>then there exist functions</u> $F_n \in E$ <u>such that</u> $\lim_{n \to \infty} \|f - F_n\| = 0$.

SKETCH OF PROOF. Let $K(s)$ be a function in E which is L^1 on $(-\infty, \infty)$ and satisfies $\int K dt = 1$. (For

instance, $K(s) = \frac{1}{\pi}\left(\frac{\sin s}{s}\right)^2$) . Then also, $K_\lambda(s) = \lambda K(\lambda s)$ belongs to E. Using Phragmen-Lindelof theorems, one can show

$$f(s;\lambda) = \int_{-\infty}^{\infty} f(u)\,\lambda K\big(\lambda(s-u)\big)\,du$$

converges uniformly on compact subsets of the complex s-plane and $f(s;\lambda)$ is of class E (this is the "hard" part of the proof). That

$$\lim_{\lambda \to \infty} |f(t;\lambda)-f(t)| = 0$$

uniformly in t, follows from Theorem 3.

We can go much further, and prove "degree of approximation" theorems similar to Jackson's. We list only the simplest:

THEOREM 44. <u>Under the hypotheses of Theorem</u> 43, <u>there exist functions</u> $F_\lambda \in E_\lambda$ <u>such that</u>

$$\|f-F_\lambda\| \le C\,\omega(f;\tfrac{1}{\lambda})$$

<u>where</u> C <u>is an absolute constant.</u>

SKETCH OF PROOF. Let K be the function whose Fourier transform is a "trapezoid function"

$$\hat{K}(x) = \begin{cases} 1, & |x| \le \frac{1}{2} \\ 0, & |x| \ge 1 \\ \text{linear, in between} \end{cases}$$

Then,

$$G(s) = \frac{1}{2\pi}\int_{-1}^{1} \hat{K}(x)e^{isx}\,dx$$

is of class E_1 and $G(t)=K(t)$ for real t. It is readily checked that K_λ is (the restriction to the real axis of) an element of E_λ, and finally so is $f * K_\lambda$, which yields the desired approximation.

Everything we proved for trigonometric approximation can be carried over, simply replacing $\mathcal{T}_n$ by E_λ. In fact, the trigonometric theory is just the special case of the latter theory when all functions have period 2π; that is, $\mathcal{T}_n$ is just the intersection of E_λ with $\tilde{C}$. For further details, we refer the reader to Timan's book.

REFERENCES

A. Approximation theory

1. N.I.Akhieser, Lectures on Approximation Theory, English edition, Ungar, 1956

2. P.L.Butzer, Zur Frage der Saturationsklassen singulärer Integral operatoren, Math. Zeitschrift 70(1958) pp.93-112

3. ———————, Representation and approximation of functions by general singular integrals, Indagationes Math. 22 (1960) pp.1-24.

4. ———————, Fourier-Transform methods in the theory of approximation, Arch. Rat. Mech. Anal. 5 (1960) 390-415.

5. G.G.Lorentz, Approximation Theory

6. A.F.Timan, Theory of Approximation of Functions of a real variable, English edition, Pergamon Press, 1963.

B. Fourier Transforms

7. S.Bochner and K.Chandrasekharan, Fourier Transforms, Princeton, 1949.

8. R.R.Goldberg, Fourier Transforms, Cambridge, 1961.

...

List of Matscience Report (1964—66)

Report No.	*Author*	*Title*
35	K. Symanzik	Lecture on a modified models of Euclidean Quantum Field Theory.
36	K. Venkatesan	Report on recent experimental data (1964).
37	A. Fujii	Lectures on Fermi dynamics.
38	M. Gourdin	Mathematical introduction to unitary symmetries.
39	J. V. Narlikar	Theories of Gravitation.
40	K. Venkatesan	Report on recent experimental data (1965).
41	K. R. Unni	Introduction to Hilbert space.
42	L. Rosenfeld	Theory of nuclear reactions.
43	K. R. Unni	Concepts in Modern Mathematics I (Algebra).
44	H. Ruegg	Relativistic generalization of SU. (6)
45	W. K. Hayman	Transfinite diameter and its applications.
46	Ph. Meyer	Selected topics in Weak Interactions.
47	J. Rzewuski	Gauge transformations in quantum field theory.
48	P. L. Kannappan	Theory of functional equations.
49	P. C. Vaidya	Selected topics in Gravitation.
50	K. Venkatesan	Report on recent experimental data (1966).
51	K. R. Unni	Concepts in modern mathematics II (Topology)
52	K. R. Unni	,, ,, III (Analysis)
53	V. Krishnamurthy	Duality Theory in locally convex spaces.
54	J. H. Williamson	Representation theory for Banach algebras and locally compact groups.
55	H. S. Shapiro	Lectures on approximation theory.