Singular Integrals and Differentiability Properties of Functions

ELIAS M. STEIN
Singular Integrals
and Differentiability Properties
of Functions
PRINCETON MATHEMATICAL SERIES
Editors: LUIS A. CAFFARELLI, JOHN N. MATHER, AND ELIAS M. STEIN

1. The Classical Groups by Hermann Weyl
2. An Introduction to Differential Geometry by Luther Pfahler Eisenhart
3. Dimension Theory by W. Hurewicz and H. Wallman
4. The Laplace Transform by D. V. Widder
5. Integration by Edward J. McShane
6. Theory of Lie Groups: I by C. Chevalley
7. Mathematical Methods of Statistics by Harald Cramer
8. Several Complex Variables by S. Bochner and W. T. Martin
9. Introduction to Topology by S. Lefschetz
10. Algebraic Geometry and Topology edited by R. H. Fox, D. C. Spencer, and A. W. Tucker
11. The Topology of Fibre Bundles by Norman Steenrod
12. Foundations of Algebraic Topology by Samuel Eilenberg and Norman Steenrod
13. Continuous Geometry by John von Neumann
14. Riemann Surfaces by L. Ahlfors and L. Sario
15. Differential and Combinatorial Topology edited by S. S. Cairns
16. Convex Analysis by R. T. Rockafellar
17. Geometric Integration Theory by H. Whitney
18. Qualitative Theory of Differential Equations by V. V. Nemytskii and V. V. Stepanov
19. Topological Analysis by Gordon T. Whyburn (revised 1964)
21. Continuous Geometry by John von Neumann
22. Representation Theory of Semisimple Groups: An Overview Based on Examples by Anthony W. Knapp
24. Spin Geometry by H. Blaine Lawson, Jr., and Marie-Louise Michelsohn
25. Topology of 4-Manifolds by Michael H. Freedman and Frank Quinn
26. Hypo-Analytic Structures: Local Theory by François Trèves
27. The Global Nonlinear Stability of the Minkowski Space by Demetrios Christodoulou and Sergiu Klainerman
29. Harmonic Analysis: Real Variable Methods, Orthogonality, and Oscillatory Integrals by Elias M. Stein
30. Topics in Ergodic Theory by Ya. G. Sinai
32. Abelian Varieties with Complex Multiplication and Modular Functions by Goro Shimura
Singular Integrals
and Differentiability Properties
of Functions

ELIAS M. STEIN

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY
1970
To Elly
Preface

This book is an outgrowth of a course which I gave at Orsay during the academic year 1966-67.* My purpose in those lectures was to present some of the required background and at the same time clarify the essential unity that exists between several related areas of analysis. These areas are: the existence and boundedness of singular integral operators; the boundary behavior of harmonic functions; and differentiability properties of functions of several variables. As such the common core of these topics may be said to represent one of the central developments in n-dimensional Fourier analysis during the last twenty years, and it can be expected to have equal influence in the future. These possibilities are further highlighted by listing some of the fields (which are not treated here) where either the particular results of this book, or ideas closely related to the techniques presented here are continuing to find significant application. These include partial differential equations, holomorphic functions of several complex variables, and analysis on other groups, either non-commutative or commutative.

In this connection we should point out that the book *Introduction to Fourier Analysis on Euclidean Spaces** details some of these applications, as well as much background and related material. It may therefore not be inappropriate to view the present volume as a companion to *Fourier Analysis*. Both books, however, may be read independently. In fact an effort has been made to make the present volume essentially self-contained, requiring only elementary facts from integration theory and Fourier transforms as prerequisites.

A brief description of the organization of the book is as follows. The first three chapters deal primarily with material which, for the most part, is beginning to find its way into several advanced texts and monographs, namely covering lemmas and maximal functions, the Marcinkiewicz interpolation theorem, singular integrals generalizing the Hilbert transform, and harmonic functions represented as Poisson integrals. In the last five chapters the topics are of a more advanced nature, including the Littlewood-Paley theory, multipliers, Sobolev spaces and their variants, extension theorems, further results about harmonic functions, and almost-everywhere differentiability theorems. Here part of the material is systematically organized for the first time, and, for example, the last two chapters contain several results whose details were hitherto unpublished.

* For the published lecture notes of this course see Stein [10].
** This work of G. Weiss and the author is referred to as *Fourier Analysis* in the rest of the text.
PREFACE

In any enterprise of this kind the author is faced with the task of balancing two aims which unfortunately are not always compatible. On the one hand there is the desire to facilitate the task of the serious student by providing all the background material and by presenting proofs in a way so that all details, no matter how unenlightening, are fully given. On the other hand there is the need to get on with the essential job of developing the basic ideas of the subject. In doing the latter it is sometimes best to be brief about certain technical details, and also at times to forego the urge to pursue various possible generalizations which could be formulated. Others may surely find fault with how I have weighed these alternatives. My justification would be based either on the ground of personal predilection (which allows no argument) or, in a more serious vein, in terms of my view of the present subject: that it has advanced to a high degree of sophistication and is still rapidly developing, but has not yet reached the level of maturity that would require it to be enshrined in an edifice of great perfection.

It is my pleasant task to acknowledge with gratitude those who have helped me in writing this book: Norman Weiss, who prepared the lecture notes (unpublished) of a course given at Princeton University in 1964-65, where an earlier version of some of this material was presented; Messrs. Bachvan and A. Somen who wrote the published lecture notes already alluded to; Misses Elizabeth Epstein and Florence Armstrong who typed the bulk of the manuscript; and Messrs. W. Beckner, C. Fefferman, and S. Gelbart who helped both mathematically and in proofreading. To all those and others unnamed, I express my thanks.

September 1970

E. M. Stein
Notation

Principal Symbols

dx—Lebesgue measure on \( \mathbb{R}^n \); also \( m(E) \)—measure of the set \( E \)

\( L^p(\mathbb{R}^n) \)—the \( L^p \) space with respect to the measure \( dx \)

\( C^k \)—the class of functions which have continuous derivatives up to and including total order \( k \)

\( \mathcal{D} \)—the space of indefinitely differentiable functions with compact support

\( \mathcal{S} \)—the space of indefinitely differential functions all of whose derivatives remain bounded when multiplied by any polynomial

\( ^c E \)—complement of the set \( E \)

The symbols that follow are listed according to their first and other principal occurrence.

Chapter I, §1

\( B(x, r) \)—ball of radius \( r \) centered at \( x \)

\( M(f) \)—maximal function

§2

\( \delta(x) = \delta(x, F) \)—distance of \( x \) from the set \( F \)

\( I(x), I_\phi(x) \)—integrals of Marcinkiewicz involving the distance function

(See also Chapter VIII, §3.)

§3

\( Q_1, \ldots, Q_k, \ldots \)—cubes; also \( \Omega = \bigcup_k Q_k \)

Chapter II, §1

\( C_0(\mathbb{R}^n), \mathcal{B}(\mathbb{R}^n) \)—continuous functions on \( \mathbb{R}^n \) vanishing at infinity and the dual space of finite Borel measures

\( \mu = \mu_1 * \mu_2 \)—the convolution of measures \( \mu_1 \) and \( \mu_2 \)

\( \hat{f}(y) = \mathcal{F}(f)(y) \)—Fourier transform of \( f \)
§3
H(f)—Hilbert transform. (See also Chapter III, §1.)

§4
$S^{n-1}, d\sigma$—the unit sphere in $\mathbb{R}^n$ and its induced element of volume

§5
$L^p(\mathbb{R}^n, \mathcal{H})$—$L^p$ space of functions which take their values in $\mathcal{H}$

Chapter III, §1
$R_s$—Riesz transforms

$C^\alpha = \nabla^\alpha$, $\alpha_i = \sum_i \alpha_i / n$

Chapter IV, §1
$g, g_1, g_2, g_k$—variants of the Littlewood-Paley $g$ functions

§2
$g^*$—another variant. (See also Chapter VII, §3.)
$\Gamma$—cone $\{(x, y) : x \in \mathbb{R}^n, |x| < y\}$ (See also Chapter VII, §1.)
$\mathcal{S}$—area integral of Lusin. (See also Chapter VII, §§2 and 3.)

§3
$M_\mu(f) = (M(f^n))^{1/\mu}, \mu \geq 1$

$T_m$—multiplier transformation with multiplier $m$  

$\mathcal{M}_p$—algebra of $L^p$ multipliers

§4
$S_p(f)$—"partial sum" operator

§5
$r_m$—Rademacher function. (See also Appendix D.)
Chapter V, §1
\( I_a \)--Riesz potential

§2
\( L^p_k(\mathbb{R}^n) \)--Sobolev space

§3
\( J_a^p(f) = G^p_a(f) \)--Bessel potential
\( \mathcal{L}^p_a(\mathbb{R}^n) \)--space of Bessel potentials
\( \omega_{p}(t) \)--\( L^p \) modulus of continuity
\( \tilde{\omega}_{p}(t) \)--second-order \( L^p \) modulus of continuity

§4
\( \Lambda_a(\mathbb{R}^n) \)--Lipschitz space

§5
\( \Lambda^{p,q}_a(\mathbb{R}^n) \)--Besov space

Chapter VI, §1
\( \Delta(x) \)--regularized distance

§2
\( \mathcal{E}_k \)--Whitney extension operators

§3
\( L^p_k(D) \)--Sobolev space for the domain \( D \)
\( \mathcal{E} \)--extension operator for the domain \( D \)

Chapter VII, §1
\( \Gamma_a(x^0) \)--cone \( \{(x, y) \in \mathbb{R}^{n+1}, |x - x^0| < \alpha y\} \)
\( \Gamma^h_a \)--truncated cone, \( \Gamma^h_a(x^0) = \Gamma_a(x^0) \cap \{0 < y < h\} \)
\( \mathcal{R} = \bigcup_{x^0 \in E} \Gamma^h_a(x^0) \). (See also Chapter VIII, §2.)

§3
\( H^p \)--space of conjugate harmonic functions satisfying an appropriate \( L^p \) inequality
\( \mathcal{E}_a \)--variant of the area integral
# Contents

**Preface**  

**Notation**  

<table>
<thead>
<tr>
<th>I. SOME FUNDAMENTAL NOTIONS OF REAL-VARIABLE THEORY</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The maximal function</td>
<td>4</td>
</tr>
<tr>
<td>2. Behavior near general points of measurable sets</td>
<td>12</td>
</tr>
<tr>
<td>3. Decomposition in cubes of open sets in $\mathbb{R}^n$</td>
<td>16</td>
</tr>
<tr>
<td>4. An interpolation theorem for $L^p$</td>
<td>20</td>
</tr>
<tr>
<td>5. Further results</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. SINGULAR INTEGRALS</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Review of certain aspects of harmonic analysis in $\mathbb{R}^n$</td>
<td>27</td>
</tr>
<tr>
<td>2. Singular integrals: the heart of the matter</td>
<td>28</td>
</tr>
<tr>
<td>3. Singular integrals: some extensions and variants of the preceding</td>
<td>34</td>
</tr>
<tr>
<td>4. Singular integral operators which commute with dilations</td>
<td>38</td>
</tr>
<tr>
<td>5. Vector-valued analogues</td>
<td>45</td>
</tr>
<tr>
<td>6. Further results</td>
<td>48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. RIESZ TRANSFORMS, POISSON INTEGRALS, AND SPHERICAL HARMONICS</th>
<th>54</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Riesz transforms</td>
<td>54</td>
</tr>
<tr>
<td>2. Poisson integrals and approximations to the identity</td>
<td>60</td>
</tr>
<tr>
<td>3. Higher Riesz transforms and spherical harmonics</td>
<td>68</td>
</tr>
<tr>
<td>4. Further results</td>
<td>77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. THE LITTLEWOOD-PALEY THEORY AND MULTIPLIERS</th>
<th>81</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Littlewood-Paley $g$-function</td>
<td>82</td>
</tr>
<tr>
<td>2. The function $g^*_1$</td>
<td>86</td>
</tr>
<tr>
<td>3. Multipliers (first version)</td>
<td>94</td>
</tr>
</tbody>
</table>
CONTENTS

4. Application of the partial sums operators 99
5. The dyadic decomposition 103
6. The Marcinkiewicz multiplier theorem 108
7. Further results 112

V. DIFFERENTIABILITY PROPERTIES IN TERMS OF FUNCTION SPACES 116

1. Riesz potentials 117
2. The Sobolev spaces, \( L^p(\mathbb{R}^n) \) 121
3. Bessel potentials 130
4. The spaces \( \Lambda_g \) of Lipschitz continuous functions 141
5. The spaces \( \Lambda^{p,q} \) 150
6. Further results 159

VI. EXTENSIONS AND RESTRICTIONS 166

1. Decomposition of open sets into cubes 167
2. Extension theorems of Whitney type 170
3. Extension theorem for a domain with minimally smooth boundary 180
4. Further results 192

VII. RETURN TO THE THEORY OF HARMONIC FUNCTIONS 196

1. Non-tangential convergence and Fatou’s theorem 196
2. The area integral 205
3. Application of the theory of \( H^p \) spaces 217
4. Further results 235

VIII. DIFFERENTIATION OF FUNCTIONS 240

1. Several notions of pointwise differentiability 241
2. The splitting of functions 246
3. A characterization of differentiability 250
4. Desymmetrization principle 257
5. Another characterization of differentiability 262
6. Further results 266

APPENDICES

A. Some Inequalities 271
B. The Marcinkiewicz Interpolation Theorem 272
C. Some Elementary Properties of Harmonic Functions 274
D. Inequalities for Rademacher Functions 276

BIBLIOGRAPHY 279

INDEX 289
Singular Integrals
and Differentiability Properties
of Functions
CHAPTER I

Some Fundamental Notions of
Real-Variable Theory

The basic ideas of the theory of real variables are connected with the concepts of sets and functions, together with the processes of integration and differentiation applied to them. While the essential aspects of these ideas were brought to light in the early part of our century, some of their further applications were developed only more recently. It is from this latter perspective that we shall approach that part of the theory that interests us. In doing so, we distinguish several main features:

(1) The theorem of Lebesgue about the *differentiation of the integral*. The study of properties related to this process is best done in terms of a "maximal function" to which it gives rise; the basic features of the latter are expressed in terms of a "weak-type" inequality which is characteristic of this situation.

(2) Certain *covering* lemmas. In general the idea is to cover an arbitrary (open) set in terms of a disjoint union of cubes or balls, chosen in a manner depending on the problem at hand. One such example is a lemma of Whitney, (Theorem 3). Sometimes, however, it suffices to cover only a portion of the set, as in the simple covering lemma, which is used to prove the weak-type inequality mentioned above.

(3) *Behavior near a "general" point* of an arbitrary set. The simplest notion here is that of point of density. More refined properties are best expressed in terms of certain integrals first studied systematically by Marcinkiewicz.

(4) The *splitting of functions* into their large and small parts. This feature which is more of a technique than an end in itself, recurs often. It is especially useful in proving $L^p$ inequalities, as in the first theorem of this chapter. That part of the proof of the first theorem is systematized in the Marcinkiewicz interpolation theorem discussed in §4 of this chapter and also in Appendix B.
1. The maximal function

1.1 According to the fundamental theorem of Lebesgue, the relation

\[ \lim_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} f(y) \, dy = f(x) \]

holds for almost every \( x \), whenever \( f \) is locally integrable function defined on \( \mathbb{R}^n \). The notation here used is that \( B(x, r) \) is the ball of radius \( r \), centered at \( x \), and \( m(B(x, r)) \) denotes its measure. In order to study the limit (1) we consider its quantitative analogue, where \( \lim \) is replaced by \( \sup \); this is the maximal function, \( Mf \). Since the properties of this function are expressed in terms of relative size and do not involve any cancellation of positive and negative values, we replace \( f \) by \( |f| \). Thus we define

\[ M(f)(x) = \sup_{r > 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y)| \, dy \]

It is to be noticed that nothing excludes the possibility that \( (Mf)(x) \) is infinite for any given \( x \).

The passage from a limiting expression to a corresponding maximal function is a situation that recurs often. Our first example here, (2), will turn out to be the most fundamental one.

1.2 We shall now be interested in giving a concise expression for the relative size of a function. Thus let \( g(x) \) be defined on \( \mathbb{R}^n \) and for each \( \alpha \) consider the set where \( |g| \) is greater than \( \alpha \),

\[ \{ x : |g(x)| > \alpha \} \]

The function \( \lambda(\alpha) \), defined to be the measure of this set, is the sought-for expression. It is the distribution function of \( |g| \).

In particular, the decrease of \( \lambda(\alpha) \) as \( \alpha \) grows describes the relative largeness of the function; this is the main concern locally. The increase of \( \lambda(\alpha) \) as \( \alpha \) tends to zero describes the relative smallness of the function “at infinity”; this is its importance globally, and is of no interest if, for example, the function is supported on a bounded set.

Any quantity dealing solely with the size of \( g \) can be expressed in terms of the distribution function \( \lambda(\alpha) \). For example, if \( g \in L^p \), then

\[ \int_{\mathbb{R}^n} |g(y)|^p \, dy = - \int_0^\infty \alpha^p \, d\lambda(\alpha) \]
§1. THE MAXIMAL FUNCTION

and if \( g \in L^\infty \), then
\[
\|g\|_\infty = \inf \{\alpha, \lambda(\alpha) = 0\}.
\]

A related fact concerning the distribution function will have immediate application. It is this: If \( g \) is integrable, then
\[
\lambda(\alpha) \leq A/\alpha \quad \text{where} \quad A = \int_{\mathbb{R}^n} |g(y)| \, dy.
\]
In fact
\[
\int_{\mathbb{R}^n} |g(y)| \, dy \geq \int_{|x| > \alpha} |g(y)| \, dy \geq \alpha \lambda(\alpha),
\]
which proves the assertion.

1.3 With these definitions we can state our first theorem. It gives the main results for the maximal function, and has as a corollary the differentiability almost everywhere of the integral, expressed in (1).

**Theorem 1.** Let \( f \) be a given function defined on \( \mathbb{R}^n \)

(a) If \( f \in L^p(\mathbb{R}^n), 1 \leq p \leq \infty \), then the function \( Mf \) is finite almost everywhere.

(b) If \( f \in L^1(\mathbb{R}^n) \), then for every \( \alpha > 0 \)
\[
m\{x: (Mf)(x) > \alpha \} \leq \frac{A}{\alpha} \int_{\mathbb{R}^n} |f| \, dx,
\]
where \( A \) is a constant which depends only on the dimension \( n \) (\( A = 5^n \) will do)

(c) If \( f \in L^p(\mathbb{R}^n) \), with \( 1 < p \leq \infty \), then \( Mf \in L^p(\mathbb{R}^n) \) and
\[
\|Mf\|_p \leq A_p \|f\|_p,
\]
where \( A_p \) depends only \( p \) and the dimension \( n \).

**Corollary 1.** If \( f \in L^p(\mathbb{R}^n), 1 \leq p \leq \infty \), or more generally if \( f \) is locally integrable, then
\[
\lim_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} f(y) \, dy = f(x),
\]
for almost every \( x \).

1.4 Before we come to the proof of the theorem we make some clarifying comments.

(i) In contrast with the case \( p > 1 \), when \( p = 1 \) the mapping \( f \to M(f) \) is not bounded on \( L^1(\mathbb{R}^n) \). Thus if \( f \) is not identically zero \( Mf \) is never integrable on all of \( \mathbb{R}^n \). This can be seen by making the simple observation
that $Mf(x) \geq c|x|^{-n}$, for $|x| \geq 1$. Moreover even if we limit our considerations to any bounded subset of $\mathbb{R}^n$, then the integrability of $Mf$ holds only if stronger conditions than the integrability of $f$ are required. (See §5.2 below.)

(ii) The result that is obtained, namely estimate (b), is weaker than the statement that $f \to M(f)$ is bounded on $L^1(\mathbb{R}^n)$, as the remarks in §1.2 show; for this reason (b) is referred to as a weak-type estimate. This estimate is the best possible (as far as order of magnitude) for the distribution function of $M(f)$, where $f$ is an arbitrary function in $L^1(\mathbb{R}^n)$. That this is so can be seen by replacing the measure $|f(y)| \, dy$ in definition (2) by the measure $d\mu$, whose total measure of one is concentrated at the origin; ($d\mu$ is the "Dirac measure"). Then $M(d\mu)(x) = c|x|^{-n}$, where $c^{-1} = \text{volume of the unit ball}$. In this case the distribution function $\lambda(\alpha)$ is exactly $1/\alpha$. But we can always find a sequence $\{f_n(x)\}$ of positive integrable functions, whose $L^1$ norm is each one, and which converge weakly to the measure $d\mu$. So we cannot expect an estimate essentially stronger than (b), since in the limit a similar stronger version would have to hold for $M(d\mu)(x)$.

1.5 Proof of Theorem 1 and its corollary. Here we shall prove the theorem and its corollary, taking for granted the covering lemma of "Vitali-type" stated in §1.6 and proved in §1.7 below. With the definition of $Mf$, and with

$$E_\alpha = \{x: Mf(x) > \alpha\}$$

then for each $x \in E_\alpha$ there exists a ball of center $x$, which we call $B_x$, so that

$$(3) \quad \int_{B_x} |f(y)| \, dy > \alpha m(B_x).$$

But on the one hand (3) gives $m(B_x) < (1/\alpha) \|f\|_1$, for all such $x$; on the other hand when $x$ runs through the set $E_\alpha$ the union of the corresponding $B_x$ covers $E_\alpha$. Thus using the covering lemma (1.6) below from this family of balls we can extract a sequence of balls, which we designate by $\{B_k\}$; these balls are mutually disjoint and have the property that

$$(4) \quad \sum_{k=0}^{\infty} m(B_k) \geq C m(E_\alpha),$$

(e.g. the bound $C = 5^{-n}$ will work). Applying (3) and then (4) to each of the mutually disjoint balls we get

$$\int \bigcup_{B_k} |f(y)| \, dy > \alpha \sum_{k} m(B_k) \geq \alpha C m(E_\alpha).$$
§1. THE MAXIMAL FUNCTION

But since the first member of this inequality is majorized by \( \|f\|_1 \), on taking \( A = 1/C \) we obtain the assertion (b) of the theorem; (and thus also part (a), when \( p = 1 \)). We shall now prove simultaneously assertion (a) (the finiteness almost everywhere of \( M(f)(x) \)), and assertion (c) (the \( L^p \) inequality), for \( 1 < p \leq \infty \). The case \( p = \infty \) is of course trivial, and here the bound is \( A_\infty = 1 \). Let us therefore suppose that \( 1 < p < \infty \). We shall use a simple example of the technique of splitting a function into its large and small parts, alluded to at the beginning of this chapter. Let us define \( f_1(x) \) by \( f_1(x) = f(x) \), if \( |f(x)| \geq \alpha/2 \), and \( f_1(x) = 0 \) otherwise. Then we have successively

\[ |f(x)| \leq |f_1(x)| + \alpha/2; \quad M(f)(x) \leq M(f_1)(x) + \alpha/2, \]

therefore

\[ \{x: M(f)(x) > \alpha\} \subseteq \{x: M(f_1)(x) > \alpha/2\}, \]

and finally

\[
m(E_\alpha) = m\{x: Mf(x) > \alpha\} \leq \frac{2A}{\alpha} \|f_1\|_1, \]

which is

\[
(5) \quad m(E_\alpha) = m\{x: Mf(x) > \alpha\} \leq \frac{2A}{\alpha} \int_{|f| > \alpha/2} |f| \, dx.
\]

The last inequality is obtained by applying conclusion (b) of the theorem which we may since \( f_1 \in L^1 \) whenever \( f \in L^p \). We now set \( g = M(f) \), and \( \lambda \) the distribution function of \( g \). Then using the observations in (1.2) together with an integration by parts we have

\[
\int_{\mathbb{R}^n} (Mf)^p \, dx = -\int_0^\infty \alpha^p \, d\lambda(\alpha) = p\int_0^\infty \alpha^{p-1} \lambda(\alpha) \, d\alpha.
\]

In particular, because of (5),

\[
\|Mf\|_p^p = p\int_0^\infty \alpha^{p-1} m(E_\alpha) \, d\alpha \leq p\int_0^\infty \alpha^{p-1} \left( \frac{2A}{\alpha} \int_{|f| > \alpha/2} |f(x)| \, dx \right) \, d\alpha.
\]

The double integral is evaluated by interchanging the orders of integration and integrating first with respect to \( \alpha \). The inner integral is then

\[
\int_0^{|f(x)|} \alpha^{p-2} \, d\alpha = \left( \frac{1}{p - 1} \right) |2f(x)|^{p-1},
\]

since \( p > 1 \). So the double integral has the value

\[
\frac{2Ap}{p - 1} \int_{\mathbb{R}^n} |f| \, |2f|^{p-1} \, dx = (A_p)^p \int_{\mathbb{R}^n} |f|^p \, dx,
\]

which proves conclusion (c). Calculating the constants we get

\[
A_p = 2\left( \frac{5^np}{p - 1} \right)^{1/p}, \quad 1 < p < \infty.
\]
It is useful, for certain applications, to observe that

\[ A_p = O\left(\frac{1}{p - 1}\right), \quad p \to 1. \]

We now come to the proof of the corollary. We easily reduce the consideration to the case \( p = 1 \), by multiplying our original function by the characteristic function of a ball, and then exhausting \( \mathbb{R}^n \) by a denumerable union of such balls. Let us denote by \( f_r \) the function

\[ f_r(x) = \frac{1}{m(B(x, r))} \int_{B(x, r)} f(y) \, dy, \quad r > 0. \]

We know* that if \( r \to 0, \| f_r - f \|_1 \to 0 \), whenever \( f \in L^1(\mathbb{R}^n) \).

Therefore \( f_{r_k} \to f \), almost everywhere for a suitable sequence \( \{r_k\} \to 0 \).

What remains to be seen, therefore, is that \( \lim_{r \to 0} f_r(x) \) exists almost everywhere. For this purpose let us denote for each \( g \in L^1 \), and \( x \in \mathbb{R}^n \)

\[ \Omega g(x) = \lim_{r \to 0} \sup_{\{g_r\}} g_r(x) - \lim_{r \to 0} \inf_{\{g_r\}} g_r(x) \]

where \( g_r \) is defined like \( f_r \). \( \Omega g \) represents the oscillation of the family \( \{g_r\} \), as \( r \to 0 \).

If \( g \) is continuous with compact support, then \( g_r \to g \) uniformly, and thus \( \Omega g \) is identically zero in this case.

Next if \( g \) is in \( L^1(\mathbb{R}^n) \), then by conclusion (b) of the theorem

\[ m\{x : 2M(g) > \varepsilon\} \leq \frac{2A}{\varepsilon} \| g \|_1. \]

However clearly \( \Omega g(x) \leq 2Mg(x) \), thus

\[ m\{x : \Omega g(x) > \varepsilon\} \leq \frac{2A}{\varepsilon} \| g \|_1, \quad g \in L^1(\mathbb{R}^n). \]

Finally any \( f \in L^1(\mathbb{R}^n) \) can be written as \( f = h + g \), where \( h \) is continuous with compact support and where the \( L^1 \) norm of \( g \) is at our disposal. But \( \Omega f \leq \Omega h + \Omega g \), and \( \Omega h \equiv 0 \) since \( h \) is continuous. Therefore (8) shows that

\[ m\{x : \Omega f(x) > \varepsilon\} \leq \frac{2A}{\varepsilon} \| g \|_1. \]

* This is a particular property of approximations of the identity. See Chapter III, §2.2 for a detailed discussion; the relevant part of that section can be used without fear of circularity.
§1. THE MAXIMAL FUNCTION

Since the norm of \( g \) can be chosen to be arbitrarily small we get \( \Omega f = 0 \) almost everywhere, which means that \( \lim_{r \to 0} f_r(x) \) exists almost everywhere.

The following summarizing comment about the proof of the corollary is worth making. The argument used was of a very general nature and occurs often. That is, the almost everywhere convergence is proved as a combination of two parts, one which is deep and already contains the essence of the result; it is expressed in terms of a maximal inequality like part (b) (or (c)) of the theorem. The second fact is usually much simpler to establish but it is just as essential. It is the convergence almost everywhere for a dense subset of the function space, in this case the continuous function on \( \mathbb{R}^n \) with compact support.

1.6 A covering lemma. We have therefore completed the proof of Theorem 1 and its corollary, save for the crucial step of the covering lemma, which we postponed until now. Not only the simplicity of its statement, or the application we use it for, but also the many variants of it that can be found in the mathematical literature attest to the fundamental character of this lemma. The reader should note that its statement and proof are closely related to a more refined but probably better known lemma of Vitali.*

**Lemma.** Let \( E \) be a measurable subset of \( \mathbb{R}^n \) which is covered by the union of a family of balls \( \{B_i\} \), of bounded diameter. Then from this family we can select a disjoint subsequence, \( B_1, B_2, \ldots B_k, \ldots \) (finite or infinite) so that

\[
\sum_k m(B_k) \geq C m(E)
\]

Here \( C \) is a positive constant that depends only on the dimension \( n \); \( C = 5^{-n} \) will do.

1.7 We begin the proof of the lemma by describing the choice of \( B_1, B_2, \ldots B_k, \ldots \). We choose \( B_1 \) so that it is essentially as large as possible; that is so that the diameter of \( B_1 \geq \frac{1}{2} \sup \{ \text{diameter } B_i \} \). Of course the choice of a \( B_1 \) satisfying these conditions, as well as the later choices of the other \( B_k, \) is not unique; but this non-uniqueness is of no consequence to us. Let us now suppose that \( B_1, B_2, \ldots B_k \) have already been chosen. We are now forced to select \( B_{k+1} \) from those balls in the family \( \{B_j\} \) which are disjoint with \( B_1, B_2, \ldots B_k \). We choose one that again is essentially as large as possible. That is we take \( B_{k+1} \) to be disjoint from \( B_1, \ldots B_k \), and the diameter of \( B_{k+1} \geq \frac{1}{2} \{ \sup \text{diameter of } B_j, \text{ with } B_j \text{ disjoint from } B_1, B_2, \ldots B_k \} \).

* The lemma of Vitali may be found in §5.4 below.
In this way we get the sequence $B_1, B_2, \ldots B_k, \ldots$ of balls. In principle this sequence could be finite, and terminate at $B_k$; this would be the case if there were no balls in $\{B_j\}$ disjoint with $B_1, B_2, \ldots B_k$.

Now two cases present themselves, depending on whether $\sum m(B_k) = \infty$ or $\sum m(B_k) < \infty$. In the first case we have attained our conclusion whether $m(E)$ is infinite or finite. Let us therefore consider the case when $\sum m(B_k) < \infty$.

For this purpose we denote by $B_k^*$ the ball having the same center as $B_k$, but whose diameter is five times as large. We claim that

\[
(9) \quad \bigcup_k B_k^* \supseteq E.
\]

To prove (9) we have to show that $U_k B_k^* \supseteq B_j$, for any fixed $B_j$ in our given family which covers $E$. We may certainly assume that our fixed $B_j$ is not one of the sequence $B_1, B_2, \ldots B_k, \ldots$, for otherwise there is nothing to prove. Since $\sum m(B_k) < \infty$, then $\operatorname{diam} (B_k) \rightarrow 0$, as $k \rightarrow \infty$, and so we take the first $k$, with the property that $\operatorname{diam} (B_{k+1}) < \frac{1}{5}(\operatorname{diam} B_j)$. Now the ball $B_j$ must intersect one of the $k$ previous balls $B_1, B_2, \ldots B_k$, or it should have been picked as the $k + 1^{th}$ ball instead of $B_{k+1}$, since its diameter is more than twice that of $B_{k+1}$. That is $B_j$ intersects $B_{j_0}$, for some $1 \leq j_0 \leq k$, and $\frac{1}{5}(\text{diameter of } B_j) \leq \text{diameter of } B_{j_0}$. From an obvious geometric consideration it is then evident that $B_j$ is contained in the ball that has the same center as $B_{j_0}$, but five times the diameter of $B_{j_0}$, i.e. $B_j \subset B_{j_0}^*$.

Thus we have proved (9), and so

\[
m(E) \leq \sum m(B_k^*) = 5^n \sum m(B_k),
\]

which proves the lemma.

1.8 Lebesgue set. The differentiation theorem just proved refers to the limits of averages taken with respect to balls. But this theorem has, as a rather simple consequence of itself, a generalization where the averages are taken over more general families of sets.

Let $\mathcal{F}$ be a family of measurable subsets of $\mathbb{R}^n$. We shall say that this family is regular, if there exists a constant $c > 0$, so that if $S \in \mathcal{F}$, then $S \subset B$, with $m(S) \geq cm(B)$, where $B$ is an appropriate open ball centered at the origin. Examples of such regular families are: (1) the family $\mathcal{F}$ of all sets of the $\delta U$, $0 < \delta < \infty$, (which are the dilations of a fixed set $U$), where $U$ is bounded and $m(U) > 0$. (2) the family of all cubes with the property that their distance from the origin is bounded by a constant multiple of their diameter. (3) any subfamily of such a family $\mathcal{F}$. In analogy with the special case of the family of all balls centered at the
§1. THE MAXIMAL FUNCTION

origin, we defined the appropriate maximal function

\[ M_f(f)(x) = \sup_{S \in \mathcal{F}} \frac{1}{m(S)} \int_S |f(x - y)| \, dy \]

Then clearly \( M_f(f)(x) \leq c^{-1} Mf(x) \), and therefore \( M_f \) satisfies the same conclusion as those in theorem 1 for \( M \). So a repetition of the argument of the corollary leads to the fact that whenever \( f \) is locally integrable

\[ \lim_{m(S) \to 0} \frac{1}{m(S)} \int_S f(x - y) \, dy = f(x), \]

for almost every \( x \).

All of this is very simple, but is not completely satisfying for the following reason. Given a fixed locally integrable function \( f \), we have proved that the relation (10) holds almost everywhere, but the exceptional set (of measure zero) depends on the given family \( \mathcal{F} \). It would be better if we could find one exceptional set of measure zero (depending on \( f \)), so that outside of it the relation (10) would hold for every regular family. This is the role of the complement of the Lebesgue set of \( f \), where the latter set is defined as those \( x \) for which

\[ \lim_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x,r)} |f(y) - f(x)| \, dy = 0. \]

(Recall that \( B(x, r) \) is the ball of radius \( r \) centered at \( x \))

To see that the limit (11) is realized almost everywhere, we consider the relation

\[ \lim_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x,r)} |f(y) - c| \, dy = |f(x) - c| \]

which holds for each constant \( c \), for almost every \( x \). That is, there is an exceptional set \( E_c \), with \( m(E_c) = 0 \), so that (11') is valid whenever \( x \notin E_c \). Let \( c_1, c_2, \ldots, c_n, \ldots \) be an enumeration of the rationals. If \( x \notin E = \bigcup_c E_c \), then (11') holds for any rational \( c \), and so by continuity for every real \( c \).

In particular, \( x \) in the complement of the set \( E \) are in the Lebesgue set of \( f \); that is, for those \( x \), (11) is valid.

But

\[ \left| \frac{1}{m(S)} \int_S f(x - y) \, dy - f(x) \right| = \left| \frac{1}{m(S)} \int_S [f(x - y) - f(x)] \, dy \right| \]

\[ \leq \frac{1}{m(S)} \int_S |f(x - y) - f(x)| \, dy \]

\[ \leq c^{-1} \frac{1}{m(B(x, r))} \int_{B(x,r)} |f(y) - f(x)| \, dy, \]
so that differentiability with respect to any regular family is established at every point of the Lebesgue set of \( f \).

For a discussion of the case of non-regular families, see §5.3 below.

2. Behavior near general points of measurable sets

2.1 In this section we wish to treat various properties of measurable sets of positive measure which confirm the observation that a "general" point of a set is almost completely surrounded by other points of the set. The simplest concrete example of this heuristic principle is contained in the notion of a point of density.

Suppose \( E \) is a given measurable set, and \( x \in \mathbb{R}^n \). Then we say that \( x \) is a point of density of \( E \), if

\[
\lim_{r \to 0} \frac{m(E \cap B(x, r))}{m(B(x, r))} = 1.
\]

Of course for general \( x \) the limit need not have value 1, or may not even exist; but if the limit in (12) has value 0, then according to our definition \( x \) is a point of density of the complement of \( E \). Let us now apply the differentiation theorem (the corollary to theorem 1 stated in §1.3), to the case when \( f = \chi_E \), the characteristic function of the set \( E \). This gives us immediately the following proposition.

**Proposition 1.** For almost every point \( x \in E \), the limit (12) holds; that is almost every point \( x \in E \) is a point of density of \( E \), and almost every point of the complement of \( E \) is not a point of density of \( E \).

Notice that if we had restricted our attention to the points of the Lebesgue set of \( \chi_E \), we should have obtained instead of Proposition 1 a similar but stronger conclusion. The balls in (12) could then have been replaced by regular families, in the sense of §1.8.

2.2 In order to continue, we shall now limit ourselves to sets \( E \) which are closed, but are still otherwise arbitrary. The reason for this limitation is obvious: In what follows the results will be expressed in terms of the distance from \( E \); if \( E \) is not closed the distance from \( E \) is in reality the distance from \( \bar{E} \), the closure of \( E \), and clearly \( E \) and \( \bar{E} \) may be quite different measure-theoretically. However, the limitation to closed sets is not a serious obstacle in applications. Closed sets are sufficiently general; in particular, any measurable set may be approximated by the closed sets it contains, so that the respective difference sets have measure as small as we wish.
§2. BEHAVIOR OF MEASURABLE SETS

To reflect our newly imposed restriction we shall denote a general closed subset of $\mathbb{R}^n$ by $F$, and we let $\delta(x) = \delta(x, F)$ represent the distance of the point $x$ from $F$. Of course $\delta(x) = 0$ if and only if $x \in F$. Now it is clear that if $x \in F$, $\delta(x + y) \leq |y|$, since $x$ is a point in $F$ whose distance from $x + y$ is equal to $|y|$. However in general, this estimate of the distance from $F$ can be improved; that is $\delta(x + y) = o(|y|)$, for most $x$ in $F$. The relation of “little $o$” means that given any $\varepsilon > 0$, there exists a $\eta = \eta_\varepsilon$, so that $\delta(x + y) \leq \varepsilon |y|$, if $|y| \leq \eta$.

Proposition 2. Let $F$ be a closed set. Then for almost every $x \in F$, $\delta(x + y) = o(|y|)$. This holds in particular if $x$ is a point of density of $F$.

We have formulated this proposition mainly because it is a simple illustration of the notion of point of density. We shall, however, also find an application for this proposition, but this is not until much later.

Figure 1. The point of density argument. The larger ball is $B(x, |y| + \varepsilon |y|)$, and the smaller ball is $B(x + y, \varepsilon |y|)$.

To prove the proposition, let $x$ be a point of density of $F$ and suppose $\varepsilon$ is given, $\varepsilon > 0$. Consider the “small” ball of center $x + y$, and radius $\varepsilon |y|$; and the “large” ball of center $x$ and radius $|y| + \varepsilon |y|$. Obviously
$B(x + y, \varepsilon |y|) \subseteq B(x, |y| + \varepsilon |y|)$. We claim that if $|y|$ is sufficiently small then there exists a $z \in F$, so that $z \in B(x + y, \varepsilon |y|)$. For otherwise $F \cap B(x + y, \varepsilon |y|) = \emptyset$, and

$$
\frac{m(F \cap B(x, |y| + \varepsilon |y|))}{m(B(x, |y| + \varepsilon |y|))} \leq \frac{m(B(x, |y| + \varepsilon |y|)) - m(B(x + y, \varepsilon |y|))}{m(B(x, |y| + \varepsilon |y|))} \\
\leq 1 - \left(\frac{\varepsilon}{1 + \varepsilon}\right)^n,
$$

which contradicts (12) if $|y|$ is small enough. Thus there exists a $z$ in $F$ which also is in $B(x + y, \varepsilon |y|)$; this means that within a distance of $\varepsilon |y|$ from the point $x + y$ we can find a point of $F$, i.e. $\delta(x + y) \leq \varepsilon |y|$.

### 2.3 Integral of Marcinkiewicz

We shall now present another expression of the principle that a general point of a measurable set is almost completely surrounded by other points of the set. This form will be independent of the theorem of differentiation, but for many problems it will have a significance which is equally important. In fact, the integrals considered below, first treated systematically by Marcinkiewicz, intervene in a decisive way in the theory of singular integrals, as discussed in the following chapter, as well as other problems treated in this book.

We consider as before a closed set, $F$; $\delta(x)$ denotes the distance of $x$ from $F$, and we shall study the integral $I(x)$ given by

$$I(x) = \int_{|y| \leq 1} \frac{\delta(x + y)}{|y|^{n+1}} \, dy.
$$

**Theorem 2.** (a) When $x \in$ complement of $F$, then $I(x) = \infty$.

(b) For almost every $x \in F$, $I(x) < \infty$.

The conclusion (a) is evident, since the complement of $F$ is an open set. Then if $x$ belongs to this complement $\delta(x + y) \geq c > 0$, for a neighborhood of the origin in $y$. The conclusion (b) is the interest of this theorem, and it states in effect that the estimate $\delta(x + y) = o(|y|)$ of Proposition 2 can be refined on the average, so as to lead to the convergence of the integral (13).

The theorem will be a simple consequence of the following lemma, which is a more quantitative expression of the same fact.

**Lemma.** Let $F$ be a closed set whose complement has finite measure. With $\delta(x)$ defined as above we let

$$I_*(x) = \int_{\mathbb{R}^n} \frac{\delta(x + y)}{|y|^{n+1}} \, dy.
$$
§ 2. Behavior of Measurable Sets

Then $I_\ast(x) < \infty$ for almost every $x \in F$. Moreover

\begin{equation}
\int_F I_\ast(x) \, dx \leq c \cdot m(cF).
\end{equation}

2.4 In proving the lemma, we observe that it suffices to prove (15) since the integrand is positive. Also by the same positivity we can interchange the orders of integration in evaluating the left side of (15). This will accomplish the proof. In detail:

$$
\int_F I_\ast(x) \, dx = \int_F \int_{\mathbb{R}^n} \frac{\delta(x + y)}{|y|^{n+1}} \, dy \, dx = \int_F \int_{\mathbb{R}^n} \frac{\delta(y)}{|x - y|^{n+1}} \, dy \, dx
$$

\begin{equation}
= \int_F \int_{\mathbb{R}^n} \frac{\delta(y)}{|x - y|^{n+1}} \, dy \, dx = \int_F \left( \int_{\mathbb{R}^n} \frac{dx}{|x - y|^{n+1}} \right) \delta(y) \, dy.
\end{equation}

Now consider

$$
\int_F \frac{dx}{|x - y|^{n+1}} \text{ with } y \in cF.
$$

The smallest value of $|x - y|$ (as $x$ varies over $F$) is of course $\delta(y)$, which is the distance of $y$ from $F$. Thus

$$
\int_F \frac{dx}{|x - y|^{n+1}} \leq \int_{|x| \leq \delta(y)} \frac{dx}{|x|^{n+1}} \leq c(\delta(y))^{-1}.
$$

This shows that

$$
\int_F I_\ast(x) \, dx \leq \int_F c(\delta(y))^{-1} \delta(y) \, dy = cm(cF),
$$

and the lemma is proved.

Theorem 2 is obtained from this lemma as follows. Let $B_m$ denote the open ball of radius $m$ center at the origin, and let $F_m = F \cup cB_m$. Then $F_m$ is closed but its complement has finite measure (since it is contained in $B_m$). Thus we can apply the lemma to $F_m$. So let $\delta_m$ denote the distance from $F_m$, and $\delta$ the distance from $F$. Observe that $\delta(x + y) = \delta_m(x + y)$, if $|y| \leq 1$ and $x \in B_{m-2}$. Hence the lemma implies that $I(x) < \infty$, for almost every $x \in F \cap B_{m-2}$. Letting $m \to \infty$ we get the desired result.

Among the several variants of the theorem and the lemma we present here one. (Another variant is discussed at the end of this chapter in §5.) We can replace $I(x)$ by

$$
I^{(2)}(x) = \int_{|y| \leq 1} \frac{\delta^2(x + y)}{|y|^{n+2}} \, dy,
$$

\end{document}
where \( \lambda > 0 \). Similarly \( I_*(x) \) can be replaced by

\[
I_*(x) = \int_{\mathbb{R}^n} \frac{\delta^\lambda(x + y)}{|y|^{n+\lambda}} \, dy, \quad \lambda > 0.
\]

In both cases similar conclusions are obtained with the above methods.

3. Decomposition in cubes of open sets in \( \mathbb{R}^n \)

3.1 The decomposition of a given set into a disjoint union of cubes (or balls) is a fundamental tool in the theory described in this chapter. We have already used this type of notion, in very rough form, in the covering lemma, §1.6.

3.1.1 We now pose ourselves the following related general problem which, however, does not involve measure theory, but deals with the geometric structure of general closed sets \( F \) in \( \mathbb{R}^n \): Can the complement of \( F \) be realized as a disjoint union of cubes in a canonical way? For \( n = 1 \) the answer is of course yes, since every open set is in a unique way the union of disjoint open intervals. For \( n \geq 2 \), the situation is no longer that simple, since we can realize an arbitrary open set in an infinity of different ways as a disjoint union of cubes (by cubes we now mean closed cubes; by disjoint we mean that their interiors are disjoint). However there are decompositions, which while not canonical, are very satisfactory and useful substitutes. We have in mind the idea first introduced by Whitney and formulated as follows.

**Theorem 3.** Let \( F \) be a non-empty closed set in \( \mathbb{R}^n \). Then its complement \( \Omega \) is the union of a sequence of cubes \( Q_k \), whose sides are parallel to the axes, whose interiors are mutually disjoint, and whose diameters are approximately proportional to their distances from \( F \). More explicitly:

(i) \( \Omega = ^cF = \bigcup_{k=1}^\infty Q_k \).

(ii) \( Q_j \cap Q_k = \emptyset \) if \( j \neq k \).

(iii) There exist two constants \( c_1, c_2 > 0 \), (we can take \( c_1 = 1 \), and \( c_2 = 4 \)), so that

\[
c_1 \text{ (diameter } Q_k) \leq \text{ distance } Q_k \text{ from } F \leq c_2 \text{ (diameter } Q_k) .
\]

3.1.2 Our intention for stating the theorem at this stage is obviously pedagogical. We shall not, strictly speaking, need to apply it until later (Chapter VI), and since its proof is a little intricate we postpone it until that point.