

NONSMOOTH CALCULUS

JUHA HEINONEN

ABSTRACT. We survey recent advances in analysis and geometry, where first order differential analysis has been extended beyond its classical smooth settings. Such studies have applications to geometric rigidity questions, but are also of intrinsic interest. The transition from smooth spaces to singular spaces where calculus is possible parallels the classical development from smooth functions to functions with weak or generalized derivatives. Moreover, there is a new way of looking at the classical geometric theory of Sobolev functions that is useful in more general contexts.

1. INTRODUCTION

The word *nonsmooth* in the title refers both to *functions* and *spaces*. *Calculus* is a field of study where infinitesimal data yields global information. Mathematicians have been practicing calculus with nonsmooth functions for over a century, but only recently in spaces that are not smooth in the traditional sense. In this article, we first survey calculus with nonsmooth functions and then move on to discuss current advances involving singular spaces.

Calculus with nonsmooth functions began in the late nineteenth century with attempts to understand the domain of validity for the fundamental theorem of calculus; these studies were not motivated by immediate practical applications. By the 1930s it had become clear that such a calculus was needed in partial differential equations, sciences, and engineering. The so-called direct methods in the calculus of variations argued for the existence of a solution (to a variational problem) which *a priori* was not smooth; the solution was found via a compactness argument in a complete function space. Regularity theory was then developed to show that *a posteriori* the solution is, under favorable circumstances, infinitely many times differentiable or even real analytic. It was also discovered that in many other cases such smoothness cannot be expected, and the concept of a generalized (or weak, distributional) solution became established in mathematics. The relevant function spaces are the *Sobolev spaces*, whose development we briefly review in this article.

Received by the editors November 7, 2005, and, in revised form, June 16, 2006.

2000 *Mathematics Subject Classification*. Primary 28A75, 49J52, 53C23, 51-02.

This paper constitutes an expanded version of the AMS invited address given by the author in Boulder, Colorado, in October 2003.

The author is grateful for the support and hospitality of MSRI and UC Berkeley, where the bulk of this paper was prepared during a visit in 2002-2003. Supported also by NSF grants DMS 0353549 and DMS 0244421.

©2007 American Mathematical Society
Reverts to public domain 28 years from publication

Calculus in nonsmooth spaces was first developed in particular cases in order to solve particular mathematical problems. Systematic theories to that end were developed later.

By the end of the 1970s it was fully recognized that much contemporary real analysis requires little structure on the underlying space. Maximal functions, differentiation theorems, functions of bounded mean oscillation, Hardy spaces, and so on, make sense in *spaces of homogeneous type*. The latter are (quasi-) metric spaces equipped with a doubling Borel measure. A nontrivial Borel measure on a metric space is said to be *doubling* if the mass of a ball (linearly) controls the mass of its double in a uniform manner. Analysis on spaces of homogeneous type is now a well developed field, with applications to many areas of mathematics.

Spaces of homogeneous type are too general to allow for calculus as defined in the beginning of this article. To that end, stronger conditions from the underlying space are required. Recent research has shown that these conditions are still rather weak. They are satisfied by spaces that have been with us for some time, but also by spaces of exotic hitherto unknown geometries. The principal new requirement is a *Poincaré type inequality*, which expresses global control of a function in terms of its derivative. Other conditions of a more geometric nature can be shown to imply such an inequality. No “final axioms” have been found; the situation is not as sharply defined as in the case of spaces of homogeneous type.¹

From a geometric point of view, spaces of homogeneous type offer no surprises. For a complete metric space, there is no obstruction for the existence of a doubling measure on it, provided that the trivial necessary condition that the space has a finite uniform covering (Assouad) dimension is satisfied. In particular, every closed subset of Euclidean space admits a doubling measure. On the other hand, the validity of a Poincaré inequality in a space begets some hidden structure. This is manifest in the remarkable fact, proved by Cheeger, that a metric space with doubling measure and a Poincaré inequality possesses a generalized cotangent structure consisting of a measurable vector bundle together with a (nontrivial) derivation from the algebra of Lipschitz functions to the bounded sections of the bundle. One can show, for example, that many of the classical fractal sets cannot possess such a derivation. Moreover, there is a striking differentiation theory of Lipschitz functions on such spaces also developed by Cheeger. This theory shows that, under the conditions of a doubling measure and a Poincaré inequality, there is strong stability in the infinitesimal behavior of Lipschitz functions at almost every point. As we do not have in such general situations standard model (linear) tangent spaces and maps between them, many fundamental notions have to be revisited.

Metric spaces with doubling measure and Poincaré inequality admit first order differential calculus akin to that in Euclidean space. Sobolev spaces and second order partial differential equations make sense in such contexts, and an elliptic regularity theory along the lines of De Giorgi and Moser can be developed. It is important to realize that all this can be carried out in spaces that do not resemble classical (Riemannian) spaces. There are compact metric spaces of Hausdorff dimension any prescribed real number larger than one whose first order calculus is in many ways indistinguishable from that of Euclidean space. (As alluded to earlier,

¹In this case, too, the situation has turned out to be more complex than previously thought; there have been recent breakthroughs in harmonic analysis on spaces with measures that are not necessarily doubling. See [139] and the references therein.

classical fractals do not serve as examples here, but more elaborate spaces need to be constructed to this end.)

As mentioned earlier, precise conditions that allow for a first order calculus do not seem clearly definable; at least the question remains open. For example, Keith has subsequently shown that measurable cotangent structures as in Cheeger's work exist also for spaces where conditions are imposed only locally or infinitesimally. On the other hand, the differentiation theory seems to require more global assumptions. (The Poincaré inequality is a semi-global condition.) An important open problem is to understand what exactly is needed for the conclusions in Cheeger's work. For applications, it is also important to know if there is, perhaps under some favorable extra hypotheses, more hidden structure in spaces with Poincaré inequality than what is currently known.

The main purpose of this article is to explain in some detail the concepts and the scope of the new nonsmooth calculus as introduced in the preceding. In order to fully appreciate the many facets of this theory, we begin with a review of the classical development. Special emphasis throughout is put on the notion of the modulus of a curve family that can be used to develop a coordinate free approach to weak differentiation.

Essays like this one are necessarily biased towards the writer's interests and expertise. Among the major omissions here are the fractional Sobolev or Besov spaces, both the classical aspects and the various generalizations to metric spaces. I will also not consider heat kernels and Laplacians on self-similar fractals; the goals and strategies of this topic of current interest are somewhat different from ours. An area that is more related to the current text, but is omitted, is the growing field of geometric measure theory in metric spaces; it would take us too far afield to discuss these interesting developments.

Some historical references appear in the text, but in most cases the sources are secondary. This is not meant to be an article on the history of mathematical analysis.

Finally, rather than interrupting the mathematical narrative with references and credits, for the most part the latter have been collected at the end of each section under **Notes**. (Exceptions occur, especially towards the end of the article.) Although the bibliography is extensive, it obviously cannot claim completeness. I apologize in advance to anyone who has been inadvertently ignored or misrepresented.

1.1. Notes. Among the classic treatises on weakly differentiable functions and their applications are [136], [169], [153]. For foundational treatments of analysis on spaces of homogeneous type, see [50], [51]. An abstract Poincaré inequality in a metric measure space as discussed in this article was formulated in [85], [86]. The work by Cheeger referred to in this introduction is [42]. For a subsequent work by Keith that was also cited earlier, see [100]. Recent monographs and surveys on analysis on metric spaces include [12], [160], [81], [13]. For the relevant literature on related but omitted topics, see [105], [180], [10], [9], and the numerous references in these works.

Acknowledgements. I would like to thank Stephen Semmes both for being a tireless advocate of the kind of mathematics discussed in this text and for years of intensive and enjoyable discussions about these topics. In addition, I have had

extensive discussions about technical as well as general points related to the material here with Mario Bonk, Jeff Cheeger, Piotr Hajlasz, Stephen Keith, Tero Kilpeläinen, Juha Kinnunen, Bruce Kleiner, Pekka Koskela, Olli Martio, Nageswari Shanmugalingam, Dennis Sullivan, Jeremy Tyson, and many others.

I am grateful to Jeff Cheeger for his criticism of an earlier version of this article that led to significant improvements and additions. Many thanks also to Bruce Hanson, David Herron, and Pekka Pankka, who read the manuscript and made useful comments.

Finally, I thank the referees for their careful reading of the paper and suggestions.

2. CLASSICAL CALCULUS

The independent discovery of calculus by Newton and Leibniz in the latter part of the seventeenth century is one of the greatest mathematical achievements of all time. The *fundamental theorem of calculus* stands as beautiful today as it did over three centuries ago; to wit,

$$(2.1) \quad f(b) - f(a) = \int_a^b f'(x) dx.$$

The theorem expresses the fact that by integrating the derivative of a function (which is infinitesimal data), one derives global information about the function itself. That is, if we know the derivative of a function at every point, then we know the change in the values of the function from one end point of an interval to the other. (This entails that we know how to compute the integral in question, but here we are interested in what, in principle, can be known.)

For over two centuries, formula (2.1) was not challenged; it served everyone well. When more attention was paid to the foundations of analysis in the late nineteenth century, many started wondering about the limitations of the fundamental theorem of calculus. Nowadays, using Lebesgue integration, the values of a function need to be known only “almost everywhere”. So what exactly suffices for (2.1) to hold?

The Cantor function $c : [0, 1] \rightarrow [0, 1]$ is a continuous nondecreasing surjective function which is constant in each complementary interval of the standard ternary Cantor set; in particular, c is almost everywhere differentiable with $c' = 0$. Therefore, the fundamental theorem of calculus cannot hold for the Cantor function c . This, and various other examples that became known early on, showed that there is more required for the formula in (2.1) than the mere almost everywhere existence and integrability of the derivative.

One of the first accomplishments of Lebesgue’s theory of integration was the clarification of the foundational issues around the fundamental theorem of calculus. The results are discussed in the next section.

2.1. Notes. One can read about the early history of infinitesimal calculus for example in [33, Chapter 17]. The preface to the book by Saks [151] contains some well known quotes by Poincaré and Hermite related to the perceived futility of the study of nonsmooth functions. Among the many texts on real analysis, the book [38] has a nice discussion of the early examples of functions for which the fundamental theorem of calculus does not hold when the right hand side of (2.1) is interpreted as a Riemann integral. See, in particular, the examples due to Volterra and Pompeiu [38, p. 54 and p. 205].

3. ABSOLUTELY CONTINUOUS FUNCTIONS OF ONE VARIABLE

The following theorem completely characterizes the situations where (2.1) holds.

Theorem 3.1. *Let $f : [a, b] \rightarrow \mathbb{R}$ be a continuous function. Then f is differentiable almost everywhere with integrable derivative such that (2.1) holds if and only if f is absolutely continuous.*

Recall that a function $f : [a, b] \rightarrow \mathbb{R}$ is *absolutely continuous* if for every $\epsilon > 0$ there is $\delta > 0$ such that

$$(3.1) \quad \sum_{i=1}^n |a_i - a_{i+1}| < \delta \Rightarrow \sum_{i=1}^n |f(a_i) - f(a_{i+1})| < \epsilon$$

whenever $[a_1, a_2], \dots, [a_n, a_{n+1}]$ are nonoverlapping subintervals of $[a, b]$.

Lipschitz functions are important examples of absolutely continuous functions. Recall that a function $f : [a, b] \rightarrow \mathbb{R}$ is *Lipschitz* if there exists $L > 0$ such that

$$(3.2) \quad |f(x) - f(y)| \leq L|x - y|$$

for all $x, y \in [a, b]$. To produce an absolutely continuous function that is not Lipschitz, we pick an arbitrary integrable function g on $[a, b]$ and define

$$(3.3) \quad f(x) = \int_a^x g(t) dt.$$

Then f is absolutely continuous, but Lipschitz only if g is bounded. In fact, up to an additive constant, all absolutely continuous functions are of the form (3.3) with $g = f'$ almost everywhere. Lipschitz functions are precisely those for which g is bounded.

It is straightforward to verify that the sum and the product of two absolutely continuous functions on $[a, b]$ are again absolutely continuous and that the usual rules of differentiation apply (in the almost everywhere sense). Most notably, the *integration by parts* formula holds: if $f, g : [a, b] \rightarrow \mathbb{R}$ are absolutely continuous, then

$$(3.4) \quad \int_a^b f'(x)g(x) dx = - \int_a^b f(x)g'(x) dx + f(b)g(b) - f(a)g(a).$$

Thus, absolutely continuous functions are precisely the functions of one real variable that allow for calculus. These functions need not be smooth; their graphs can have corners or other bad behavior that prevent pointwise differentiability at many points, as the representation in (3.3) shows. For example, for any given set E of measure zero in $[a, b]$, there exists a Lipschitz function $f : [a, b] \rightarrow \mathbb{R}$ that fails to be differentiable at every point in E .

To place absolutely continuous functions in a proper context, it is necessary to recall another important concept of classical real analysis. A function $f : [a, b] \rightarrow \mathbb{R}$ is said to be of *bounded variation* if

$$(3.5) \quad \sup \sum_{i=1}^n |f(a_i) - f(a_{i+1})| < \infty,$$

where the supremum is taken over all nonoverlapping subintervals $[a_1, a_2], \dots, [a_n, a_{n+1}]$ of $[a, b]$. Functions of bounded variation are precisely the functions that can be written as a difference of two increasing functions. Absolutely continuous functions are of bounded variation, but not conversely; the Cantor function is of

bounded variation but not absolutely continuous. Functions of bounded variation are almost everywhere differentiable, but the fundamental equality (2.1) may fail for them. There is an important criterion that identifies absolutely continuous functions among general functions of bounded variation: *a continuous function of bounded variation is absolutely continuous if and only if it maps sets of measure zero to sets of measure zero.* That the Cantor function fails to be absolutely continuous can be seen using this criterion as well; it maps the Cantor set onto $[0, 1]$.

There is one more important way to think about absolutely continuous functions of one variable. It follows from (3.4) that if $f : [a, b] \rightarrow \mathbb{R}$ is absolutely continuous, and if $\varphi : [a, b] \rightarrow \mathbb{R}$ is a smooth function that vanishes at the end points of the interval, then

$$(3.6) \quad \int_a^b f'(x)\varphi(x) dx = - \int_a^b f(x)\varphi'(x) dx .$$

The converse is true as well, in the following improved form, where continuity need not be assumed: if $f : [a, b] \rightarrow \mathbb{R}$ is an integrable function such that

$$(3.7) \quad \int_a^b g(x)\varphi(x) dx = - \int_a^b f(x)\varphi'(x) dx$$

for some integrable $g : [a, b] \rightarrow \mathbb{R}$ and for all infinitely many times differentiable, or *smooth*, functions $\varphi : [a, b] \rightarrow \mathbb{R}$ such that $\varphi(a) = 0 = \varphi(b)$, then f agrees almost everywhere with an absolutely continuous function; moreover, $f' = g$ almost everywhere in this case.

Using the language of the theory of distributions, the preceding fact can be stated as follows: *a (continuous) function is absolutely continuous if and only if its distributional derivative is represented by integration against an integrable function.* We discuss distributions briefly at the end of Section 4.

3.1. Notes. The facts presented in this section typically belong to a first graduate course in real analysis and measure theory. They can be found in all good textbooks on the subject. The classic treatise by Saks [151] is still worth reading. A nice recent treatment of real analysis of one variable, with many historical comments and examples, is the book [38].

4. ABSOLUTE CONTINUITY IN MANY VARIABLES

What is the right definition of absolute continuity for a function of more than one variable? More to the point, what is the largest class of functions of many variables that allow for calculus? In this section, we attempt to answer these questions in a coordinate dependent manner. The coordinate free discussion is deferred to Section 7.

We let \mathbb{R}^n denote Euclidean n -space, $n \geq 1$. Lebesgue n -measure is denoted by m_n and we often abbreviate $dm_n(x) = dx$. For brevity, we consider only functions that are defined on all of \mathbb{R}^n , although much of the discussion remains valid for functions defined on an open subset of \mathbb{R}^n .

Let us begin with the following definition.

Definition 4.1. A function $u : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be *absolutely continuous on lines* if u is absolutely continuous on almost every line segment parallel to the coordinate axes.

The definition should be clear: a function is absolutely continuous on lines if the collection of all line segments, parallel to any given coordinate axis, on which the function fails to be absolutely continuous projects to a set of Lebesgue $(n - 1)$ -measure zero on the subspace orthogonal to the given axis.

When $n = 1$, we recover the earlier definition given in Section 3. If $n \geq 2$, functions absolutely continuous on lines need not be continuous. The function

$$(4.1) \quad u_\alpha(x) = \frac{1}{|x|^\alpha}, \quad \alpha > 0,$$

is absolutely continuous on lines in \mathbb{R}^n , $n \geq 2$, but not continuous at the origin. (The value at the origin can be assigned arbitrarily.) Note that u_α is locally integrable if and only if $\alpha < n$. Similarly, if (a_i) is a dense countable set in \mathbb{R}^n , $n \geq 2$, then the function

$$(4.2) \quad u(x) = \sum_i 2^{-i} \cdot |x - a_i|^{-\alpha}$$

is absolutely continuous on lines and nowhere continuous (with arbitrary assignment of values at points a_i); it is locally integrable if and only if $\alpha < n$.

If u is absolutely continuous on lines, then the partial derivatives

$$\partial_i u := \frac{\partial u}{\partial x_i}, \quad i = 1, \dots, n,$$

exist almost everywhere in \mathbb{R}^n ; in particular, the *gradient*

$$(4.3) \quad \nabla u := (\partial_1 u, \dots, \partial_n u)$$

is a well defined function almost everywhere in this case. This follows from the one variable result (Theorem 3.1) and Fubini's theorem. For the function in (4.1) we have that

$$\nabla u_\alpha(x) = -\alpha x |x|^{-\alpha-2}, \quad x \neq 0.$$

Note that ∇u_α is locally integrable if and only if $\alpha < n - 1$.

The preceding examples show the difference between dimensions $n = 1$ and $n \geq 2$. An absolutely continuous function in \mathbb{R} is continuous with locally integrable derivative; a function absolutely continuous on lines in \mathbb{R}^n , $n \geq 2$, need not be continuous or even integrable, and even if it is integrable, its gradient as defined in (4.3) need not be integrable. This means that Definition 4.1 needs to be strengthened so as to become amenable to calculus.

Definition 4.2. Let $1 \leq p < \infty$. A real-valued function $u \in L^p(\mathbb{R}^n)$ is said to of class $\text{ACL}_p(\mathbb{R}^n)$ if it is absolutely continuous on lines with $\nabla u \in L^p(\mathbb{R}^n)$.

Note that the membership in the class $\text{ACL}_p(\mathbb{R}^n)$ as defined here requires three conditions from a function: absolute continuity on lines, and p -integrability for both the function and its gradient (where the latter is defined as in (4.3)). For functions of one variable, the last two requirements (more precisely, their local versions) follow from the first.

For the sake of simplicity, the p -integrability condition is imposed both on the function and its gradient. For much of the ensuing discussion, it would be enough to assume local integrability in both cases. Either way, the measurability of both u and ∇u is implied.

Functions as in (4.2) (when properly modified near infinity so as to guarantee global integrability) show that members in the class $\text{ACL}_p(\mathbb{R}^n)$ need not be differentiable anywhere. By using the one-dimensional result and Fubini's theorem, we have the following integration by parts for functions in $\text{ACL}_p(\mathbb{R}^n)$.

Theorem 4.3. *Let $u \in \text{ACL}_p(\mathbb{R}^n)$ and let $\varphi \in C_0^\infty(\mathbb{R}^n)$. Then*

$$(4.4) \quad \int_{\mathbb{R}^n} \partial_i u \varphi \, dx = - \int_{\mathbb{R}^n} u \partial_i \varphi \, dx$$

for each $i = 1, \dots, n$.

It follows from Theorem 4.3 that the (almost everywhere defined) partial derivatives $\partial_i u$ of a function u in the class $\text{ACL}_p(\mathbb{R}^n)$ are the distributional partial derivatives of the distribution determined by u . It is now appropriate to recall the basics of distribution theory.

4.1. Distributions. The vector space $C_0^\infty(\mathbb{R}^n)$, consisting of all infinitely many times differentiable functions with compact support in \mathbb{R}^n , can be given a natural topology so that addition and scalar multiplication are continuous operations. A linear map

$$T : C_0^\infty(\mathbb{R}^n) \rightarrow \mathbb{R}$$

that is continuous in this topology is called a *distribution*. In fact, T is continuous if and only if for each compact set $K \subset \mathbb{R}^n$ there exist a number $C > 0$ and an integer $m \geq 0$ such that

$$|T(\varphi)| \leq C \max_{x \in K, |\alpha| \leq m} |\partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} \varphi(x)|$$

for all functions $\varphi \in C_0^\infty(\mathbb{R}^n)$ with support in K and for all multi-indices $\alpha = (\alpha_1, \dots, \alpha_n)$ with length $|\alpha| = \alpha_1 + \dots + \alpha_n \leq m$. Every distribution T has an i th partial derivative $\partial_i T$, which is another distribution with action

$$\partial_i T(\varphi) := -T(\partial_i \varphi).$$

A locally integrable function u determines a distribution via integration,

$$\varphi \mapsto \int_{\mathbb{R}^n} u \varphi \, dx,$$

with partial derivatives

$$\varphi \mapsto - \int_{\mathbb{R}^n} u \partial_i \varphi \, dx.$$

The statement made after equation (4.4) is now clear: the distributional i th partial derivative of a function u in $\text{ACL}_p(\mathbb{R}^n)$ is the distribution that is determined by the locally integrable function $\partial_i u$.

The formal theory of distributions was preceded both by the theory of functions that are absolutely continuous on lines, as discussed above, and by the theory of weak derivatives, which will be discussed in the next section.

4.2. Notes. Functions that are absolutely continuous on lines were considered early on in Lebesgue's theory, especially by the Italian school. Among the first papers were [126], [177]. There are several excellent sources for the theory of distributions, for example [153], [150], [93].

5. CLASSICAL SOBOLEV SPACES

For functions in the class $\text{ACL}_p(\mathbb{R}^n)$, the integration by parts formula (4.4) is a conclusion. In the theory of Sobolev spaces, it is taken as a definition.

Let $1 \leq p < \infty$ and let $u \in L^p(\mathbb{R}^n)$.² Then u is said to belong to the *Sobolev space* $W^{1,p}(\mathbb{R}^n)$ if there exists, for each $i = 1, \dots, n$, a function $v_i \in L^p(\mathbb{R}^n)$ such that the distributional i th partial derivative of u is determined by v_i via integration; that is,

$$(5.1) \quad \int_{\mathbb{R}^n} v_i \varphi \, dx = - \int_{\mathbb{R}^n} u \partial_i \varphi \, dx$$

for each $\varphi \in C_0^\infty(\mathbb{R}^n)$. It is easy to see that such a function v_i , if it exists, is unique as an L^p -function, and we set

$$(5.2) \quad \partial_i u := v_i.$$

Notation notwithstanding, *a priori* the function $\partial_i u$ has nothing to do with the partial derivative of u . Let us discuss how appropriate this notation is.

Every function in $\text{ACL}_p(\mathbb{R}^n)$ is in $W^{1,p}(\mathbb{R}^n)$ essentially by definition (and Theorem 4.3), and for functions in $\text{ACL}_p(\mathbb{R}^n)$ notation (5.2) certainly is justified. Now the definition for a Sobolev function entails only that it belongs to $L^p(\mathbb{R}^n)$, and so is strictly speaking an equivalence class of functions whose values can be assigned arbitrarily in sets of measure zero. In particular, a function in $W^{1,p}(\mathbb{R}^n)$ can have a representative that is not absolutely continuous on lines. For example, one can take a function and change its values on a hyperplane. But it would be silly to do so. The following theorem shows that there is no essential difference in the two approaches introduced above.

Theorem 5.1. *Every function in $W^{1,p}(\mathbb{R}^n)$ agrees almost everywhere with a function in $\text{ACL}_p(\mathbb{R}^n)$.*

We will prove Theorem 5.1 later in a stronger form (Theorem 7.4). Before this, let us discuss an alternative way to define Sobolev spaces.

The Sobolev space $W^{1,p}(\mathbb{R}^n)$ is a vector space, and it can be normed by

$$(5.3) \quad \|u\|_{1,p} := \|u\|_{L^p(\mathbb{R}^n)} + \|\nabla u\|_{L^p(\mathbb{R}^n)}.$$

It is easy to see that the norm in (5.3) is complete, making $W^{1,p}(\mathbb{R}^n)$ into a Banach space. There is an indirect but succinct description of the Sobolev space as the closure of smooth functions under the norm (5.3). Namely, consider the vector subspace of $C^\infty(\mathbb{R}^n)$ consisting of all functions $\varphi \in C^\infty(\mathbb{R}^n)$ with $\|\varphi\|_{1,p} < \infty$. In this case, the gradient can be understood in classical terms. The completion of this normed space as a metric space is denoted by $H^{1,p}(\mathbb{R}^n)$. Because $W^{1,p}(\mathbb{R}^n)$ is a Banach space, the space $H^{1,p}(\mathbb{R}^n)$ is (isometrically) contained in $W^{1,p}(\mathbb{R}^n)$. It is not difficult to show that in fact equality holds.

Theorem 5.2.

$$H^{1,p}(\mathbb{R}^n) = W^{1,p}(\mathbb{R}^n).$$

Theorem 5.2 implies the useful fact that every function in $W^{1,p}(\mathbb{R}^n)$ can be approximated in the norm (5.3) by smooth, infinitely many times differentiable functions. This approximation is accomplished by a straightforward mollification

²For simplicity, we leave out the case $p = \infty$, as here and there it would require a separate treatment. See, however, Remark 6.1 (b).

argument as follows. Fix a nonnegative function $\eta \in C_0^\infty(\mathbb{R}^n)$ with total integral one, $\epsilon > 0$, and let $u \in W^{1,p}(\mathbb{R}^n)$. Consider the convolution

$$(5.4) \quad u_\epsilon(x) := u * \eta_\epsilon(x) = \int_{\mathbb{R}^n} u(y) \eta_\epsilon(x-y) dy,$$

where

$$\eta_\epsilon(x) := \epsilon^{-n} \eta(x/\epsilon).$$

Then $u_\epsilon \in C^\infty(\mathbb{R}^n)$ with $\partial_i u_\epsilon = u * \partial_i \eta_\epsilon$, and it is not difficult to show that both $u_\epsilon \rightarrow u$ and $\partial_i u_\epsilon \rightarrow \partial_i u$ in $L^p(\mathbb{R}^n)$ as $\epsilon \rightarrow 0$.

To summarize, the elements in the Sobolev space $W^{1,p}(\mathbb{R}^n)$ can be viewed in three different ways: as p -integrable functions with p -integrable distributional partial derivatives; as p -integrable functions that are absolutely continuous on lines with p -integrable partial derivatives; or as limits of smooth functions in the norm (5.3).

5.1. Sobolev spaces with other measures. There is one subtlety in the preceding definition for $H^{1,p}(\mathbb{R}^n)$ that is easily overlooked. Namely, $H^{1,p}(\mathbb{R}^n)$ consists of equivalence classes of Cauchy sequences of smooth functions in the norm (5.3). A Cauchy sequence (φ_i) in $H^{1,p}(\mathbb{R}^n)$ gives rise to two Cauchy sequences in $L^p(\mathbb{R}^n)$, the sequences (φ_i) and $(\nabla \varphi_i)$. There are unique L^p -limits φ and v for these two sequences, where $v : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a vector valued L^p -function. Obviously, two equivalent Cauchy sequences produce the same L^p -limit φ . By using integration by parts, it is straightforward to see that v is also independent of the particular choice of the Cauchy sequence (φ_i) amongst all equivalent Cauchy sequences. It follows that we can assign a unique “gradient” to each element of $H^{1,p}(\mathbb{R}^n)$ by setting $\nabla u := v$.

A similar construction makes sense in situations where integration by parts is not available. For example, we can replace Lebesgue measure in (5.3) by an arbitrary Radon measure μ in \mathbb{R}^n , and consider the norm

$$(5.5) \quad \|\varphi\|_{1,p,\mu} := \left(\int_{\mathbb{R}^n} |\varphi|^p d\mu \right)^{1/p} + \left(\int_{\mathbb{R}^n} |\nabla \varphi|^p d\mu \right)^{1/p}$$

for every $\varphi \in C^\infty(\mathbb{R}^n)$ such that the quantity in (5.5) is finite. The closure $H^{1,p}(\mathbb{R}^n; \mu)$ of smooth functions under the norm (5.5) is a kind of Sobolev space. Its members are equivalence classes of Cauchy sequences of functions in $L^p(\mathbb{R}^n; \mu)$, but in general one cannot assign a unique gradient to such a Cauchy sequence; in particular, the space $H^{1,p}(\mathbb{R}^n; \mu)$ is not a space of functions in general. By using the language of functional analysis, asking for the uniqueness is equivalent to asking if the operator $\nabla : C^\infty(\mathbb{R}^n) \cap H^{1,p}(\mathbb{R}^n; \mu) \rightarrow L^p(\mathbb{R}^n; \mu)$ is *closable*.

For example, a Sobolev space $H^{1,2}(\mathbb{R}^n; \mu)$ arises as a natural space for solutions to degenerate elliptic equations of the type

$$(5.6) \quad -\operatorname{div}(w(x)\nabla u(x)) = 0, \quad d\mu(x) = w(x) dx,$$

for some locally integrable nonnegative function or *weight* w in \mathbb{R}^n . A solution to (5.6) is a function $u \in H_{\text{loc}}^{1,2}(\mathbb{R}^n; \mu)$ that satisfies

$$(5.7) \quad \int_{\mathbb{R}^n} w(x)\nabla u(x) \cdot \nabla \varphi dx = 0$$

for each $\varphi \in C_0^\infty(\mathbb{R}^n)$. In other words, the *divergence* of the (vector-valued) distribution $w(x)\nabla u(x)$ is zero. However, before one can speak about equation (5.6), it should be decided what ∇u means for elements $u \in H^{1,2}(\mathbb{R}^n; \mu)$. This is the closability problem for the gradient just discussed.

The author does not know whether the class of measures with the preceding closability property has been characterized by other means. For sufficient conditions, see [83], [79, Section 13]. For an example of a measure that is absolutely continuous with respect to Lebesgue measure but for which the gradient operator is not closable, see [63, pp. 91–92].

Remark 5.3. For brevity, we have restricted our definition for Sobolev functions and Sobolev spaces to functions defined on all of \mathbb{R}^n . Much of what we have said goes over to the Sobolev spaces $W^{1,p}(\Omega)$, where $\Omega \subset \mathbb{R}^n$ is an arbitrary open set. We also ignore the higher order Sobolev spaces in this article.

5.2. Notes. The theory of Sobolev spaces dates from the 1930s; see [167], [168]. Brief historical comments can be found for example in [136, Ch. 1.8, p. 19] and [134, p. 29]. Among the excellent modern sources for the material in this section are [4], [62], [192]. For the theory of Sobolev spaces with weights, see [83].

6. SOBOLEV-POINCARÉ INEQUALITIES

The fundamental theorem of calculus gives global information about a function after integrating its derivative. There are important several variable versions of this phenomenon; these are the various *Sobolev-Poincaré inequalities*.

Let u be a smooth real-valued function on \mathbb{R}^n . Fix two points $x, y \in \mathbb{R}^n$. We can apply the fundamental theorem of calculus on the line segment $[x, y]$ connecting x and y , and obtain from the chain rule that

$$(6.1) \quad u(y) - u(x) = \int_0^1 \nabla u(ty + (1-t)x) \cdot (y-x) dt.$$

Inserting absolute values in (6.1) yields the estimate

$$|u(y) - u(x)| \leq \int_{[x,y]} |\nabla u| ds$$

for the oscillation of the function at the end points in terms of a scalar line integral. The above reasoning holds not only for the straight line segment from x to y but for every rectifiable path γ joining the two points; we have the estimate

$$(6.2) \quad |u(y) - u(x)| \leq \int_\gamma |\nabla u| ds$$

for each such γ . Now comes a technical point, which however should be conceptually clear. Namely, by averaging over a suitable family of curves γ , the right hand side in (6.2) can be turned into a volume integral over all of \mathbb{R}^n . After some manipulation, we obtain

$$(6.3) \quad |u(x) - u(y)| \leq C(n) \left(\int_{\mathbb{R}^n} \frac{|\nabla u(z)|}{|z-x|^{n-1}} dz + \frac{|\nabla u(z)|}{|z-y|^{n-1}} dz \right),$$

where $C(n) > 0$ is a dimensional constant.

The exponent $n - 1$ in the denominator of (6.3) can be explained as follows. Near both points x and y , one considers a radial family of curves over which (6.2) is averaged; a polar coordinate computation in a small patch near x , say, changes

$$\int \frac{|\nabla u(z)|}{|z - x|^{n-1}} dz$$

into a number of line integrations over radial segments residing in this patch.

Inequality (6.3) can be thought of as a substitute for the fundamental theorem of calculus for functions in \mathbb{R}^n . It is called a *Riesz potential estimate*. The expression in the parentheses on the right in (6.3) is a sum of two convolutions with the *Riesz kernel* $|z|^{1-n}$.

We emphasize the fact that in estimate (6.2), and consequently in (6.3), only the *pointwise norm* $|\nabla u|$ of the gradient ∇u is used. The same holds in the ensuing fundamental inequalities. This fact is a basis for the various generalizations of the Sobolev-Poincaré type inequalities that we will discuss later in this article; there nonnegative functions are used that are to be thought of as pointwise defined bounds for the *size* of a derivative. (See Section 11.)

Returning to Riesz potentials, the convolution integral operator

$$f \mapsto |z|^{1-n} * f =: I_1 f$$

has the following mapping properties:

$$(6.4) \quad \|I_1 f\|_{L^{np/(n-p)}(\mathbb{R}^n)} \leq C(n, p) \|f\|_{L^p(\mathbb{R}^n)}, \quad 1 < p < n,$$

and

$$(6.5) \quad \|I_1 f\|_{L^\infty(B)} \leq C(n, p) (\text{diam}(B))^{1-n/p} \|f\|_{L^p(B)}, \quad n < p \leq \infty,$$

where in estimate (6.5) it is moreover assumed that f is supported in a ball B in \mathbb{R}^n . The second estimate (6.5) is a straightforward consequence of Hölder's inequality. The proof for the first estimate (6.4) is more complicated involving maximal functions.

We forgo, for a moment, estimates for the borderline case $p = n$ (see Remark 6.1 (a)).

Estimates (6.4) and (6.5) can be used to obtain integral bounds for functions in a Sobolev space.

To that end, we first explain one more technical point. When the points x and y are given and contained in an open ball B in \mathbb{R}^n , the argument leading to (6.3) can be made so as to use curves that lie in B . In particular, we can restrict the gradient $|\nabla u|$ to the ball B , whence the estimate

$$(6.6) \quad |u(x) - u(y)| \leq C(n) (I_1(|\nabla u| \cdot \chi_B)(x) + I_1(|\nabla u| \cdot \chi_B)(y)),$$

where χ_B is the characteristic function of B .

By using estimates (6.6) and (6.4), and integrating over x and y separately, we arrive at the following *Sobolev-Poincaré inequality*, valid for $1 < p < n$:

$$(6.7) \quad \left(\int_B |u - u_B|^{np/(n-p)} dx \right)^{(n-p)/np} \leq C(n, p) (\text{diam}(B)) \left(\int_B |\nabla u|^p dx \right)^{1/p},$$

where u_B denotes the *average* of u in the ball B , as does the barred integral sign:

$$(6.8) \quad v_B := \int_B v := \frac{1}{m_n(B)} \int_B v dx.$$

(An analogous notation will be used later in other metric spaces and for other measures.) In (6.7), the pointwise estimates have all disappeared, and hence by an approximation argument using convolutions as in (5.4) we conclude that the *Sobolev-Poincaré inequality* (6.7) is valid for each function $u \in W^{1,p}(\mathbb{R}^n)$ whenever B is a ball in \mathbb{R}^n and $1 < p < n$.

In fact, (6.7) is true also when $p = 1$ by a different argument which we forgo here.

For future reference, let us record the following consequence of (6.7), known simply as the *Poincaré inequality*. With notation as in (6.7), we have that

$$(6.9) \quad \int_B |u - u_B|^p dx \leq C(n,p)(\text{diam}(B))^p \int_B |\nabla u|^p dx,$$

whenever $u \in W^{1,p}(\mathbb{R}^n)$ and $1 \leq p < \infty$.

Let us now turn to the case $p > n$. As before, we obtain from (6.5) that

$$(6.10) \quad |u(x) - u(y)| \leq C(n,p) |x - y|^{1-n/p} \left(\int_B |\nabla u|^p dz \right)^{1/p}$$

whenever B is a ball containing x and y . It follows that u is *Hölder continuous* of order $1 - n/p > 0$. This *a priori* bound for smooth functions can be used, via density, to pass to a similar inequality for functions in $W^{1,p}(\mathbb{R}^n)$ for $p > n$. It follows that *every function in $W^{1,p}(\mathbb{R}^n)$, $p > n$, can be changed in a set of measure zero so as to become Hölder continuous and satisfy (6.10) in every ball B .*

By the discussion in the preceding sections, we know that functions in the Sobolev space $W^{1,p}(\mathbb{R}^n)$ have regularity slightly beyond their mere definition as L^p -functions; they have representatives that are absolutely continuous on almost all lines. In this section, we have seen a different kind of improvement. Namely, for the range $1 \leq p < n$, a function in $W^{1,p}(\mathbb{R}^n)$ is locally integrable to a higher degree than required by the definition. For $p > n$, we even have (Hölder) continuity of the functions in $W^{1,p}(\mathbb{R}^n)$. These statements are independent of the previous discovery of absolute continuity on lines.

Finally, the above estimates and embeddings are all sharp, as can be seen by considering functions that behave like the functions u_α in (4.1) near the origin, for appropriate values α (also for positive α in the case $p > n$).

Remark 6.1. (a) As alluded to earlier, the borderline case $p = n$ is somewhat delicate. The integral operator I_1 maps $L^n(\mathbb{R}^n)$ to $\text{BMO}(\mathbb{R}^n)$, the space of *functions of bounded mean oscillation*. In particular, we get local exponential integrability for the functions in $W^{1,n}(\mathbb{R}^n)$. Such integrability is sharp, because functions with singularities of the type

$$(6.11) \quad \log \log |x|, \quad x \sim 0,$$

belong to $W^{1,n}(\mathbb{R}^n)$. By moving the singularities around as in (4.2), one can exhibit nowhere continuous functions in $W^{1,n}(\mathbb{R}^n)$. The borderline case $p = n$ exhibits many interesting and unique phenomena, but it would take us too far afield to discuss them here.

(b) We have not included the case $p = \infty$ in the preceding discussion. One can show that functions in $W^{1,\infty}(\mathbb{R}^n)$ are Lipschitz continuous (in the sense of having representatives). A converse holds as well: a bounded Lipschitz function $u : \mathbb{R}^n \rightarrow \mathbb{R}$ belongs to $W^{1,\infty}(\mathbb{R}^n)$. The definition for the Sobolev space $W^{1,\infty}(\mathbb{R}^n)$ is as for $W^{1,p}(\mathbb{R}^n)$ using the space $L^\infty(\mathbb{R}^n)$ in place of $L^p(\mathbb{R}^n)$.

6.1. **Notes.** For more information about Sobolev-Poincaré inequalities, potential estimates, and related issues, see, for example, [3], [171], [172], [70], [62], [134], [192].

7. CAPACITY AND MODULUS

To reiterate, functions in $W^{1,p}(\mathbb{R}^n)$ are integrable beyond the initial requirement, and they have representatives that are absolutely continuous on lines. In addition, they can be chosen to be everywhere continuous if $p > n$. For $1 \leq p \leq n$, functions in $W^{1,p}(\mathbb{R}^n)$ need not be continuous anywhere, but there is more that can be said about the continuity in this case, too. Moreover, the absolute continuity on lines can be improved in all cases to include curves other than line segments that are parallel to the coordinate axes. In this section, we discuss such improvements. The goal is a coordinate free description of Sobolev functions.

One improvement in the continuity properties of a Sobolev function can be described by saying that a function in $W^{1,p}(\mathbb{R}^n)$ is continuous when restricted to the complement of a small set. This is a *Lusin type phenomenon*. In the present context, it is called *p-quasicontinuity*; the smallness of the exceptional set is tied to the integrability exponent. Although an important property of Sobolev functions, the quasicontinuity plays only an implicit role in our story. We concentrate more on the improvement of the absolute continuity on lines property; it is this aspect that matters more when the theory is generalized. However, the latter cannot fully be discussed without the former.

7.1. **Capacity.** Let $1 \leq p < \infty$. The *p-capacity* of a subset $E \subset \mathbb{R}^n$ is the number

$$(7.1) \quad \text{cap}_p(E) := \inf \int_{\mathbb{R}^n} (|u|^p + |\nabla u|^p) dx,$$

where the infimum is taken over all functions $u \in W^{1,p}(\mathbb{R}^n)$ such that $u \geq 1$ almost everywhere on a neighborhood of E . If no such function u exists, we set $\text{cap}_p(E) = \infty$. There are various alternatives and generally equivalent ways to define capacity, but we choose this one.

It is not too hard to verify that the set function

$$E \mapsto \text{cap}_p(E)$$

is an outer measure defined on all subsets of \mathbb{R}^n . That is, it takes values in $[0, \infty]$; it is monotone,

$$\text{cap}_p(E) \leq \text{cap}_p(F), \quad E \subset F;$$

it is countably subadditive,

$$\text{cap}_p \left(\bigcup_{i=1}^{\infty} E_i \right) \leq \sum_{i=1}^{\infty} \text{cap}_p(E_i);$$

and it assigns the value zero to the empty set. A subset E of \mathbb{R}^n is said to be of *zero p-capacity* if $\text{cap}_p(E) = 0$. Capacity measures sets in a more refined way than Lebesgue measure. Hausdorff measures and contents are close relatives of capacity, and the following relations are valid:

$$(7.2) \quad \text{cap}_p(E) = 0 \quad \Rightarrow \quad \mathcal{H}_{n-p+\epsilon}(E) = 0 \quad \text{for all } \epsilon > 0,$$

and

$$(7.3) \quad \mathcal{H}_{n-p}(E) < \infty \quad \Rightarrow \quad \text{cap}_p(E) = 0$$

whenever $1 \leq p \leq n$.³ Moreover, singletons have positive p -capacity if $p > n$. Here and later \mathcal{H}_s denotes the s -dimensional Hausdorff measure. (See [133] for the definition.)

Theorem 7.1. *Let $1 \leq p \leq n$. Every function in $W^{1,p}(\mathbb{R}^n)$ has a representative that is p -quasicontinuous in the following sense: for every $\epsilon > 0$ there is an open set $U \subset \mathbb{R}^n$ such that $\text{cap}_p(U) < \epsilon$ and the representative is continuous in $\mathbb{R}^n \setminus U$. Moreover, such a p -quasicontinuous representative is well defined pointwise, and unique, up to an ambiguity of a set of p -capacity zero.*

By combining Theorem 7.1 with (7.2) and (7.3), we have that every function in $W^{1,p}(\mathbb{R}^n)$ has an essentially unique representative with pointwise defined values outside a set of Hausdorff dimension $n - p$. This is a definite improvement for the initially only almost everywhere defined function. It means, among other things, that every function in $W^{1,p}(\mathbb{R}^n)$ has an (essentially) well defined trace on smooth surfaces of dimension larger than $n - p$.

The proofs for the Hausdorff measure estimates in (7.2) and (7.3) are based on covering lemmas and maximal function estimates. The proof for the existence of a quasicontinuous representative follows the standard proof of Lusin’s theorem. The key fact is that the mollified functions u_ϵ converge in $W^{1,p}(\mathbb{R}^n)$. Then one chooses a sufficiently rapidly converging subsequence and computes that this subsequence converges uniformly outside a given open set of arbitrarily small capacity. This leads both to the quasicontinuity and to the quasieverywhere pointwise defined representative.

The uniqueness of the quasicontinuous representative is not obvious. A direct proof of the uniqueness can be given, but it can also be deduced from the following interesting fact: for every function $u \in W^{1,p}(\mathbb{R}^n)$ the limit

$$(7.4) \quad \lim_{r \rightarrow 0} \int_{B(x,r)} u(y) \, dy$$

exists for x outside a set of zero p -capacity, and this limit defines the quasicontinuous representative of u . Thus, in principle, quasicontinuous representatives of Sobolev functions can be identified via (7.4).

We now turn to the second type of improvement of Sobolev functions, promised in the beginning of this section. The principal question is the following: on how many curves is a given Sobolev function $u \in W^{1,p}(\mathbb{R}^n)$ absolutely continuous? If we are content with an arbitrary $\text{ACL}_p(\mathbb{R}^n)$ -representative of u (which exists by Theorem 5.1), then nothing more can be said than what the definition for $\text{ACL}_p(\mathbb{R}^n)$ dictates. Indeed, we can take any set E of n -measure zero such that the projection of E to each of the coordinate hyperplanes has $(n - 1)$ -measure zero and then change the values of u arbitrarily on E . But again, it would be silly to do this.

In the definition for $\text{ACL}_p(\mathbb{R}^n)$ -functions, a fixed coordinate system was used, but membership in a Sobolev space is easily seen to remain intact after linear (say) coordinate changes. Is it reasonable to expect that every Sobolev function has a representative that is, *simultaneously*, absolutely continuous on almost every line, in every given direction in \mathbb{R}^n ? This indeed is the case, and even more is true. To explain the full answer here, it is necessary to introduce the concept of the modulus of a curve family.

³In the borderline case $p = n$, something more precise can be said, but we forgo it here.

7.2. Modulus. Let $\Gamma = \{\gamma\}$ be a family of rectifiable curves in \mathbb{R}^n , where a *curve* in \mathbb{R}^n is a continuous map $\gamma : [a, b] \rightarrow \mathbb{R}^n$, and a curve is *rectifiable* if it is (componentwise) of bounded variation. A Borel measurable function $\rho : \mathbb{R}^n \rightarrow [0, \infty]$ is said to be an *admissible function*, or *density*, for Γ if

$$(7.5) \quad \int_{\gamma} \rho \, ds \geq 1$$

for each $\gamma \in \Gamma$. The *p-modulus* of Γ is the number

$$(7.6) \quad \text{mod}_p(\Gamma) := \inf \int_{\mathbb{R}^n} \rho^p \, dx,$$

where the infimum is taken over all admissible functions ρ . The definition can be extended, in a routine way, to all curve families consisting of locally rectifiable curves. The *p-modulus* of the family of all curves that are not locally rectifiable is defined to be zero.

It is not hard to see that the set function

$$\Gamma \mapsto \text{mod}_p(\Gamma)$$

is an outer measure on the collection of all curve families in \mathbb{R}^n ; it takes values in $[0, \infty]$; is monotone,

$$\text{mod}_p(\Gamma') \leq \text{mod}_p(\Gamma), \quad \Gamma' \subset \Gamma;$$

countably subadditive,

$$\text{mod}_p\left(\bigcup_{i=1}^{\infty} \Gamma_i\right) \leq \sum_{i=1}^{\infty} \text{mod}_p(\Gamma_i);$$

and the empty family has zero modulus.

There is another property that follows directly from the definitions: if Γ' and Γ are two curve families such that every curve in Γ' has a subcurve in Γ , then

$$(7.7) \quad \text{mod}_p(\Gamma') \leq \text{mod}_p(\Gamma).$$

A family Γ is said to be *p-exceptional* if $\text{mod}_p(\Gamma) = 0$. If a property of curves holds outside a *p-exceptional* family, then it is said to hold on *p-almost every curve*.

Property (7.7) implies the following useful result.

Lemma 7.2. *A family of curves is p-exceptional if each curve in the family has a subcurve in a fixed p-exceptional family.*

The next lemma can be used as an alternative definition for *p-exceptionality*. The proof is an easy exercise.

Lemma 7.3. *A family Γ of curves in \mathbb{R}^n is p-exceptional if and only if there exists a Borel function $\rho : \mathbb{R}^n \rightarrow [0, \infty]$ such that $\rho \in L^p(\mathbb{R}^n)$ and*

$$\int_{\gamma} \rho \, ds = \infty$$

for each locally rectifiable $\gamma \in \Gamma$.

We now state the following main relation between modulus and Sobolev functions.

Theorem 7.4. *Every function in $W^{1,p}(\mathbb{R}^n)$ has a representative that is absolutely continuous on p-almost every curve.*

This result provides a coordinate free good representative for a Sobolev function. We will consider the uniqueness and pointwise ambiguity for such representatives after we outline the proof of Theorem 7.4.

The proof is an adaptation of the real analysis argument, which shows that an L^p -convergent sequence has a pointwise almost everywhere convergent subsequence. We have the following general result.

Lemma 7.5 (Fuglede’s lemma). *If a sequence of Borel functions (g_k) converges in $L^p(\mathbb{R}^n)$ to a Borel function g , then there is a subsequence (g_{k_j}) such that*

$$\int_{\gamma} g_{k_j} ds \rightarrow \int_{\gamma} g ds$$

for p -almost every curve γ in \mathbb{R}^n .

Proof. We may assume that $g = 0$ and pick a subsequence (g_{k_j}) such that

$$(7.8) \quad \|g_{k_j}\|_{L^p(\mathbb{R}^n)} \leq 2^{-j}.$$

Fix $l \geq 1$, and let Γ_l denote the family of all curves $\gamma : [a, b] \rightarrow \mathbb{R}^n$ such that there are infinitely many functions g_{k_j} with

$$\int_{\gamma} g_{k_j} ds \geq 1/l, \quad \gamma \in \Gamma_l.$$

By the subadditivity of modulus, it suffices to show that

$$(7.9) \quad \text{mod}_p(\Gamma_l) = 0.$$

Indeed, the function

$$g = \sum_{j=1}^{\infty} g_{k_j}$$

is p -integrable by (7.8) and satisfies

$$\int_{\gamma} g ds = \infty$$

for each $\gamma \in \Gamma_l$. Thus (7.9) follows from Lemma 7.3, and the proof is complete. \square

To prove Theorem 7.4, we use Lemma 7.5. Recall that the mollifications u_{ϵ} of a function $u \in W^{1,p}(\mathbb{R}^n)$ satisfy $u_{\epsilon} \rightarrow u$ and $|\nabla u_{\epsilon}| \rightarrow |\nabla u|$ in $L^p(\mathbb{R}^n)$, as $\epsilon \rightarrow 0$. Now fix an arbitrary Borel representative of $|\nabla u|$ (so that line integrations can be performed). We have that the relations

$$(7.10) \quad |u_{\epsilon}(\gamma(a)) - u_{\epsilon}(\gamma(b))| \leq \int_{\gamma} |\nabla u_{\epsilon}| ds$$

and

$$(7.11) \quad \lim_{\epsilon_j \rightarrow 0} \int_{\gamma} |\nabla u_{\epsilon_j}| ds = \int_{\gamma} |\nabla u| ds$$

hold except for a p -exceptional family of curves $\gamma : [a, b] \rightarrow \mathbb{R}^n$, for some subsequence ϵ_j . In particular, in view of Lemma 7.2, p -almost every curve in \mathbb{R}^n has the property that both (7.10) and (7.11) hold for the curve and for all its subcurves.

If the sequence (u_{ϵ}) converged pointwise to u , then (7.10) and (7.11) would hold true for u and for p -almost every curve γ along with all subcurves of γ . The desired absolute continuity then follows, because $|\nabla u|$ is integrable on p -almost every curve

by Lemma 7.3. The problem one encounters here is that such everywhere pointwise convergence need not hold. However, as explained right after Theorem 7.1, we can pass to a rapidly converging subsequence to assure that pointwise convergence takes place *outside an exceptional set E of zero p -capacity*. The key observation now is that if E has zero p -capacity, then the collection of all curves in \mathbb{R}^n that meet E is p -exceptional. (We omit the proof of this, but see, however, the proof of Theorem 10.5.) This fact accepted, the above argument, where pointwise convergence everywhere was assumed, can be used; we simply ignore one more p -exceptional curve family.

This finishes the outline of the proof for Theorem 7.4.

The preceding proof reveals that for the good representative in Theorem 7.4 one can choose the p -quasicontinuous representative; conversely, one can show that every such representative is p -quasicontinuous and well defined up to a set of p -capacity zero.

We close this section by providing the following alternative, coordinate free description of Sobolev functions in \mathbb{R}^n .

Theorem 7.6. *A function $u \in L^p(\mathbb{R}^n)$ has a representative in $W^{1,p}(\mathbb{R}^n)$ if and only if there exists a Borel function $\rho \in L^p(\mathbb{R}^n)$ such that the inequality*

$$(7.12) \quad |u(\gamma(a)) - u(\gamma(b))| \leq \int_{\gamma} \rho \, ds$$

holds for the representative, for p -almost every curve $\gamma : [a, b] \rightarrow \mathbb{R}^n$.

Proof. The necessity part follows from the proof outlined for Theorem 7.4. One can take ρ to be any Borel representative of $|\nabla u|$.

To prove the sufficiency, one uses Fubini's theorem and proves that for almost all lines parallel to the coordinate axes, the function u is absolutely continuous and satisfies

$$|\partial_i u| \leq \rho$$

almost everywhere on such a line. It follows that the partial derivatives are p -integrable and hence that $u \in W^{1,p}(\mathbb{R}^n)$. \square

A refinement of the final argument in the preceding proof shows that the length of the distributional gradient $|\nabla u|$ is the (almost everywhere) smallest function ρ that satisfies (7.12) for p -almost every curve γ . It follows that the norm in $W^{1,p}(\mathbb{R}^n)$ can be defined as

$$(7.13) \quad \|u\|_{1,p} = \|u\|_{L^p(\mathbb{R}^n)} + \inf \|\rho\|_{L^p(\mathbb{R}^n)},$$

where the infimum is taken over all Borel functions ρ satisfying (7.12) for p -almost every curve γ .

This characterization of Sobolev functions requires no smooth structure of the underlying space. One uses the metric structure for the line integration and measure for the modulus. This lead is followed later when Sobolev spaces in arbitrary metric measure spaces are discussed in Section 10. In preparation for this discussion, we first give examples of nonsmooth spaces that are particularly suitable for a general Sobolev space theory or where such a theory is needed for external reasons.

7.3. Notes. The relationship between various capacities and pointwise behavior of functions was first studied in classical potential theory; see [123], [58]. For general estimates between capacities and Hausdorff measures, see [123], [3], [134]. Elementary proofs for the particular estimates in (7.2) and (7.3) can be found in [83, Chapter 2]. There is a large literature on fine behavior of Sobolev functions; see, for example, [134], [3], [132], and the many references there. For a slick proof of the uniqueness of the quasicontinuous representative in Theorem 7.1, see [106]. Modulus was first used as a tool in geometric function theory, especially by Beurling and Ahlfors [23], [5]. Its usefulness in real analysis was probably first realized by Fuglede [69]; in particular, Theorem 7.4 is due to him. Theorem 7.6 and equality (7.13) are due to Shanmugalingam [163].

8. SINGULAR SPACES

We give definitions for a “smooth” and a “singular” metric space, but for the purposes of this article only. Several examples of singular spaces and their role in analysis and geometry are discussed.

8.1. Smooth spaces. Let us call a metric space *smooth* if it is locally isometric to a finite dimensional Riemannian manifold. Thus, a space is smooth if it locally looks like an open region in some \mathbb{R}^n , equipped with a metric

$$(8.1) \quad d(x, y) := \inf_{\gamma} \int_{\gamma} \sqrt{\langle G\gamma', \gamma' \rangle} ds,$$

where G is a smooth symmetric matrix-valued function on the region, and the infimum is taken over all smooth curves γ joining x and y in the region. We also call a metric space *smooth at a point* if the point has a neighborhood that is smooth.

Metric spaces of this kind were first considered by Riemann in his famed 1854 *Habilitationsschrift*. They are very special metric spaces, as Riemann himself was aware of. Here we quote a passage from Riemann’s lecture (in Spivak’s translation [170, p. 152]):

Still more complicated relations can occur if the line element cannot be represented, as was presupposed, by the square root of a differential expression of the second degree. Now it seems that the empirical notions on which the metric determinations of Space are based, the concept of a solid body and that of a light ray, lose their validity in the infinitely small; it is therefore quite definitely conceivable that the metric relations of Space in the infinitely small do not conform to the hypotheses of geometry; and in fact one ought to assume this as soon as it permits a simpler way of explaining phenomena.

A crucial feature of metrics as in (8.1) is that infinitesimally they become Euclidean. More precisely, if x_0 is a point in a smooth metric space (X, d) , then the pointed metric spaces

$$X_{\epsilon} := (X, \epsilon^{-1}d, x_0)$$

converge, as $\epsilon \rightarrow 0$, to some finite dimensional Euclidean (Hilbert) space. Convergence of metric spaces is an important concept in synthetic geometry, and we will discuss it in more detail in subsection 8.6. In the present context of smooth spaces, the convergence $X_{\epsilon} \rightarrow \mathbb{R}^n$ means that, given any $R > 0$ and any $\lambda > 0$,

there exists $\epsilon_\lambda > 0$ such that the closed ball $\overline{B}_\epsilon(x_0, R)$ in the metric space X_ϵ is $(1 + \lambda)$ -bi-Lipschitz equivalent to the closed ball $\overline{B}(0, R)$ in \mathbb{R}^n whenever $\epsilon < \epsilon_\lambda$.

We now discuss Lipschitz maps.

8.2. Lipschitz maps. Lipschitz maps form the right substitute for smooth maps in metric geometry.

Recall that a map $f : X \rightarrow Y$ between two metric spaces is said to be *L-Lipschitz*, $L \geq 1$, if

$$d_Y(f(a), f(b)) \leq L d_X(a, b)$$

whenever $a, b \in X$. An embedding $f : X \rightarrow Y$ is said to be *L-bi-Lipschitz*, $L \geq 1$, if

$$L^{-1} d_X(a, b) \leq d_Y(f(a), f(b)) \leq L d_X(a, b)$$

whenever $a, b \in X$. We use the terms *Lipschitz*, or *bi-Lipschitz*, if the parameter L is not important for the discussion at hand.

A 1-bi-Lipschitz embedding is called an *isometric embedding*. If there is a bi-Lipschitz bijection between two metric spaces, then these spaces are said to be *bi-Lipschitz equivalent*. Spaces that are bi-Lipschitz equivalent through a 1-bi-Lipschitz map are called *isometric*. The terms *locally bi-Lipschitz equivalent* and *locally isometric* should be clear.

Metric spaces that are locally bi-Lipschitz equivalent to a region in \mathbb{R}^n are called (*n-dimensional*) *Lipschitz manifolds*.⁴

In our definition for the Lipschitz manifold the metric is given first. It is a different issue that a topological manifold may be metrized so as to become a Lipschitz manifold in our sense. (The latter holds in fact for all topological manifolds outside dimension four; see subsection 11.4.) Recall here that a (*topological*) *n-manifold* is a metrizable separable space where every point has a neighborhood that is homeomorphic to \mathbb{R}^n . We will call 2-manifolds *surfaces*.

8.3. Singular spaces. The main message of this article is that calculus is possible in spaces that are not smooth (in the sense defined in subsection 8.1 for example). We restrict our discussion to a first order calculus, which should be preserved by bi-Lipschitz maps. In particular, calculus is possible on Lipschitz manifolds.

The preceding understood, let us call a metric space X *singular at a point* p , $p \in X$, if no neighborhood of p is bi-Lipschitz equivalent to a smooth space. We also say that p is a *singular point of* X in this case. A metric space is *singular* if it has at least one singular point.⁵

Notice that our terms “smooth” and “singular” are not complementary concepts. Easy examples show that a metric space can be simultaneously nowhere singular and nowhere smooth. (Consider the graph of a nowhere smooth Lipschitz function.)

It is in general a difficult task to recognize whether a given point in a metric space is singular or not. Equivalently, it is in general a difficult task to recognize whether a given metric space is a Lipschitz manifold as defined in the preceding. We will return to this question later in subsection 16.5.

We next display some examples of singular spaces and their constructions that appear in analysis and geometry.

⁴Lipschitz manifolds are typically assumed to be separable, in addition, but as we are mainly interested in local questions this amplification is immaterial.

⁵By this definition, every point in an infinite dimensional Banach space is singular, which may be a cause of objection to the terminology. However, the focus of this article is on finite dimensional spaces.

8.4. Alexandrov spaces. These were, quite likely, the first singular spaces to be systematically studied in geometry. There are two essentially different classes of spaces here. The first class is defined via a triangle comparison; it allows one to say whether a space has curvature bounded from either above or below by some constant. The spaces in the second class are topological surfaces, where the total curvature makes sense as a measure.⁶

The starting point for the first class is the model 2-dimensional simply connected Riemannian surface M_k^2 of constant Gaussian curvature k , where k is a given real number. Thus, M_k is the Euclidean plane if $k = 0$, or a properly scaled 2-sphere, or hyperbolic plane, if $k > 0$, or $k < 0$, respectively. Assume now that X is a metric space that is *locally geodesic* in the sense that every point in X has a neighborhood whose points can be joined by a geodesic. A *geodesic* $[a, b]$ joining two points $a, b \in X$ is the image of an isometric map of the interval $[0, d_X(a, b)] \subset \mathbb{R}$ to X .⁷ In such a space, we can speak about a *triangle* $\Delta(a, b, c)$, which is a union of geodesics $[a, b]$, $[b, c]$, and $[c, a]$. In the model space, we have a unique (up to isometry) triangle with side lengths determined by the lengths of the three geodesics. Now X is said to have *curvature at most k* , or *at least k* , if all sufficiently small triangles in X are thinner, or thicker, respectively, than the corresponding model triangle in M_k . More precisely, X has curvature at most k , or at least k , if every point in X has a neighborhood such that for every geodesic triangle $\Delta(a, b, c)$ in the neighborhood the distance from every point $z \in [a, b]$ to c is at most, or at least, the distance between the corresponding points in the model triangle in M_k .

(In the general theory of spaces with bounded curvature, as defined in the preceding, some additional a priori assumptions are typically placed on the metric space X . For example, a standard assumption is that X be a length space.⁸ Also, local compactness or completeness is often assumed.)

One obtains singular spaces with bounded curvature, for example, by *cone* or *gluing constructions*. To describe the former, let X be a metric space with $\text{diam}(X) \leq \pi$. We can define a metric on $X \times (0, \infty)$ by the formula

$$d_C((x, s), (y, t)) := \sqrt{s^2 + t^2 - 2st \cos(d_X(x, y))}.$$

The *cone over X* , $\text{Cone}(X)$, is the completion of $X \times (0, \infty)$ in the metric d_C . The unique point that will be added to $X \times (0, \infty)$ in the completion is called the *cone point*. For many spaces X with curvature at least 1, or at most 1, $\text{Cone}(X)$ is a space of curvature at least 0, or at most 0, respectively [40, p. 91, p. 132]. Typically, $\text{Cone}(X)$ is not a manifold at the cone point.

Examples of larger singular sets can be constructed by using gluing methods. For example, *Reshetnyak's gluing theorem* asserts that if two complete and locally compact spaces with curvature at most k are glued together along isometric convex subsets, then the resulting space has curvature at most k [40, p. 316].

The concept of a metric space with curvature bounded from above or below generalizes the notion of a Riemannian manifold with similar bounds for the sectional

⁶Despite being singular in appearance, the spaces in this second class are in fact Lipschitz manifolds; that is, they are locally bi-Lipschitz equivalent to open regions in \mathbb{R}^2 . This follows from a recent result of Bonk and Lang [32].

⁷More generally, a *geodesic metric space* is one where every pair of points can be joined by a geodesic.

⁸A metric space is a *length space* if the distance between every pair of points in the space is the infimum of the lengths of paths joining the points.

curvature. In fact, the concept exactly specializes to that of sectional curvature bounds in case of Riemannian manifolds. Many classical theorems of Riemannian geometry, describing the effect of sectional curvature bounds on global geometry, have been generalized to spaces with bounded curvature. Such generalizations have applications also to the study of Riemannian manifolds via *compactness arguments*, as discussed later in subsection 8.8. Many spaces with bounded curvature are obtained as limits of smooth manifolds; it is an open problem to what extent the converse is true.⁹

The second class of Alexandrov spaces is defined by using triangles as in the preceding, but now computing their “excesses”; the outcome is turned into a global quantity in the form of a signed measure. We give an alternate but equivalent definition. Let X be a metric space that is also a topological 2-manifold. Then X is said to be a *surface of bounded curvature (in the sense of Alexandrov)* if it is locally isometric to a domain $G \subset \mathbb{R}^2$ that is conformally deformed by a density e^u , where u is a locally integrable function in G whose distributional Laplacian is a signed measure of finite total mass with every point mass less than 2π . Let us make this definition more precise. First, the assumption on u means that u satisfies

$$(8.2) \quad \int_G u \Delta \varphi \, dx = \int_G \varphi \, d\mu$$

for all functions $\varphi \in C_0^\infty(G)$, where μ is a signed measure on G with

$$|\mu|(G) < \infty \quad \text{and} \quad \mu(\{z\}) < 2\pi$$

for every $z \in G$. (Here $|\mu| = \mu^+ + \mu^-$ denotes the total variation of μ .) Second, given such a function u , by a *conformal deformation* of G by the function e^u we mean a metric d_u on G that is defined by

$$(8.3) \quad d_u(a, b) := \inf \int_\gamma e^u \, ds,$$

where the infimum is taken over all rectifiable curves γ in G joining a and b . Although u need not be smooth, one can show that the requirement (8.2) implies that u has a Borel representative for which the line integration in (8.3) is well defined and for which d_u defines a metric in G . Indeed, a function u as in (8.2) is a *δ -subharmonic function*, i.e., a difference of two subharmonic functions u^+ and u^- that satisfy $\Delta u^+ = \mu^+$ and $\Delta u^- = \mu^-$ in the sense of distributions for the positive measures μ^+ and μ^- . It follows from the classical potential theory that every δ -subharmonic function is well defined outside a set of 2-capacity zero, in particular outside a set of Hausdorff dimension zero; cf. (7.2).

If u is smooth, then a conformal deformation of the Euclidean metric in G by e^u as in (8.3) leads to a smooth metric whose Gaussian curvature K at a point $x \in G$ is given by $K(x) = -e^{-2u(x)} \Delta u(x)$. Because the volume element in this deformed space is $dV = e^{2u} \, dx$, we see that

$$(8.4) \quad \int K \, dV = - \int \Delta u \, dx.$$

Thus, surfaces of bounded curvature are metric surfaces where the classical Gaussian curvature makes sense not pointwise but as a measure, replacing the integral on the right in (8.4).

⁹See the discussion in [45, p. 469ff.] or in [40, p. 402].

Examples of surfaces of bounded curvature are polyhedral surfaces with uniform bounds on the vertex complexity. It is a difficult theorem that all surfaces of bounded curvature arise as limits of polyhedral surfaces with curvature concentrated as atoms in the vertices, with appropriately bounded total mass. Thus, surfaces of bounded curvature can be viewed as a closure of polyhedral spaces under an appropriate kind of convergence.

As mentioned in a footnote in 8.4, surfaces of bounded curvature in the sense of Alexandrov are not singular as defined in 8.3, but this is only a recently discovered fact.

8.5. Sub-Riemannian spaces. The starting point is a (connected) Riemannian manifold M and a proper subbundle H of the tangent bundle TM . We assume that H satisfies *Hörmander's condition*, meaning that sufficiently many brackets of vector fields with values in H generate the tangent space at each point. Under this assumption, a metric can be defined on M by the formula as in (8.1), but now the infimum is taken over a smaller class of curves, namely those whose tangent vectors lie in H . Such a metric is often called a *Carnot metric* on M , and the resulting metric space a *Carnot-Carathéodory space*.

The Hausdorff dimension of a Carnot-Carathéodory space is always larger than the dimension of M , so we have a singular space.¹⁰ The local geometry of a Carnot-Carathéodory space is rather complicated and currently the subject of extensive study. Carnot-Carathéodory spaces can be viewed as limits of Riemannian manifolds, where one lets the length of the vectors that are perpendicular to the subbundle H grow without bound.

Special cases of Carnot-Carathéodory spaces are *Carnot groups*, where the underlying manifold M is a nilpotent Lie group and the Carnot metric is invariant under left translations. In this case, there is also a natural one-parameter family of *dilations* on M that behave with respect to the Carnot metric much like the Euclidean dilations $x \mapsto \lambda x$ for $\lambda > 0$. These dilations provide a mild form of symmetry in otherwise very singular Carnot groups.

In proving his celebrated rigidity results, Mostow was forced to perform some delicate calculus on Carnot groups that naturally arise as boundaries of rank one symmetric spaces. Thus, after Alexandrov spaces, Carnot-Carathéodory spaces were perhaps the next singular spaces where analysis and geometry were considered.

Carnot-Carathéodory spaces also arise as natural model geometries for hypoelliptic partial differential equations; in the general case, they play the role played by Euclidean or Riemannian geometries that can be recovered from the associated Laplace equation.

8.6. Precompactness and convergence of metric spaces. It is a simple but extremely effective idea to consider convergence of metric spaces. Many singular spaces arise naturally this way. For subsets of a fixed space, we have the *Hausdorff convergence*, or *Hausdorff distance*. The general case of (separable metric) spaces can either be reduced to the preceding case or can be handled in a different manner applicable to all metric spaces.

Let us first consider compact spaces. The *Gromov-Hausdorff distance* between two compact metric spaces X and Y is defined as

$$(8.5) \quad d_{\text{GH}}(X, Y) := \inf \text{dist}_H^Z(X, Y),$$

¹⁰The Hausdorff dimension of a metric space is a bi-Lipschitz invariant.

where the infimum is taken over all compact metric spaces Z that contain (isometric) copies of X and Y and where $\text{dist}_H^Z(X, Y)$ is the Hausdorff distance between X and Y in Z . Recall that the latter is defined as the infimum of the positive numbers $\epsilon > 0$ such that X lies in the ϵ -neighborhood of Y in Z and Y lies in the ϵ -neighborhood of X in Z . (Note that we have suppressed the isometries from the notation in (8.5) and consider X and Y as subsets of Z .)

Unlike the classical Hausdorff distance, which uses a fixed underlying reference space, the Gromov-Hausdorff distance assigns value zero for each pair of mutually isometric spaces.

The family of isometry classes of compact metric spaces equipped with the Gromov-Hausdorff distance is a complete, separable, and contractible metric space $\mathcal{C} = (\mathcal{C}, d_{\text{GH}})$. The contractibility is obvious, for $X_t = (X, t d_X)$, $0 \leq t \leq 1$, provides a path from X to the one point space. The separability follows from the fact that the isometry classes of finite metric spaces are dense in \mathcal{C} .

The completeness of \mathcal{C} is a consequence of *Gromov's compactness theorem*, which asserts that a family $\mathcal{M}(D, N) \subset \mathcal{C}$ is precompact if two geometric quantities are uniformly bounded for all members in the family. These bounds are described by a positive number $D > 0$ and by a positive function $N : (0, \infty) \rightarrow (0, \infty)$. We require for each space X in the family $\mathcal{M}(D, N)$ that the diameter of X does not exceed D and that the cardinality of every maximal ϵ -net¹¹ in X does not exceed $N(\epsilon)$. Gromov's theorem states that every sequence from such a family $\mathcal{M}(D, N)$ has a convergent subsequence in \mathcal{C} .

It is more important to have a notion of *convergence of spaces* rather than an actual distance. In particular, one can define a suitable convergence notion in the case of noncompact spaces that restricts to the convergence in \mathcal{C} for compact spaces. To this end, we consider pointed proper¹² metric spaces $X = (X, x)$.

One would like to say that a sequence of pointed and proper spaces $X_n = (X_n, x_n)$ converges to a pointed proper space $X_\infty = (X_\infty, x_\infty)$ if the closed balls $\overline{B}_n(x_n, r) \subset X_n$ converge in the Gromov-Hausdorff distance to $\overline{B}(x_\infty, r) \subset X_\infty$ for each $r > 0$. This definition is indeed correct for a large class of spaces, for example for length spaces. Unfortunately, in general such a definition would not agree with the convergence in the Gromov-Hausdorff distance for compact spaces. (Consider the sequence $X_n = (\{-1/n, 1 + 1/n\}, -1/n)$ of pointed subsets of \mathbb{R} .)

The correct general definition stipulates the following. We say that a sequence of pointed proper metric spaces $X_n = (X_n, x_n)$ *Gromov-Hausdorff converges* to a pointed proper metric space $X_\infty = (X_\infty, x_\infty)$ if for every $r > 0$ and for every $\epsilon > 0$ there exists n_0 such that, whenever $n \geq n_0$, there exist maps

$$f_n : B(x_n, r) \rightarrow X_\infty, \quad f_n(x_n) = x_\infty,$$

such that the ball $B(x_\infty, r - \epsilon)$ belongs to the ϵ -neighborhood of the image $f_n(B(x_n, r))$ in X_∞ and such that

$$d_{X_n}(a, b) - \epsilon \leq d_{X_\infty}(f_n(a), f_n(b)) \leq d_{X_n}(a, b) + \epsilon$$

for every $a, b \in B(x_n, r)$. We also say that the pointed space X_∞ is a *Gromov-Hausdorff limit* of the pointed spaces X_n in this case.

¹¹A maximal ϵ -net, $\epsilon > 0$, in a metric space is a maximal subset in the space with the property that distinct points in the subset lie at least distance ϵ apart.

¹²A metric space is *proper* if every closed metric ball in the space is compact. The term *boundedly compact* is also used.

If (X_n, x_n) Gromov-Hausdorff converges to (X_∞, x_∞) , and if (y_n) is a sequence of points such that $y_n \in X_n$ for each n , we say that $y_n \rightarrow y \in X_\infty$ if for every $r > 0$ such that $y_n \in B(x_n, r)$ and for every $\epsilon > 0$ there exist n_0 and maps f_n as in the preceding such that $d_\infty(f_n(y_n), y) < \epsilon$.

In the next subsection, we discuss how the Gromov compactness theorem can be used to show the existence of (weak) tangent spaces to general metric spaces.

8.7. Tangent spaces. In general, there are many tangent spaces at a given point, depending on the small scales considered. A convenient class of metric spaces, where tangent spaces can be expected, is the class of doubling spaces.

A metric space is said to be *doubling* if there exists a constant $C \geq 1$ such that every ball in the space can be covered by at most C balls of half the radius. The doubling condition promotes itself to the following ostensibly stronger requirement: there exist constants $C \geq 1$ and $\alpha > 0$ such that the cardinality of every maximal ϵR -net, $0 < \epsilon \leq \frac{1}{2}$, in a ball of radius $R > 0$ does not exceed $C \epsilon^{-\alpha}$. The infimal α such that the preceding condition holds in a doubling metric space for some C is called the *Assouad dimension* of the metric space.¹³ It is easy to see that the Assouad dimension is always at least the Hausdorff dimension and can be larger.

Suppose now that $X = (X, d)$ is a complete doubling space. Then X is proper, and by employing Gromov’s compactness theorem, it is not hard to see that the sequence

$$(8.6) \quad X_\epsilon := (X, \epsilon^{-1}d, x), \quad \epsilon > 0, \quad x \in X,$$

of pointed proper spaces has a subsequence (X_{ϵ_n}) which Gromov-Hausdorff converges to a pointed proper space (X_∞, x_∞) as $\epsilon_n \rightarrow 0$. Moreover, X_∞ is doubling as well. Every such space X_∞ is called a *tangent space*¹⁴ of X at x . Roughly speaking, such tangent spaces exist because doubling spaces can be approximated (locally) by finite spaces with cardinality bounds independent of the given scale.

Although in the definition of Gromov-Hausdorff convergence we assume the spaces to be proper, it is clear that in the present context the definition can be altered so as to apply to locally compact doubling spaces.¹⁵

It follows that tangent spaces always exist for X locally compact and doubling. If X is a Riemannian manifold, then these spaces are always isometric to some Euclidean space. For a Carnot-Carathéodory space (at least under some additional assumptions [135]) each tangent space is a Carnot group, and for a Carnot group each tangent space is the same Carnot group.

More generally, a *weak tangent space* (or *cone*) of a proper metric space (X, d) is a Gromov-Hausdorff limit (X_∞, x_∞) of pointed proper metric spaces $(X, \lambda_n d, x_n)$, where $\lambda_n > 0$ and $x_n \in X$. Note that here we do not impose any additional requirement for the numbers λ_n ; in particular, X (with a base point) is its own weak tangent space always. Weak tangent spaces play an important role in the study of the local structure of Riemannian manifolds with curvature bounds and in geometric rigidity questions, for example. However, we will not need weak tangent spaces in this article.

¹³Other names are also used, e.g., a *uniform covering dimension*.

¹⁴The term *tangent cone* is also used in this context, although this term has a different meaning in geometric measure theory [64, 3.1.21].

¹⁵Without this remark, noncomplete Riemannian manifolds, for example, would be left out of the discussion.

Along with tangent spaces, we need to discuss tangent functions. Given any metric space $X = (X, d)$ and an L -Lipschitz function $f : X \rightarrow \mathbb{R}$, we form the *rescaled functions*

$$(8.7) \quad f_{x,r}(y) := \frac{f(x) - f(y)}{r}, \quad x \in X, \quad r > 0.$$

Every such rescaled function $f_{x,r}$ is an L -Lipschitz function on the rescaled (pointed) metric space

$$X_{x,r} := (X, r^{-1}d, x)$$

and satisfies $f_{x,r}(x) = 0$.

Now, let X be a locally compact doubling space, let $x \in X$, and let $X_\infty = (X_\infty, x_\infty)$ be a tangent space to X at x ; i.e., X_∞ is a Gromov-Hausdorff limit of a sequence $X_{x,r_j} = (X, r_j^{-1}d, x)$ as $r_j \rightarrow 0$. Standard Arzelà-Ascoli type arguments imply that there is a subsequence (r_{j_i}) of (r_j) such that the rescaled functions $f_{x,r_{j_i}}$ converge (in a suitable sense) to a Lipschitz function

$$(8.8) \quad f_\infty : X_\infty \rightarrow \mathbb{R}, \quad f_\infty(x_\infty) = 0.$$

Such a Lipschitz function f_∞ is called a *tangent function* to f at x .

Differential analysis on metric spaces deals with the structure of tangent spaces and tangent functions on them.

For example, smooth functions on smooth manifolds have unique tangent functions that are linear functions from the (unique, linear) tangent spaces to \mathbb{R} . This fact is a direct consequence of the definitions. There is a deeper fact, proved by Rademacher in 1919: for every Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the *rescaled functions* $f_{x,r}$ as in (8.7) converge, as $r \rightarrow 0$, to a linear function at almost every $x \in \mathbb{R}^n$. In its classical setting, one easily forgets that there are two separate statements to Rademacher's theorem: the almost everywhere existence of a limit and the special structure of the limit function. We will discuss some far-reaching extensions of Rademacher's theorem in Section 12.

8.8. Singular spaces as limit spaces. Uniform bounds on curvature and diameter for a collection of compact Riemannian manifolds ensure that the collection is precompact in the Gromov-Hausdorff metric. For example, given $n \geq 2$, $D > 0$, and $k \in \mathbb{R}$, we have that every collection $\mathcal{R}(n, D, k)$ of closed Riemannian n -manifolds with diameter not exceeding D and with sectional curvature at least k at every point is a precompact family. In fact, one can show that the spaces in each such collection are uniformly doubling.

Gromov-Hausdorff limits of manifolds in $\mathcal{R}(n, D, k)$ need not be smooth Riemannian, but they are metric spaces with curvature bounded from below in the sense of Alexandrov, as defined in 8.4. One can draw a parallel to the Sobolev space theory. A collection of smooth functions in \mathbb{R}^n with uniform bound in the Sobolev norm $\|\varphi\|_{1,p}$, $1 < p < \infty$, as given in (5.3) is precompact in the Sobolev space in the weak topology.¹⁶ The limit functions need not be smooth, but as members of $W^{1,p}(\mathbb{R}^n)$ they possess a certain degree of regularity. On the other hand, every Sobolev function is a (strong) limit of smooth functions. As mentioned earlier in subsection 8.4 it is not known whether every space with curvature bounded from below in the sense of Alexandrov is a limit of smooth spaces.

¹⁶We need $1 < p < \infty$ here for the reflexivity of the pertinent Sobolev spaces. A variant of this result for $p = 1$ would involve functions whose distributional gradients are measures.

The precompactness for a collection $\mathcal{R}(n, D, k)$ as in the preceding is true under weaker bounds on the curvature. The main point is to obtain a uniform bound for the doubling constant for each member in the family, and for this it suffices to assume a uniform lower bound for the Ricci curvature by classical volume comparison theorems. In other words, every collection $\mathcal{R}(n, D, k)$ of closed Riemannian n -manifolds with diameter not exceeding D and with Ricci curvature at least k at every point is a precompact family.¹⁷

Bounds on Ricci curvature are weaker than bounds on sectional curvature, and more complicated limit spaces occur than those that have curvature bounded in the sense of Alexandrov. However, limits of Riemannian manifolds with uniform lower bound on Ricci curvature are known to possess enough geometric structure so as to allow for calculus. This is discussed more in subsection 16.3. In general, the structure of these spaces is not as well understood as the structure of spaces with bounded curvature in the sense of Alexandrov.

8.9. Geometrized decomposition spaces. Geometric topology has produced many fascinating examples of spaces that are “nearly manifolds”. We now discuss some of them as they pertain to calculus.

Given a collection $G = \{F\}$ of closed sets in a topological space X , one can form a quotient space X/G , where the members of G are crushed to points. Under mild conditions on the collection G , the resulting quotient space equipped with the quotient topology is metrizable if X is metrizable. Such quotient spaces are called *decomposition spaces*.

The Bing school in topology created extraordinarily sophisticated techniques to study decomposition spaces. One of the primary questions was to understand the conditions under which X/G is an n -manifold if X is an n -manifold. Even if X/G is not a manifold, it is in many cases a *homology manifold*, which means that it cannot be distinguished from a manifold by means of local algebraic topology. More precisely, if X is a topological n -manifold and G is an *upper semicontinuous decomposition* of X into *cell-like* compact sets, then $Y := X/G$ satisfies

$$(8.9) \quad H_*(Y, Y \setminus \{y\} : \mathbb{Z}) = H_*(\mathbb{R}^n, \mathbb{R}^n \setminus \{0\} : \mathbb{Z}),$$

where H_* stands for singular homology.¹⁸ Condition (8.9) is sufficient for many geometric and analytic arguments, for the degree theory of mappings, for example.¹⁹

Traditionally, in the theory of decomposition spaces, little attention has been paid to metric questions. Recently, Semmes used some of the classical decomposition spaces to construct interesting examples of singular metric spaces where calculus is possible. In these examples, the (metric or topological) singularities that are formed by crushing the members of G to a point turn out to be inconsequential from the point of view of differential analysis, provided the construction is done with care. Semmes’s spaces are limits of smooth manifolds, where certain metric constraints are kept within bounds. At the end, calculus, like algebraic topology, cannot tell whether X/G is smooth or even a manifold.

¹⁷By the lower bound for the Ricci curvature we mean that the Ricci curvature tensor Ric satisfies $\text{Ric} \geq kg$, where g is the Riemannian metric.

¹⁸See [54, p. 191] for this result as well as for the terminology.

¹⁹In general, a locally compact space that satisfies the local homological condition (8.9) for some positive integer n is called a *homology n -manifold*.

Classical geometric topology provides many challenging problems of a metric nature where analysis may play a role.

8.10. Boundaries at infinity. A source of examples of singular spaces, where it is desirable to develop calculus for external reasons, is provided by the theory of *negatively curved*, or *Gromov hyperbolic*, groups. With every such group one can associate a *boundary at infinity* whose geometry reflects the asymptotic geometry of the group. More generally, one can consider negatively curved Gromov hyperbolic metric spaces and obtain a similar asymptotic theory. Examples of such spaces are simply connected Riemannian manifolds with negative curvature bounded away from zero and fundamental groups of negatively curved closed Riemannian manifolds. Some Carnot groups appear in this role as boundaries of negatively curved *homogeneous spaces* (in the sense of Riemannian geometry). Exotic geometries appear also on the boundaries of certain hyperbolic *buildings*.

In many interesting cases, it is unknown whether the boundary carries a nice geometric structure or not. For example, it seems to be unknown what kind of geometries can be carried on the boundary of the fundamental group of the negatively curved 4-manifolds constructed in [75]. Spaces with interesting features arise also in the case of compact hyperbolic manifolds with totally geodesic boundaries or when closed hyperbolic manifolds are glued along totally geodesic submanifolds [98], [122]. The structure of the boundary of their universal covers is unknown.

8.11. Other examples. Examples of exotic singular metric spaces where calculus is possible can be constructed by hand. Most notable in this respect are the spaces constructed by Laakso in [121]. He showed that for every given Hausdorff dimension at least one, there are topologically one-dimensional spaces with this given Hausdorff dimension that are suitable for calculus. These spaces cannot resemble the classical fractals; they must contain lots of rectifiable curves, for example.

8.12. Notes. Spivak's book [170] contains a nice analysis of Riemann's lecture. Of course, there is a huge literature on this topic. The most immediate generalization of a Riemannian metric leads to *Finsler geometry*, where one equips each tangent space of a smooth manifold with an arbitrary (smoothly varying) Banach norm; see, for example, [21]. Basic references to Lipschitz manifolds are [131], [175], [181].

The monographs [37], [40] contain much information about spaces with bounded curvature in the sense of Alexandrov. See also the survey [144]. For the theory of surfaces of bounded curvature in the sense of Alexandrov, see [6], [1], [148]. The work [32] by Bonk and Lang on bi-Lipschitz parametrizations of these surfaces was cited in subsection 8.4. A weaker result along these lines (under a small mass assumption) was proved earlier by Fu [68]. We return to the general question of bi-Lipschitz parametrizations later in subsection 16.5.

Carnot-Carathéodory spaces are currently under extensive study. For the origins of the analytic aspects of the theory, related to hypoelliptic partial differential equations and complex analysis, see [172] and the references there. Excellent sources for more geometric aspects of the theory are the articles [22], [72]. For Mostow rigidity, see [137], [138]. The contemporary literature on harmonic analysis, partial differential equations, and geometric measure theory in sub-Riemannian contexts is large and growing.

For a thorough discussion of the Gromov compactness theorem and convergence of metric spaces, see [73, Chapter 5], [40, Chapters 7 and 8]. Cheeger's lectures [43]

contain a survey of some of the applications of these ideas to Riemannian geometry, together with further references. Weak tangent spaces and tangent functions have been extensively used by Cheeger and Colding [45], [46], [47], Cheeger [42], David and Semmes [56], Keith [100], Bonk and Kleiner [29], [30], and others. The concepts of a doubling metric space and Assouad dimension (under a different name) can be found in Larman's paper [124] from 1967, but they have received more attention after *Assouad's embedding theorem* [14] became a well known and effective tool in metric geometry. See [56], [159], [81], [160] and the references there.

Rademacher's theorem was proved in 1919 [145]. A proof can be found, e.g., in [62], [64].

An accessible account of the decomposition space theory is Daverman's book [54], where references to the original works of Bing and others can be found. The constructions of Semmes, alluded to in 8.9, can be found in [156], [157]. Although the spaces by Semmes are singular as defined in this article, they still admit "branched Lipschitz parametrizations" by Euclidean space [89].

The fundamental source for Gromov hyperbolic spaces is [71]. See also [40], [37], and references there. For the boundaries of homogeneous spaces and buildings, see [137], [91], [143], [35], [36].

Further examples or references related to singular spaces can be found in [160], [81].

9. SPACES OF HOMOGENEOUS TYPE

At this juncture, it is instructive to briefly discuss spaces of homogeneous type. These are spaces where much of the classical real analysis (without derivatives) is possible.

Let $X = (X, d, \mu)$ be a separable metric space equipped with a Borel regular measure μ such that the measure of every metric ball in X is finite and positive. Henceforth, such a triple is termed a *metric measure space*.

A metric measure space $X = (X, d, \mu)$ is said to be a *space of homogeneous type* if the measure satisfies a *doubling condition*: there exists a constant $C \geq 1$ such that

$$(9.1) \quad \mu(B(x, 2r)) \leq C \mu(B(x, r))$$

for every ball $B(x, r)$ in X .

In fact, the term "homogeneous space" usually stands for something more general; the distance function $d : X \times X \rightarrow [0, \infty]$ is only assumed to be a *quasimetric*, where both the triangle inequality and the symmetry condition are required to hold up to a multiplicative constant. We forgo these more general conditions here. Moreover, we choose to use the term *doubling metric measure space* for a metric measure space X as above if μ satisfies (9.1). Finally, in this case, we call μ a *doubling measure*.

By the 1970s it had become clear that much of the basic zeroth order analysis can be done in spaces of homogeneous type, in particular in doubling metric measure spaces. By "zeroth order" analysis we mean analysis involving functions only, and not their derivatives in any disguise. As a prime example, we have that the Hardy-Littlewood maximal operator,

$$Mf(x) := \sup_{r>0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} f \, d\mu,$$

maps $L^1(X)$ to weak- $L^1(X)$, and by interpolation $L^p(X)$ to $L^p(X)$ for $1 < p < \infty$. The maximal operator and its mapping properties provide one of the most important technical tools in classical harmonic analysis in Euclidean spaces. Behind the basic properties of the maximal operator are fundamental covering theorems of *Vitali type*, valid in spaces of homogeneous type.

From the preceding maximal function theorem one derives some standard consequences, such as the Lebesgue differentiation theorem: if u is a locally integrable function on a doubling metric measure space X , then

$$\lim_{r \rightarrow 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |u(y) - u(x)| d\mu(y) = 0$$

for almost every $x \in X$.

These two examples give a flavor of what can be done in metric measure spaces satisfying the doubling condition (9.1). It is not the purpose of this article to go further in this direction, but it should be emphasized that much of the zeroth order analysis in homogeneous spaces, such as the maximal theorem, is crucial in the development of the (first order) calculus in metric measure spaces to be discussed later.

The structure of a doubling metric measure space is not strong enough to allow for calculus as defined in the beginning of this article. For example, consider the “snowflake” doubling metric measure space

$$(9.2) \quad X = ([0, 1], |x - y|^{1/2}, \mathcal{H}_2),$$

where \mathcal{H}_2 is the 2-dimensional Hausdorff measure. The function $f(x) = x$ is Lipschitz on X , with

$$\frac{|f(x) - f(y)|}{|x - y|^{1/2}} = |x - y|^{1/2} \rightarrow 0$$

as $y \rightarrow x$ for every $x \in [0, 1]$. Thus, if a reasonable notion of a derivative existed for Lipschitz maps on X , it would vanish for this f . Such infinitesimal information would then imply that f is constant, which is not the case.

The preceding example presupposes a requirement that should be consequential to any reasonable calculus. From the point of view of this paper, the space X in (9.2) fails this requirement in a fundamental way because it has no nonconstant rectifiable curves.

A similar phenomenon persists on many classical fractals, which have some but not enough rectifiable curves. In such cases, the analysis is more subtle.

Finally, we point out that there are essentially no geometric obstructions for a metric space to carry a nontrivial doubling measure. If such a measure exists, then it is easy to see that X as a metric space is doubling as defined in subsection 8.7. It turns out that every complete doubling metric space carries a nontrivial doubling measure.

Later we will see what extra conditions are needed so that a metric measure space would allow for calculus. These conditions impose geometric restrictions on the space.

9.1. Notes. A fundamental work on spaces of homogeneous type is the book [50] by Coifman and Weiss; see also the survey [51]. For an excellent account on more recent developments as well as connections to geometric problems, see Semmes’s article [159]. For constructions of doubling measures on doubling spaces, see [185],

[130], [190], [81]. For the use of the Vitali covering theorem in this connection, see e.g. [159], [81].

10. SOBOLEV SPACES IN METRIC MEASURE SPACES

We finally begin to explain more precisely what is meant by calculus on metric measure spaces. We will see in this section that with each metric measure space $X = (X, d, \mu)$ there is canonically associated a first order Sobolev space $N^{1,p}(X)$ for every $1 \leq p < \infty$. This Sobolev space is a Banach space and agrees with the classical Sobolev space $W^{1,p}(\mathbb{R}^n)$ if $X = \mathbb{R}^n$.

The letter N in the terminology is used to separate $N^{1,p}(X)$ from other Sobolev spaces that can abstractly be defined, even though some of these may actually coincide. This is in keeping with the tradition of using the notations $H^{1,p}(\mathbb{R}^n)$ and $W^{1,p}(\mathbb{R}^n)$, as discussed earlier.

First we make some remarks about rudimentary calculus on metric measure spaces based on rectifiable curves.

10.1. Modulus in metric measure spaces. Curves and their rectifiability, and consequently line integration of nonnegative Borel functions, can be considered in every metric space. Every rectifiable curve $\gamma : [a, b] \rightarrow X$ has an arc length parametrization

$$\gamma_0 : [0, \text{length}(\gamma)] \rightarrow X,$$

and we define

$$\int_{\gamma} \rho \, ds := \int_0^{\text{length}(\gamma)} \rho(\gamma_0(t)) \, dt$$

whenever $\rho : X \rightarrow [0, \infty]$ is a Borel function.

In particular, the definition and basic properties of modulus as described in Section 7.2 transfer over to general metric measure spaces. We can talk about a property holding on p -almost every curve in a metric measure space X . The basic lemmas 7.2, 7.3, and 7.5 remain valid with the same proofs.

Much of the ensuing discussion can be seen as a reduction to one dimensional calculus, with modulus as an outer measure exerting some global control.

10.2. Upper gradients. Let $X = (X, d)$ be a metric space. A Borel function $\rho : X \rightarrow [0, \infty]$ is said to be an *upper gradient* of a function $u : X \rightarrow \mathbb{R}$ if

$$(10.1) \quad |u(a) - u(b)| \leq \int_{\gamma} \rho \, ds$$

whenever $a, b \in X$ and γ is a rectifiable curve in X with end points a and b .

As trivial examples, we note that $\rho \equiv \infty$ is an upper gradient of every function, and if X contains no rectifiable curves, then $\rho \equiv 0$ is an upper gradient of every function on X . If $f : X \rightarrow \mathbb{R}$ is Lipschitz, then $\rho \equiv \text{Lip}(f)$ is an upper gradient, where

$$(10.2) \quad \text{Lip}(f) := \sup_{a \neq b} \frac{|f(a) - f(b)|}{d(a, b)}$$

is the *Lipschitz number* of f .

More interesting examples of upper gradients are the pointwise infinitesimal Lipschitz numbers

$$(10.3) \quad \text{Lip}f(x) := \limsup_{r \rightarrow 0} \sup_{d(x,y) < r} \frac{|f(x) - f(y)|}{r}$$

and

$$(10.4) \quad \text{lip}f(x) := \liminf_{r \rightarrow 0} \sup_{d(x,y) < r} \frac{|f(x) - f(y)|}{r},$$

for a Lipschitz function $f : X \rightarrow \mathbb{R}$. It is not hard to see that indeed the functions $\text{Lip}f$ and $\text{lip}f$ are upper gradients of f .

By standard measure theoretic arguments, and by using the fundamental theorem of calculus, one verifies that $\rho = |\nabla u|$ is an upper gradient of every smooth function on \mathbb{R}^n (or more generally on a Riemannian manifold), where $|\nabla u|$ denotes the length of the (Riemannian) gradient of u .

The definition for upper gradients is purely metric, but the concept is most useful in the context of metric measure spaces. In fact, even in \mathbb{R}^n we cannot expect $\rho = |\nabla u|$ to be an upper gradient of an arbitrary Sobolev function u . To counter this problem, we next introduce a weaker notion of an upper gradient. The modulus of a curve family is an essential notion here.

10.3. p -weak upper gradients. Let $X = (X, d, \mu)$ be a metric measure space and let $1 \leq p < \infty$. A Borel function $\rho : X \rightarrow [0, \infty]$ is said to be a p -weak upper gradient of a function $u : X \rightarrow \mathbb{R}$ if the inequality in (10.1) holds for p -almost every curve γ in X .

If u is a p -quasicontinuous function in $W^{1,p}(\mathbb{R}^n)$, then it follows from the proofs for Theorems 7.4 and 7.6 that every Borel representative ρ of $|\nabla u|$ is a p -weak upper gradient of u . In fact, it is necessary to take the quasicontinuous representative here, for every locally integrable function in \mathbb{R}^n that possesses a p -integrable p -weak upper gradient is p -quasicontinuous. We will see this momentarily.

Functions that have p -integrable p -weak upper gradients are the analogs of Sobolev functions on metric measure spaces. The precise definition, presented momentarily, requires a little care because of issues related to the pointwise definition. Let us first study functions with p -integrable upper gradients without further formalism. Thus, all functions are assumed to be defined pointwise everywhere, unless otherwise stipulated.

First we observe, by Lemma 7.3, that every p -integrable p -weak upper gradient of a function can be approximated in $L^p(X)$ by upper gradients. In fact, if ρ is p -integrable and satisfies (10.1) outside a p -exceptional family Γ of curves, there exists a p -integrable Borel function $\sigma : X \rightarrow [0, \infty]$ such that

$$\int_{\gamma} \sigma \, ds = \infty$$

for every $\gamma \in \Gamma$. Consequently,

$$(10.5) \quad \rho_{\epsilon} := \rho + \epsilon \sigma, \quad \epsilon > 0,$$

is an upper gradient of u and $\rho_{\epsilon} \rightarrow \rho$ in $L^p(X)$ as $\epsilon \rightarrow 0$.

In view of the preceding, the difference between upper gradients and weak upper gradients seems small, at least if the latter exist and are appropriately integrable. However, there is no viable theory based on upper gradients alone, even for functions

in $W^{1,p}(\mathbb{R}^n)$, as should be clear to the reader by now. Modulus provides the precise conceptual framework in dealing with the failure of the fundamental theorem of calculus on all curves and is especially indispensable in spaces with no rectilinear coordinate system, or system of “preferred curve families”.

A function $u : X \rightarrow [0, \infty]$ is said to be *absolutely continuous on a rectifiable curve* γ if $u \circ \gamma_0 : [0, \text{length}(\gamma)] \rightarrow \mathbb{R}$ is absolutely continuous, where γ_0 is the arc length parametrization of γ .

The next proposition is essentially built into the definitions; cf. Theorem 7.4.

Proposition 10.1. *A real-valued function that possesses a p -integrable p -weak upper gradient on a metric measure space is absolutely continuous on p -almost every curve.*

10.4. Minimal p -weak upper gradients. Generally a given function has infinitely many upper gradients; but if there exists one that is p -integrable, then there exists one that is minimal.

Proposition 10.2. *Suppose that a function $u : X \rightarrow \mathbb{R}$ has a p -integrable p -weak upper gradient. Then there exists a minimal p -weak upper gradient ρ_u characterized by the following two properties: ρ_u is a p -integrable p -weak upper gradient of u ; and if ρ is another p -integrable p -weak upper gradient of u , then $\rho_u \leq \rho$ almost everywhere.*

Proof. First one proves (we omit the details) that p -integrable p -weak upper gradients form a lattice in the sense that if τ and σ are two p -integrable p -weak upper gradients of a given function, then so is $\min\{\tau, \sigma\}$. Consequently, any sequence (ρ_i) of p -weak upper gradients of a given function u satisfying

$$\lim_{i \rightarrow \infty} \|\rho_i\|_{L^p(X)} = \inf_{\rho} \|\rho\|_{L^p(X)},$$

where the infimum is taken over all p -weak upper gradients ρ of u , can be chosen to be pointwise decreasing:

$$\rho_1 \geq \rho_2 \geq \dots$$

Clearly, then, (ρ_i) converges in $L^p(X)$ to a Borel function ρ_u whose L^p -norm assumes the above infimum. By Fuglede’s lemma 7.5, ρ_u is a p -weak upper gradient of u . It is also clear from the lattice property of upper gradients that ρ_u is minimal as asserted. \square

It is not difficult to verify, by using Fubini’s theorem and the fundamental theorem of calculus, that if u is a function on \mathbb{R}^n that possesses a p -integrable p -weak upper gradient, then the partial derivatives of u exist and $\rho_u = |\nabla u|$ almost everywhere.

A similar remark holds in other situations, where $|\nabla u|$ makes sense more classically, e.g. in Riemannian or sub-Riemannian contexts.

In conclusion, we can canonically attach a p -integrable function ρ_u with a given function u in an arbitrary metric measure space, provided that u has one p -integrable upper gradient. The function ρ_u is unique up to a set of measure zero, and it measures the size of the rate of change of u along most curves in the space. Moreover, ρ_u agrees with the more classical notion of the length of the gradient whenever the latter is available.

The minimal upper gradient satisfies the following *locality* property satisfied by the usual gradients.

Lemma 10.3. *Suppose that a function $u : X \rightarrow \mathbb{R}$ has a p -integrable p -weak upper gradient. If $A \subset X$ is a Borel set and $u|_A = c$ for some constant $c \in \mathbb{R}$, then $\rho_u|_A = 0$ almost everywhere.*

The lemma is not hard to derive from the definitions.

Remark 10.4. As defined, ρ_u depends *a priori* on the fixed parameter $p \in [1, \infty)$, although this dependence is suppressed in the notation. In many situations, although not always, ρ_u is known to be independent of p (see Section 12).

10.5. Sobolev spaces $N^{1,p}(X)$. Consider the vector space $\tilde{N}^{1,p}(X)$ consisting of all functions $u : X \rightarrow \mathbb{R}$ such that u is in $L^p(X)$ and there exists an upper gradient ρ of u in $L^p(X)$. Then we can define a seminorm in $\tilde{N}^{1,p}(X)$ by

$$(10.6) \quad \|u\|_{1,p} := \|u\|_{L^p(X)} + \|\rho_u\|_{L^p(X)},$$

where ρ_u is the minimal p -weak upper gradient of u defined in the previous subsection. The Sobolev space $N^{1,p}(X)$ is obtained by dividing out the elements of $\tilde{N}^{1,p}(X)$ of zero norm. That is,

$$N^{1,p}(X) := \tilde{N}^{1,p}(X) / \sim,$$

where

$$u \sim v \text{ if and only if } \|u - v\|_{1,p} = 0.$$

As is customary, we will talk about *functions* in $N^{1,p}(X)$ rather than equivalence classes.

The Sobolev space $N^{1,p}(X)$ normed by (10.6) is a Banach space, although this fact is not obvious from the definitions.

To show the power of the modulus method in connection with Sobolev spaces $N^{1,p}(X)$, we formulate and sketch a proof for the following property of weak compactness type.

Theorem 10.5. *Let $X = (X, d, \mu)$ be a metric measure space and let $1 < p < \infty$. Then every bounded sequence (u_i) of functions in the Banach space $N^{1,p}(X)$ contains a subsequence (u_{i_k}) that converges weakly in $L^p(X)$ to a function $u \in N^{1,p}(X)$. Moreover, we have that*

$$(10.7) \quad \|u\|_{1,p} \leq \liminf_{k \rightarrow \infty} \|u_{i_k}\|_{1,p}.$$

Proof. (For the standard functional analytic facts in the ensuing argument, see [191, pp. 120 and 126] for example.) Because $L^p(X)$ is reflexive for $1 < p < \infty$, and because the $L^p(X)$ -norms of the functions u_i are uniformly bounded, we can find a weakly convergent subsequence (u_{i_k}) . By passing to a further subsequence if necessary, we may assume that the corresponding sequence of minimal p -weak upper gradients (ρ_{i_k}) also converges weakly in $L^p(X)$. Let u denote the weak limit of the sequence (u_{i_k}) in $L^p(X)$, and let ρ be the weak limit of the sequence (ρ_{i_k}) in $L^p(X)$. We claim that an appropriate representative of u belongs to $N^{1,p}(X)$ and that ρ is a p -weak upper gradient of u . Then (10.7) follows from the lower semicontinuity of norms under weak convergence in Banach spaces.

To prove the claim, we first invoke *Mazur's lemma* asserting that from every weakly convergent sequence of elements in a Banach space one can form a sequence of convex combinations that converges strongly or in the norm. Let (v_j) be such a sequence of convex combinations of functions u_{i_k} . By passing to a further subsequence if necessary we may also assume that the correspondingly indexed sequence

σ_j of convex combinations of the functions ρ_{i_k} converges in $L^p(X)$ to ρ . The upper gradient inequality (10.1) is preserved under convex combinations in the sense that if v and w are functions with upper gradients ρ and σ , and if $0 < \lambda < 1$, then $\lambda\rho + (1 - \lambda)\sigma$ is an upper gradient of $\lambda u + (1 - \lambda)w$. It follows therefore that for each j , we have that σ_j is a p -weak upper gradient of v_j . Moreover, by (10.5), we can assume that σ_j is an actual upper gradient of v_j . By Fuglede’s lemma 7.5,

$$(10.8) \quad \lim_{j \rightarrow \infty} \int_{\gamma} \sigma_j \, ds = \int_{\gamma} \rho \, ds$$

for p -almost every curve γ in X . Next, we will show that for p -almost every curve γ we have both that (10.8) holds and that

$$(10.9) \quad \lim_{j \rightarrow \infty} v_j(x) = u(x) \in \mathbb{R}$$

for every x that is an end point of γ , provided that we properly choose a representative of u . In fact, we can pass to another subsequence and assume that (10.9) holds outside a set E of measure zero. Then set

$$(10.10) \quad u(x) := \frac{1}{2}(\limsup_{j \rightarrow \infty} v_j(x) + \liminf_{j \rightarrow \infty} v_j(x))$$

whenever this expression makes sense; for other points x we can set $u(x) = 0$. Because p -almost every curve γ meets E in a set of linear (Hausdorff 1-measure) zero (use as a test function $\infty \cdot \chi_E$ for a measure zero Borel superset of E and Lemma 7.3), we have for some $y \in \gamma$ that $y \notin E$. In particular, if also γ satisfies (10.8) and $x \in \gamma$ is arbitrary, then

$$|v_j(x)| - |v_j(y)| \leq |v_j(x) - v_j(y)| \leq \int_{\gamma} \sigma_j \, ds,$$

which gives that $\limsup_{j \rightarrow \infty} |v_j(x)| < \infty$. In particular, with the definition (10.10), we have that

$$|u(a) - u(b)| \leq \limsup_{j \rightarrow \infty} |v_j(a) - v_j(b)| \leq \limsup_{j \rightarrow \infty} \int_{\gamma} \sigma_j \, ds = \int_{\gamma} \rho \, ds$$

for p -almost every curve γ , where a and b are the end points of γ . It follows that ρ is a p -weak upper gradient of the representative of u given in (10.10) as required, and the theorem follows. \square

Remark 10.6. Theorem 10.5 does not assert that every bounded sequence in $N^{1,p}(X)$ contains a weakly convergent subsequence. In fact, this would imply that $N^{1,p}(X)$ is reflexive by the Eberlein-Shmulian theorem [59, Theorem 1, p. 430, and Theorem 7, p. 425]. However, $N^{1,p}(X)$ is not always reflexive; an example will be given in subsection 12.5. The general lack of reflexivity can be compensated by the result in Theorem 10.5.

The combination of the lemmas of Fuglede and Mazur as in the preceding proof has turned out to be a powerful tool in the analysis on metric measure spaces based on modulus.

10.6. Cheeger’s definition. The fact that inequality (6.2) need not hold for every rectifiable curve, for a given Sobolev function u in \mathbb{R}^n , is a source of technical problems if one wants to found a Sobolev space theory on upper gradients. We preferred to deal with this problem by using the p -modulus (cf. Theorem 7.6). Cheeger proposed a different approach which we next describe.

Let $X = (X, d, \mu)$ be a metric measure space, and let $u : X \rightarrow \mathbb{R}$ be a function in $L^p(X)$, $1 \leq p < \infty$. We say that a p -integrable Borel function $\rho : X \rightarrow [0, \infty]$ is a *generalized p -upper gradient*²⁰ of u if there exists a sequence (u_i) of functions in $L^p(X)$ and a corresponding sequence of upper gradients (ρ_i) such that $u_i \rightarrow u$ in $L^p(X)$ and that $\rho_i \rightarrow \rho$ in $L^p(X)$. The *Cheeger-Sobolev space*²¹ $C^{1,p}(X)$ consists of all L^p -functions u for which there exists a generalized p -upper gradient. Then $C^{1,p}(X)$ is a Banach space with the norm

$$(10.11) \quad \|u\|_{1,p} := \|u\|_p + \inf \|\rho\|_p,$$

where the infimum is taken over all generalized p -upper gradients of u .

In $N^{1,p}(X)$, for $p > 1$, there is a kind of weak compactness stemming from Fuglede's lemma 7.5 and Mazur's lemma as explained in Theorem 10.5. In $C^{1,p}(X)$, for all $p \geq 1$, this property and the lower semicontinuity of the $C^{1,p}$ -norm are built into the definition. The equality between $N^{1,p}(X)$ and $C^{1,p}(X)$ for $p > 1$ essentially follows from Theorem 10.5.

Theorem 10.7. *We have that*

$$C^{1,p}(X) = N^{1,p}(X)$$

isometrically when $1 < p < \infty$.

Proof. If $u \in N^{1,p}(X)$, then obviously $u \in C^{1,p}(X)$. Moreover, the embedding is norm nonincreasing because for every p -weak upper gradient ρ of u we can find a sequence of upper gradients converging to ρ in $L^p(X)$ as explained around (10.5). Conversely, it follows from Theorem 10.5 that every function $u \in C^{1,p}(X)$ has a representative that belongs to $N^{1,p}(X)$ and that this embedding also is norm nonincreasing. \square

When $p = 1$, the two spaces are generally different. Whereas $N^{1,1}(X)$ typically is a substitute for the classical Sobolev space, $C^{1,1}(X)$ is a substitute for the space BV of functions whose distributional derivatives are measures.

Cheeger also defined a *minimal generalized p -upper gradient*²² of a function $u \in L^p(X)$ to be a generalized p -upper gradient ρ_u such that

$$\|u\|_{1,p} = \|u\|_p + \|\rho_u\|_p.$$

Minimal generalized upper gradients can directly be shown to exist for every function $u \in C^{1,p}(X)$, for $1 < p < \infty$, by using the uniform convexity of $L^p(X)$. On the other hand, it follows from the preceding discussion that minimal generalized p -upper gradients are precisely the minimal p -weak upper gradients for $p > 1$, so the existence also follows from Proposition 10.2.

It should be emphasized that all the results about the space $N^{1,p}(X)$, and about minimal p -weak upper gradients, that are attributed to Cheeger later in this paper were actually proved in the context of the space $C^{1,p}(X)$ and in the context of minimal generalized upper gradients, that is, by using the definitions given in this subsection.

Even though the two spaces $N^{1,p}(X)$ and $C^{1,p}(X)$ agree (isometrically) for $p > 1$, there is one fundamental difference between them that we wish to stress. Namely,

²⁰Cheeger omitted the prefix p from the terminology.

²¹Cheeger [42] used the notation $H_{1,p}(X)$ for this space, but we prefer to reserve the letter H for spaces that are defined by a completion operation; cf. Theorem 5.2 and subsection 5.1.

²²Again, Cheeger did not use the prefix p in this terminology.

the functions in $N^{1,p}(X)$ are automatically “good” representatives analogous to Theorems 7.1 and 7.4. We next discuss the pointwise behavior of functions in $N^{1,p}(X)$.

A capacity theory can be developed in metric measure spaces in analogy with that in Section 7.1. If $1 \leq p < \infty$ and $E \subset X$, we define the p -capacity of E to be the number (possibly infinite)

$$(10.12) \quad \text{cap}_p(E) := \inf(\|u\|_{L^p(X)}^p + \|\rho_u\|_{L^p(X)}^p),$$

where the infimum is taken over all $u \in N^{1,p}(X)$ such that $u \geq 1$ on a neighborhood of E . We can talk about sets of zero p -capacity on X , as well as p -quasicontinuity of functions. It turns out that functions in $N^{1,p}(X)$ are automatically p -quasicontinuous. The preceding procedure of forming the Sobolev space $N^{1,p}(X)$ automatically produces good representatives for the Sobolev functions. Here is a precise statement for \mathbb{R}^n .

Theorem 10.8. *If $u \in N^{1,p}(\mathbb{R}^n)$, then u is p -quasicontinuous and belongs to $W^{1,p}(\mathbb{R}^n)$. Conversely, every p -quasicontinuous representative u of a Sobolev function in $W^{1,p}(\mathbb{R}^n)$ belongs to $N^{1,p}(\mathbb{R}^n)$. Moreover, the corresponding Sobolev norms agree.*

As before, we have a corresponding statement about Sobolev functions and spaces in Riemannian and sub-Riemannian contexts.

10.7. Nontriviality of $N^{1,p}(X)$. The Sobolev space $N^{1,p}(X)$ is a vector subspace of $L^p(X)$, and the identity map gives a continuous embedding,

$$(10.13) \quad N^{1,p}(X) \rightarrow L^p(X).$$

If the metric measure space X is lacking in rectifiable curves, it may happen that this embedding is onto. If X has no nonconstant rectifiable curves, then $\rho_u \equiv 0$ is the minimal p -upper gradient of every function, and we have that $N^{1,p}(X) = L^p(X)$ isometrically. This is an example of a degenerate situation, where the Sobolev space $N^{1,p}(X)$ carries no information beyond measure theory. Let us call $N^{1,p}(X)$ *nontrivial* if the embedding in (10.13) is not surjective.

One can establish the following criterion for nontriviality of $N^{1,p}(X)$. A metric measure space is said to have *nontrivial p -modulus* if there exists a curve family in the space with positive p -modulus.

Proposition 10.9. *The Sobolev space $N^{1,p}(X)$ is nontrivial if and only if X has nontrivial p -modulus.*

Proof. The necessity of the assertion is obvious. For the sufficiency, one uses the separability of X and finds that under the hypothesis of the nontriviality of the p -modulus, there must be a ball whose characteristic function cannot belong to $N^{1,p}(X)$. We omit the details. \square

Thus, the nontriviality of the Sobolev space $N^{1,p}(X)$ carries some geometric information about the underlying metric measure space. It is known that some classical fractals, such as the Sierpinski carpet and gasket, have trivial p -modulus for every $p \geq 1$, so the associated Sobolev spaces are trivial as well. These examples are interesting in that the pertinent spaces have rectifiable curves, just not enough of them.

To obtain a setting where calculus is possible, we need a space that has nontrivial modulus in a uniform scale invariant manner. Such an assumption is made precise in the next section.

10.8. Notes. Upper gradients were introduced and studied by Koskela and the author in [85], [86], [87]. The concept of a weak upper gradient was defined by Koskela and MacManus in [116]; they also noted the approximation procedure in (10.5).

Sobolev spaces $C^{1,p}(X)$ for $1 \leq p < \infty$ were introduced by Cheeger in [42]. The important concept of a minimal upper gradient was first introduced by Cheeger, in the context of the space $C^{1,p}(X)$. As mentioned in the text, Cheeger used the term *minimal generalized upper gradient*. He proved an existence and uniqueness theorem for minimal generalized upper gradients in [42, Theorem 2.10]; this result preceded the theory of minimal p -weak upper gradients. Minimal p -weak upper gradients for $p > 1$ were considered by Shanmugalingam in her thesis [162, Corollary 2.3.4], [164, Corollary 3.7]. She proved that Cheeger's minimal generalized upper gradients agree with minimal p -weak upper gradients. The proof given in this section for the existence of a minimal weak upper gradient (Proposition 10.2) is due to Hajlasz [78], who also observed that minimal weak upper gradients exist for $p = 1$ as well.

We note that Cheeger's definition for the Sobolev space does not rely on the modulus of a curve family. Independently of Cheeger, a systematic study of Sobolev spaces based on the concept of modulus and weak upper gradients was made by Shanmugalingam in [162], [163]. Theorems 10.5, 10.8, and the equality $C^{1,p}(X) = N^{1,p}(X)$, for $1 < p < \infty$, are due to her; see also [97]. (The assertion for the equality $C^{1,p}(X) = N^{1,p}(X)$ appears in [163, Theorem 4.10], but full details can only be found in [162, Chapter 2].)

I do not know if the spaces $C^{1,1}(X)$ and $N^{1,1}(X)$ have been studied more systematically. For studies of BV-functions in metric spaces, see [7], [8], [11].

The definition for (weak) upper gradients, and hence that for Sobolev spaces, readily extends to functions with values in a Banach space. For such a theory, see [88].

The question of nontriviality of modulus in a metric measure space has interesting applications beyond Sobolev spaces, especially to geometric rigidity questions. See [102], [31]. Proposition 10.9 should be well known to the experts, but may not have been explicitly recorded before.

The recent survey [78] by Hajlasz contains a more detailed discussion of Sobolev spaces $N^{1,p}(X)$. A full treatment of Sobolev spaces of general Banach-valued functions will appear in a forthcoming monograph by Koskela, Shanmugalingam, Tyson, and the author.²³ Finally, we note that the term *Newtonian space* is often used for $N^{1,p}(X)$ (because the definition is ultimately based on the fundamental theorem of calculus.)

11. POINCARÉ INEQUALITY IN METRIC MEASURE SPACES

We discuss a principal condition on a metric measure space $X = (X, d, \mu)$ that implies that X allows for calculus. This condition is an axiomatization of the classical Poincaré inequality (6.9). Recall the fact, emphasized in Section 6, that

²³*Sobolev spaces on metric measure spaces: an approach based on upper gradients.* A monograph in preparation.

the fundamental inequalities stemming from the potential estimate in (6.3) employ only the size of the gradient of a function.

11.1. Poincaré inequality. We say that a metric measure space X supports a p -Poincaré inequality, $1 \leq p < \infty$, if there exist constants $C > 0$ and $\lambda \geq 1$ such that

$$(11.1) \quad \int_B |u - u_B| d\mu \leq C \operatorname{diam}(B) \left(\int_{\lambda B} \rho^p d\mu \right)^{1/p}$$

for every ball B in X , for every function $u : X \rightarrow \mathbb{R}$, and for every upper gradient ρ of u . (Recall the mean value notation from (6.8).)

To require that (11.1) holds in X is to require that X has plenty of rectifiable curves, uniformly at all scales. The parameter p measures in a subtle way the amount of curves; it is akin to the parameter p in the definition of modulus. In fact, the validity of a Poincaré inequality can sometimes be stated in terms of modulus. In general, the connection is more suggestive than formal.

We have seen in Section 6 that \mathbb{R}^n supports a p -Poincaré inequality for each $p \geq 1$. On the other hand, the snowflake space X in (9.2) cannot support a p -Poincaré inequality for any p .

It follows from Hölder’s inequality that if a space supports a p -Poincaré inequality, it supports a q -Poincaré inequality for each $q \geq p$. On the other hand, given any two numbers $1 \leq q < p$, one can construct spaces that support a p -Poincaré inequality, but not a q -Poincaré inequality. Such constructions can be accomplished, for example, by gluing two copies of \mathbb{R}^n along an appropriate Cantor set. The size of the Cantor set dictates how many curves can connect the two copies, and this is reflected in the exponent of the Poincaré inequality.

It is a rather deep fact that in some instances the validity of a Poincaré inequality for some $p > 1$ implies a similar inequality for some smaller $q < p$. This will be discussed more in subsection 11.7.

11.2. Doubling p -Poincaré spaces. Let us call a metric measure space $X = (X, d, \mu)$ a doubling p -Poincaré space if μ is a doubling measure and if X supports a p -Poincaré inequality.

A doubling p -Poincaré space provides fertile soil for a first order calculus. In this regard, the stipulation of (11.1), in addition to the doubling condition (9.1), seems to be a correct specialization down from general spaces of homogeneous type. In subsequent sections, we will see more precisely what can be accomplished under these hypotheses.

Notice that condition (11.1) depends on the parameter p . However, many consequences of this condition seem to be independent of this parameter. The picture is not yet completely clear.

How does one recognize doubling p -Poincaré spaces? Do such spaces, apart from certain trivial or standard examples, occur naturally in mathematics? The answer to the second question is a resounding *yes*. For example, many of the singular spaces introduced in Section 8 are examples of doubling p -Poincaré spaces. The answer to the first question is more complicated. There exist techniques that can be employed here; some are similar to those which we used earlier to prove that a Poincaré inequality holds in \mathbb{R}^n . On the other hand, most of the currently known techniques are quite *ad hoc*, and there is room for improvement.

The analogous questions for spaces of homogeneous type are much easier to answer. As mentioned earlier, in Section 9, every complete doubling metric space carries a nontrivial doubling measure.

To summarize, there are rather exotic metric measure spaces, of fractional Hausdorff dimension for example, that support a Poincaré inequality and hence allow for calculus. Some of these spaces arise naturally in other areas of mathematics, and calculus is needed to solve specific problems. Some examples of metric measure spaces have been constructed by hand in order to show the robustness of condition (11.1). These latter examples raise further questions: where exactly lie the boundaries for first (or higher) order calculus?

11.3. Singular spaces supporting a Poincaré inequality. Several of the singular spaces presented in Section 8 support a Poincaré inequality.

The validity of a Poincaré inequality in Alexandrov spaces with curvature bounded from below does not seem to be well known, so I will record the following result.

Theorem 11.1. *Let X be a complete and geodesic metric space of finite Hausdorff dimension. If X has curvature at least zero, then the Hausdorff dimension of X is an integer, the corresponding Hausdorff measure is doubling, and the associated metric measure space supports a 1-Poincaré inequality.*

The measure theoretic conclusions in Theorem 11.1 can be found in [40, Chapter 10], and a proof of the Poincaré inequality in [146] (see also [120]).

Surfaces of bounded curvature in the sense of Alexandrov also support (locally) a 1-Poincaré inequality. This follows from the Bonk-Lang parametrization theorem cited earlier.

Carnot-Carathéodory spaces typically support a Poincaré inequality. See [95], [184], [79, Section 11].

The validity of a Poincaré inequality carries over to Gromov-Hausdorff limits of metric measure spaces, where a convergence of measures has to be incorporated in the definition. We will discuss this in more detail in subsection 11.6. This leads to many nontrivial examples of singular spaces supporting a Poincaré inequality, in light of the discussion in subsection 8.8 and the fact that every complete Riemannian manifold with nonnegative Ricci curvature supports a 1-Poincaré inequality [41]. See [47] for applications.

Some singular spaces that arise as boundaries of Gromov hyperbolic groups support a Poincaré inequality; in many cases, it is not known whether or not a Poincaré inequality holds. A better understanding of the issue here would be of great importance for geometric rigidity questions.

The Laakso spaces alluded to in subsection 8.11 support a Poincaré inequality.

We discuss the decomposition spaces together with topological manifolds in the next subsection.

11.4. Manifolds and the Poincaré inequality. Many of the decomposition spaces mentioned in subsection 8.9 support a Poincaré inequality. This follows from the metric nature of the constructions and from the following further result of Semmes [155].

Theorem 11.2. *Let X be a metric space that is an n -homology manifold and of Hausdorff dimension n , where $n \geq 2$. Suppose that the Hausdorff n -measure \mathcal{H}_n*

satisfies the following Ahlfors regularity condition: there exists a constant $C > 0$ such that

$$(11.2) \quad \frac{1}{C} r^n \leq \mathcal{H}_n(B(x, r)) \leq C r^n$$

for all balls $B(x, r) \subset X$ with $r < \text{diam } X$. Suppose further that X satisfies the following linear local contractibility condition: there exists a constant $C \geq 1$ such that every metric ball $B(x, r)$ in X with $r < C^{-1} \text{diam } X$ can be contracted to a point in $B(x, Cr)$. Then X supports a 1-Poincaré inequality (with respect to the Hausdorff measure).

This theorem is highly nontrivial; it also carries remarkable consequences. The hypotheses are (quantitative) topological and measure theoretic. The outcome is that significant first order calculus is possible on a space that satisfies these hypotheses. There are also geometric consequences.

According to a theorem of Sullivan, alluded to earlier in 8.2, every topological manifold outside dimension four is homeomorphic to a metric space that is a Lipschitz manifold as defined in subsection 8.2. In particular, it follows easily that every compact topological manifold outside dimension four can be metrized so as to satisfy the hypotheses in Theorem 11.2.

It was known to Whitney in the 1950s that Lipschitz charts on a manifold can be used to set up a measurable or L^∞ de Rham theory [189]. Other similar analytic tools have been developed on Lipschitz manifolds. For example, using Lipschitz (or, more generally, quasiconformally) invariant notions of geometric analysis together with ideas from noncommutative geometry, Connes, Sullivan, and Teleman constructed local formulas for certain characteristic classes on quasiconformal manifolds [53]. The consequences of Theorem 11.2 are of a somewhat different nature; they allow for global quantitative calculus on a manifold, including second order partial differential equations.

A four dimensional topological manifold need not admit a Lipschitz or even quasiconformal structure. More precisely, there are compact simply connected four manifolds that cannot be metrized so as to become Lipschitz or even quasiconformal manifolds. This follows because the Donaldson gauge theory can be extended to the setting of Lipschitz or quasiconformal manifolds [57]. However, the lack of Lipschitz structures does not preclude the possibility that metrics as in Theorem 11.2 exist on a given compact four manifold. The following question was first raised by Semmes in [155, p. 282].

11.5. Question. *Is every compact four manifold homeomorphic to a metric space that satisfies the conditions of Ahlfors 4-regularity and local linear contractibility as defined in Theorem 11.2?*

In the preceding, by Ahlfors 4-regularity we mean that (11.2) holds for $n = 4$.

An affirmative answer to this question would bring all four manifolds into the realm of analysis. Moreover, if the answer indeed is in the affirmative, it would mean that there are limits to the calculus made possible by a Poincaré inequality. Namely, one cannot hope to set up a Donaldson type gauge theory, as in the presence of Lipschitz or quasiconformal structures. In order to address Question 11.5,

one could start by investigating some of the specific simply connected four manifolds constructed by Freedman that are known not to have smooth or Lipschitz structures.²⁴

Little is known about metric structures in general on nonsmoothable four manifolds. For example, as a first step towards answering Question 11.5, one could ask if a given four manifold can be metrized so that it has locally finite Hausdorff 4-measure.²⁵ There is a conjecture, attributed to Freedman, stating that every four manifold admits Hölder continuous charts mentioned above [2, Problem 4.71]. This conjecture is distantly related to Question 11.5.

Finally, Semmes has also asked what kind of metric structures can exist on the exotic homology n -manifolds, $n \geq 5$, constructed by Bryant, Ferry, Mio, and Weinberger [39]. Can they be metrized so that the hypotheses of Theorem 11.2 are valid or just so as to have locally finite Hausdorff n -measure?

11.6. Poincaré inequality and Gromov-Hausdorff convergence. We first require the concept of a measured Gromov-Hausdorff convergence. Let $X = (X, d, \mu, x)$ be a locally compact doubling p -Poincaré space with a base point $x \in X$. For $\epsilon > 0$ we consider the pointed renormalized metric measure spaces

$$(11.3) \quad X_\epsilon := (X, \epsilon^{-1}d, \mu_\epsilon, x),$$

where

$$(11.4) \quad \mu_\epsilon(E) := \mu(B_\epsilon(x, 1))^{-1} \mu(E)$$

for $E \subset X$. Here and later B_ϵ refers to a ball in the metric $\epsilon^{-1}d$. Thus, (11.3) is analogous to rescalings of X as in (8.6), with an additional renormalization of the measure as in (11.4).

For the purposes of this article, we adapt the following definition for a measured Gromov-Hausdorff convergence.²⁶ A sequence (X_{ϵ_n}) , where $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, of pointed renormalized metric measure spaces is said to *converge in the measured Gromov-Hausdorff sense* to a pointed metric measure space $X_\infty = (X_\infty, d_\infty, \mu_\infty, x_\infty)$, if the pointed metric spaces $X_{\epsilon_n} = (X, \epsilon_n^{-1}d, x)$ Gromov-Hausdorff converge to $(X_\infty, d_\infty, x_\infty)$ as explained in subsection 8.6 and if, in addition, the following holds: for every sequence (y_n) of points such that $y_n \in X_{\epsilon_n}$ and $y_n \rightarrow y \in X_\infty$, and for every $r > 0$, we have that $\mu_{\epsilon_n}(B_{\epsilon_n}(y_n, r)) \rightarrow \mu_\infty(B_\infty(y, r))$ as $n \rightarrow \infty$, where B_∞ refers to a ball in X_∞ .

We will call every such limit space a *measured Gromov-Hausdorff limit space based at $x \in X$* .

Under the current hypotheses, one can show that measured Gromov-Hausdorff limits exist for every choice of a base point $x \in X$ (cf. subsection 8.7). Moreover, every such limit space is a proper and doubling p -Poincaré space. This latter result is fundamental for many applications of the theory; cf. subsection 12.2.

²⁴See [66], [67] for constructions of such manifolds.

²⁵Every noncompact four manifold is smoothable [67, p. 116], so the question is interesting only for compact manifolds.

²⁶Here we follow [42, Section 9]. The given definition in the context in which it is used here can be shown to coincide with the more standard definition as in [56, Section 8.6] for example.

11.7. Self-improving property of the Poincaré inequality. There are well known phenomena in harmonic analysis, where uniform and scale invariant conditions “self-improve”. The first and perhaps best known of such instances is *Gehring’s lemma*, which asserts that if a nonnegative locally integrable function in \mathbb{R}^n satisfies a uniform *reverse Hölder inequality* for some exponent, then it satisfies a similar inequality for some larger exponent. An important consequence of Gehring’s lemma is the fact that the so-called A_p -condition is an *open-ended condition*. To describe the latter, recall that a nonnegative function $w \in L^1_{loc}(\mathbb{R}^n)$ is said to be an A_p -weight, $1 \leq p < \infty$, if there exists a constant $C \geq 1$ such that

$$(11.5) \quad \int_B w \, dx \leq C \left(\int_B w^{1/(1-p)} \, dx \right)^{1-p}$$

for each ball B in \mathbb{R}^n . (For $p = 1$, the integral in the right hand side is taken to mean the essential infimum of w over the ball B .) Obviously, an A_p -weight is an A_q -weight for every $q \geq p$, but it is also true that every A_p -weight, $p > 1$, is an A_q -weight for some $q < p$ (depending on the weight). This latter fact is nontrivial.

Self-improving inequalities of the Gehring lemma type are widely used in many proofs for the regularity properties of solutions to partial differential equations, for example. Weights that satisfy a reverse Hölder inequality also have a connection to Poincaré inequalities in \mathbb{R}^n . Namely, if w is an A_p -weight in \mathbb{R}^n , then $d\mu = w \, dx$ determines a doubling measure that satisfies a p -Poincaré inequality. By the open-ended property, we have that $d\mu$ satisfies a q -Poincaré inequality for some $q < p$ if $p > 1$.

Recently, Keith and Zhong proved that such a self-improving property holds in the context of a Poincaré inequality in general [104]. More precisely, they proved the following.

Theorem 11.3. *Let $p > 1$ and let X be a complete and doubling p -Poincaré space. Then there exists $q < p$ such that X supports a q -Poincaré inequality. The statement is quantitative in the sense that q and other parameters involved in the conclusion depend only on the parameters in the hypotheses.*

Theorem 11.3 has several important consequences; we will mention some of them later. The proof for Theorem 11.3 is rather intricate involving several rescaling and limiting arguments for metric spaces. Moreover, the result is nontrivial already in \mathbb{R}^n , where it is still not known whether a doubling measure that supports a p -Poincaré inequality has to be absolutely continuous with respect to Lebesgue measure. In fact, all known (to the author at least) doubling measures μ in \mathbb{R}^n that satisfy a p -Poincaré inequality for some $p \geq 1$ are of the form $d\mu = w \, dx$ for a weight w in some A_q (the exponents p and q need not be the same here).

11.8. The Poincaré inequality and removability. Theorem 11.3 is not true without the assumption that X be complete as a metric space. Given any number $p > 1$ there exists a locally compact doubling metric measure space X that supports a p -Poincaré inequality but does not support a q -Poincaré inequality for any $q < p$. One can choose X to be a domain in Euclidean space, with Euclidean distance and Lebesgue measure. This remark follows from the following theorem of Koskela, which illustrates how the concept of a Poincaré inequality sheds new light on classical questions as well [113].

We say that a compact set $K \subset \mathbb{R}^n$ is *removable for the Sobolev space* $W^{1,p}$ if every function in $W^{1,p}(\mathbb{R}^n \setminus K)$ has an extension to a function in $W^{1,p}(\mathbb{R}^n)$.

Theorem 11.4. *Let K be a compact set of measure zero in \mathbb{R}^n , $n \geq 2$, and let $p > 1$. Then $\mathbb{R}^n \setminus K$ supports a p -Poincaré inequality if and only if K is removable for the Sobolev space $W^{1,p}$. Moreover, there are compact sets of measure zero that are removable for $W^{1,p}$, but not for $W^{1,q}$ for any $q < p$.*

11.9. Notes. An abstract Poincaré inequality as in (11.1) was introduced together with upper gradients by Koskela and the author [85], [86]. Naturally, Poincaré type inequalities in various situations have a long history in analysis. There are several rather deep results in geometric analysis that are known to require only a Poincaré type inequality and a doubling measure. See, for example, [152], [52], [16], and the many references there. See also [115].

For examples of doubling p -Poincaré spaces, see [86], [79], and references therein. Semmes's article [155] introduced an important method to verify the validity of a Poincaré inequality; this method has been used later e.g. in [35], [121]. For an earlier article in this regard, see [55]. There are by now several equivalent ways to express inequality (11.1). See [86], [79], [87], [99], [103].

References related to singular spaces and Poincaré inequalities have been given in the text of this section. The fact that a Poincaré inequality passes to limit spaces was first proved by Cheeger [42, Theorem 9.6, p. 486] in a slightly weaker form than what is formulated in subsection 11.6. Later Cheeger (unpublished), as well as Keith and Koskela, independently, found the optimal statement; see [99, Theorem 3, p. 260], [114].

For a survey on Gehring's lemma and its use in analysis, see [94]. For the use of A_p -weights and Poincaré inequalities in partial differential equations, see [83].

For removability questions for Sobolev functions in abstract settings, see [118].

12. MEASURABLE COTANGENT STRUCTURES AND DIFFERENTIABILITY

In this section, we discuss Cheeger's work which implies that every doubling p -Poincaré space admits a measurable cotangent structure. The existence of such a structure is an example of a geometric restriction that spaces that support a Poincaré inequality must satisfy. A crucial aspect of this work is a careful analysis of the infinitesimal behavior of Lipschitz functions.

I reiterate what was already mentioned in subsection 10.6 (see also Notes to Section 10) that in the ensuing results Cheeger used a different definition both for the Sobolev space and for the minimal p -weak upper gradient. For $p > 1$, the concepts are the same, however, and we will make this standing assumption for the rest of this section. (In fact, as will become clear, such an assumption entails no loss of generality in what follows.)

12.1. Rademacher's theorem. Recall the definition of rescaled functions and tangent functions from (8.7) and (8.8). If X is \mathbb{R}^n , then Rademacher's classical theorem (cf. subsection 8.7) implies that every tangent function of a Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a unique linear map at almost every point.

In doubling Poincaré spaces, we have the following remarkable extension of Rademacher's result due to Cheeger.

Theorem 12.1. *In a doubling p -Poincaré space, there are, up to a decomposition of the space into countably many measurable parts, finitely many Lipschitz functions whose tangent functions generate the vector space of the tangent functions of arbitrary Lipschitz functions almost everywhere. The cardinality of the generating set of functions has an upper bound depending only on the data in the hypotheses of the theorem.*

It is necessary to make the assertion in the preceding theorem more precise. To this end, recall the two pointwise infinitesimal Lipschitz constants from (10.3) and (10.4). Also recall the notation for the minimal p -weak upper gradient from subsection 10.4.

Theorem 12.2. *Let $X = (X, d, \mu)$ be a doubling p -Poincaré space. Then there exists a countable decomposition of X into disjoint measurable sets U_α , $\alpha \in \mathcal{A}$, together with Lipschitz mappings $X_\alpha : U_\alpha \rightarrow \mathbb{R}^{N(\alpha)}$ such that the following holds: with every Lipschitz function $f : X \rightarrow \mathbb{R}$ there is associated a collection of bounded functions $d_\alpha f : U_\alpha \rightarrow \mathbb{R}^{N(\alpha)}$, $\alpha \in \mathcal{A}$, such that*

$$(12.1) \quad \lim_{y \rightarrow x} \frac{f(x) - f(y) - \langle d_\alpha f(x), X_\alpha(x) - X_\alpha(y) \rangle}{d(x, y)} = 0$$

for almost every $x \in U_\alpha$. Moreover, for almost every $x \in U_\alpha$, we have

$$(12.2) \quad \text{Lip}f(x) = \text{lip}f(x) = \rho_f(x)$$

and there is a norm $|\cdot|_x$ in $\mathbb{R}^{N(\alpha)}$ such that

$$(12.3) \quad |d_\alpha f(x)|_x = \rho_f(x).$$

Finally, the integers $N(\alpha)$ are bounded from above by a finite constant depending only on the constants associated with the Poincaré inequality and the doubling condition.

The information in Theorem 12.2 can be reformulated as follows.

Theorem 12.3. *With every doubling p -Poincaré space X there is canonically associated a measurable Banach space bundle T^*X with fibers of uniformly bounded dimension such that every Lipschitz function $f : X \rightarrow \mathbb{R}$ determines a bounded measurable section df of the bundle and that*

$$(12.4) \quad |df(x)|_x = \text{Lip}f(x) = \rho_f(x)$$

for almost every $x \in X$.

The preceding results have several notable consequences. The existence of a nontrivial differential as in (12.4) when coupled with the validity of a Poincaré inequality allows for differential analysis on X akin to that in classical situations. One can set up second order differential equations on X and run the basic elliptic regularity arguments of De Giorgi and Moser type, for example. We make more remarks on this in Section 15.

Note that the exponent p in the hypotheses of the preceding theorems has no role in the conclusions. (Except in formulas (12.3) and (12.4), where the minimal p -weak upper gradient depends a priori on p . The conclusion is of course that it is independent of p .) In subsection 12.3, we discuss a partial extension of these results, where the hypotheses are weakened so as not to depend on p .

There are several interesting open questions about the cotangent bundle T^*X , especially about the measure theoretic properties of the *Cheeger coordinates* $X_\alpha : U_\alpha \rightarrow \mathbb{R}^{N(\alpha)}$. There should be more information available than what is currently known, perhaps under some additional assumptions on X . In this direction, Keith has shown that if X is a complete and doubling p -Poincaré space, then the components of the Cheeger coordinates can be taken to be distance functions [101].

We will discuss the proof of Theorem 12.2 in the next two subsections.

12.2. Differentiability and infinitesimal structure of Lipschitz functions.

Cheeger's theorem 12.2 has more information than just the existence of a nontrivial cotangent structure. Namely, it implies that we have stability in the infinitesimal Lipschitz constants and that these agree with the minimal upper gradient; this is expressed in (12.2). Moreover, analogous to the Euclidean case, the tangent functions of a Lipschitz function have a special structure called *generalized linearity*. These facts are perhaps the deepest among Cheeger's findings in [42]. We will discuss them briefly in this subsection.

Let $X = (X, d, \mu)$ be a metric measure space and let $1 < p < \infty$. A function $u : X \rightarrow \mathbb{R}$ is said to be *p -harmonic in X* if, for every bounded open set $\Omega \subset X$, we have that $u \in N^{1,p}(\Omega)$ and that

$$(12.5) \quad \int_{\Omega} \rho_u^p d\mu \leq \int_{\Omega} \rho_v^p d\mu$$

for every $v \in N^{1,p}(\Omega)$ such that $u - v$ vanishes in a neighborhood of $X \setminus \Omega$. If $X = \mathbb{R}^n$, then u is 2-harmonic in the above sense if and only if u is a harmonic function in the classical sense (up to a redefinition in a set of measure zero). This follows from the remarks made in Section 10.4 and from the classical interpretation of harmonic functions as local minimizers of the Dirichlet energy functional

$$u \mapsto \int |\nabla u|^2.$$

More generally, if $X = \mathbb{R}^n$, then u is p -harmonic as defined in the preceding if and only if u is in $W^{1,p}(\Omega)$ for every bounded open set Ω and satisfies the *p -harmonic equation*

$$(12.6) \quad -\operatorname{div}(|\nabla u|^{p-2} \nabla u) = 0$$

in the sense of distributions.

A Lipschitz function $L : X \rightarrow \mathbb{R}$ is called *generalized p -linear* if the following three conditions hold: (i) either $L \equiv 0$ or $L(X) = \mathbb{R}$, (ii) L is p -harmonic, and (iii) $\rho_L \equiv \text{constant}$.

One can easily show that the generalized p -linear functions in \mathbb{R}^n are precisely the ordinary linear functions. Indeed, a linear function $L : \mathbb{R}^n \rightarrow \mathbb{R}$ obviously satisfies (i) – (iii). On the other hand, if $L : \mathbb{R}^n \rightarrow \mathbb{R}$ is generalized p -linear, then $\rho_L = |\nabla L| \equiv c$ and we have that

$$0 = -\operatorname{div}(|\nabla L|^{p-2} \nabla L) = -\operatorname{div}(c^{p-2} \nabla L) = -c^{p-2} \Delta L,$$

so that L is harmonic and hence smooth. Then a direct computation using the harmonicity gives that

$$(12.7) \quad 0 = \Delta |\nabla L|^2 = 2|\operatorname{Hess}(L)|^2,$$

where Hess denotes the Hessian and where the Euclidean norm for the matrix Hess(L) is used. It follows that L must be linear. (Note that (12.7) is the so-called *Bochner formula* in the special case of \mathbb{R}^n . Compare [43, p. 19].)

A proof of the preceding statement can be given without using the smoothness of generalized linear functions. See [42, Theorem 8.11, p. 483]. The fact that this can be done ultimately means that Cheeger’s proof for the Rademacher theorem is essentially different from the previously known proofs even in \mathbb{R}^n .

Next, we have the following theorem.

Theorem 12.4. *Let $X = (X, d, \mu)$ be a locally compact doubling p -Poincaré space and let $f : X \rightarrow \mathbb{R}$ be a Lipschitz function. Then for almost every point $x \in X$ and for every measured Gromov-Hausdorff limit X_∞ based at x , every tangent function $f_\infty : X_\infty \rightarrow \mathbb{R}$ of f is generalized p -linear. Moreover, the minimal p -weak upper gradient of f_∞ satisfies $\rho_{f_\infty} \equiv \text{Lip}f(x)$.*

In the preceding formulation of the theorem, it is understood that the tangent function f_∞ is obtained through the same sequence of scalings of (X, d, μ, x) as the measured Gromov-Hausdorff limit (cf. subsections 8.7 and 11.6).

One should note the remarkable fact that the minimal p -weak upper gradient of f_∞ is independent of the particular limit space, tangent function, and p .

Next I will briefly explain the ideas behind the proof of Theorem 12.4. The actual rigorous proof is quite intricate in [42], and we have to ignore many technical points.

The first step in showing that the tangent function f_∞ is generalized p -linear is to establish that the Lipschitz function f asymptotically minimizes the p -energy at almost all points in X . More precisely, this means that

$$(12.8) \quad \lim_{r \rightarrow 0} \left(\int_{B(x,r)} \rho_f^p d\mu - \inf_{v_r} \int_{B(x,r)} \rho_{v_r}^p d\mu \right) = 0$$

for almost every $x \in X$, where the infimum is taken over all functions $v_r \in N^{1,p}(X)$ such that $f - v_r$ vanish in a neighborhood of $X \setminus B(x, r)$.²⁷ To prove that a Lipschitz function satisfies (12.8) almost everywhere, one relies on two ingredients: a Vitali covering theorem and the lower semicontinuity property for the Sobolev norm (10.7). The idea is that if (12.8) failed in a set of positive measure, then by adding essentially disjoint small bump functions to f one could create a bounded sequence of functions in the Sobolev space, converging to f in L^p , such that the Sobolev norm for the members in the sequence would be less than the Sobolev norm of f . Such conclusion contradicts (10.7). This elegant argument is very general and works for every metric space where a Vitali covering theorem holds.²⁸

The asymptotic p -harmonicity as in the preceding can be shown to imply the global p -harmonicity for the tangent function, but crucial in that argument is the knowledge that we have stability in the infinitesimal Lipschitz constant and that this constant agrees almost everywhere with the minimal p -weak upper gradient; see (12.2). It is at the points of density of the minimal p -weak upper gradient that tangent functions are considered. We cannot enter the technical discussion of this argument here. It suffices to say that the Poincaré inequality is used decisively in this argument.

²⁷Functions satisfying (12.8) are called *asymptotically p -harmonic at x* in [42, p. 446].

²⁸According to Cheeger, it was the demands of this proof in [42, Theorem 3.7, p. 447] that led him to define Sobolev spaces such that the lower semicontinuity of norms as in (10.7) holds automatically; see (10.11).

For a technically well versed reader, I would like to emphasize that the Poincaré inequality enters the proof in the form of a “telescoping argument” which gives semi-global Lipschitz type control on Sobolev functions in terms of the upper gradient. In particular, one obtains that locally Lipschitz functions are dense in the Sobolev space and moreover that the minimal p -weak upper gradient of a Sobolev function can be approximated in L^p by continuous upper gradients of an approximating sequence of Lipschitz functions. Note that for continuous upper gradients ρ the inequality $\text{Lip}f(x) \leq \rho(x)$ is immediate; it is a rather elaborate approximation procedure that gives the same in general. See [42, Section 10] for the details.

12.3. Discussion of proof for Theorem 12.2. I will discuss a part of the proof of Theorem 12.2. (In particular, (12.2) will not be discussed; cf. the preceding subsection.) The ensuing argument is from Keith’s paper [100], although the underlying ideas are similar to those of [42].

The starting point is the following property of Lipschitz functions on doubling p -Poincaré spaces: there exists a constant $K \geq 1$ such that

$$(12.9) \quad \text{Lip}f(x) \leq K \text{lip}f(x)$$

for all Lipschitz functions $f : X \rightarrow \mathbb{R}$, for almost every $x \in X$.

It is instructive to explain why a p -Poincaré inequality implies (12.9). Because $\text{lip}f$ is an upper gradient of a Lipschitz function f , (11.1) implies

$$\frac{1}{\text{diam}B} \int_B |f - f_B| d\mu \leq C \left(\int_{\lambda B} (\text{lip}f)^p d\mu \right)^{1/p}.$$

By the Lebesgue differentiation theorem, the right hand side of this inequality tends to $\text{lip}f(x)$ for almost every $x \in X$, where $B = B(x, r)$ and $r \rightarrow 0$. On the other hand, by using the theorems of Egoroff and Lusin, one can show that for almost every point x the mean oscillation of a Lipschitz function, or the left hand side of the preceding inequality, is close to $\text{Lip}f(x)$ for all small enough balls centered at x . In fact, the same argument shows that $\text{Lip}f(x) \leq C \rho_f(x)$ for almost every $x \in X$ for some constant depending only on the constants associated with the Poincaré inequality and the doubling condition. With a slightly different argument, using maximal functions, this was first proved in [42, Proposition 4.26].

Now we have the following partial generalization of Theorem 12.2.

Theorem 12.5. *Let $X = (X, d, \mu)$ be a doubling metric measure space such that condition (12.9) holds for some $K \geq 1$. Then there exists a countable decomposition of X into disjoint measurable sets U_α , $\alpha \in \mathcal{A}$, together with Lipschitz mappings $X_\alpha : U_\alpha \rightarrow \mathbb{R}^{N(\alpha)}$ such that the first assertion including (12.1) in Theorem 12.1 holds. The integers $N(\alpha)$ are bounded from above by a finite constant depending only on K and the constant associated with the doubling condition.*

There is also an analog of Theorem 12.3 under the hypotheses of Theorem 12.5. On the other hand, there is no analog of (12.2) in this generality, for the hypothesis (12.9) allows for spaces where upper gradients may degenerate; it allows for totally disconnected spaces, for example.

It is not well understood what exactly is required for nontrivial cotangent structures as in Theorem 12.3 to exist. An abstract functional analytic approach to this question is described in the next section.

Let us now return to the idea of proof based on (12.9). First one shows by a rescaling argument that condition (12.9) passes on to tangent functions as a global

condition: if $f : X \rightarrow \mathbb{R}$ is Lipschitz, then

$$(12.10) \quad \text{Lip}f(x) \simeq \text{Lip}(f_\infty) \simeq \text{var}(f_\infty : B(z, r))$$

for every ball $B(z, r) \subset X_\infty$. Here $\text{Lip}(f_\infty)$ denotes the (global) Lipschitz number of f_∞ as defined in (10.2),

$$(12.11) \quad \text{var}(f_\infty : B(z, r)) := \sup_{y \in B(z, r)} \frac{|f_\infty(z) - f_\infty(y)|}{r}$$

is the *variation* of the map f_∞ in a ball, and \simeq means that the three numbers in (12.10) are comparable up to multiplicative constants that depend on the original data only. Moreover, we can only ascertain that (12.10) holds for tangent spaces and maps at almost every point x in X .

By employing the fact that the tangent spaces to X are also doubling, a packing type argument shows that the vector space spanned by tangent functions satisfying (12.10) must be finite dimensional, with an upper bound on the dimension depending only on the original data.

From these finite conclusions, by a maximality argument, one further obtains that (almost everywhere) every tangent function of a Lipschitz function must lie in a finite dimensional space as required in the theorem.

In sum, the existence of (nontrivial) cotangent structures, as in the preceding results, owes to rescaling and packing arguments, where one repeatedly uses the precompactness of doubling metric spaces with measure theory mixed in. The scale invariant assumptions, whether the doubling condition, Poincaré inequality, or condition (12.9), give stability in the difference quotients formed as in (8.7).

We reiterate that proofs for the infinitesimal stability of Lipschitz functions on doubling p -Poincaré spaces, e.g. property (12.2) and Theorem 12.4, require more sophisticated analysis that we must forgo here.

In the next section, we will see that there is canonically associated with every metric measure space a (possibly degenerate) cotangent structure. This conclusion is achieved by pure functional analytic methods.

12.4. Density of Lipschitz functions. One can show that in a doubling p -Poincaré space X , the Sobolev space $N^{1,p}(X)$ for $1 \leq p < \infty$ equals the closure of locally Lipschitz functions $f : X \rightarrow \mathbb{R}$ in the Sobolev norm. In this case, we have that

$$\|f\|_{1,p} = \left(\int_X |f(x)|^p d\mu(x) \right)^{1/p} + \left(\int_X |df(x)|_x^p d\mu(x) \right)^{1/p}$$

for such functions by (12.3) and (12.4). I do not know whether locally Lipschitz functions are always dense in $N^{1,p}(X)$ for an arbitrary metric measure space X .

12.5. Reflexivity of Sobolev spaces. If X is a doubling p -Poincaré space, it follows from Theorem 12.3 that the Sobolev spaces $N^{1,p}(X)$ are reflexive for $p > 1$. Indeed, the norms on fibers have uniformly bounded dimension and (the upper gradient part of) the Sobolev norm is an integral norm using the pointwise defined norms on fibers. The reflexivity of the Sobolev space $N^{1,p}(X)$ is used in a crucial way in the proof of (12.2) in [42, Section 5].

While the p -Poincaré hypothesis reflects the fact that X must have lots of rectifiable curves, it is not correct to view the reflexivity in this light. Indeed, if X has no rectifiable curves, or few such, then $N^{1,p}(X) = L^p(X)$ and hence is reflexive for

$p > 1$. (See Proposition 10.9.) Moreover, it follows from Theorem 12.5 (as in the previous paragraph) that $N^{1,p}(X)$ for $p > 1$ is reflexive under the hypotheses of that theorem.

It is not clear exactly under what assumptions on X we have reflexivity of the Sobolev space $N^{1,p}(X)$. The next example shows that this need not always be the case. Let

$$X = \prod_{n \geq 0} [0, a_n]$$

for some sequence $a = (a_n)$ of real numbers such that $a_n \rightarrow 0$ as $n \rightarrow \infty$. We view X as a compact subset of the Banach space c_0 consisting of all sequences $x = (x_n)$ of real numbers, $\lim_{n \rightarrow \infty} x_n = 0$, with $\|x\| := \max |x_n|$. There is a natural product (probability) measure μ on X , that is the product of normalized Lebesgue measures on each factor $[0, a_n]$. Now consider a sequence (f_n) of functions $f_n : X \rightarrow \mathbb{R}$ that are the projections on the factors; i.e., $f_n(x) = x_n$ for $x = (x_n) \in X$. Because each function f_n is 1-Lipschitz on X , and bounded by $\|a\|$, the sequence (f_n) is a bounded sequence in $N^{1,p}(X)$. Assuming that the Sobolev space in question is reflexive, a weakly convergent subsequence could be found with weak limit f . By passing to Mazur's lemma [191, p. 126], we would further find a sequence (g_m) consisting of convex combinations of the functions f_n ,

$$g_m = \lambda_{m,1} f_{m_1} + \cdots + \lambda_{m,k_m} f_{m_{k_m}},$$

where $0 \leq \lambda_{m,j}$ and $\lambda_{m,1} + \cdots + \lambda_{m,k_m} = 1$, with $m_{k_i} \rightarrow \infty$ as $m \rightarrow \infty$, such that $g_m \rightarrow f$ strongly in $N^{1,p}(X)$. It follows that $f \equiv 0$. Moreover, from the norm structure it is easy to check that the minimal upper gradient of each g_m is identically $\lambda_{m,1} + \cdots + \lambda_{m,k_m} = 1$, which contradicts the fact that $g_m \rightarrow 0$ in $N^{1,p}(X)$.

Stephen Keith has asked the interesting question whether $N^{1,p}(X)$ is always reflexive for $p > 1$ if X is a doubling metric space.

12.6. Notes. The classical theorem of Rademacher dates from 1919 [145]. For the standard proof, see e.g. [64], [62]. As mentioned in the text, Theorems 12.1 – 12.4 are due to Cheeger [42]. The concept of an (abstract) p -harmonic function and a generalized p -linear function was also introduced by Cheeger in [42]. For the Euclidean theory of p -harmonic functions, see [83]. Theorem 12.5 is due to Keith [100]. Unlike the proof of Theorem 12.5, the proofs for (12.2) and for Theorem 12.4 use the Poincaré inequality in a substantial way. It is not known if there are essentially weaker hypotheses on X (than what is assumed in Theorem 12.2) that give (12.2).

There is an important classical generalization of Rademacher's theorem, due to Stepanov [64]. For an extension of this theorem to metric measure spaces, see [20].

Cheeger's proof of Rademacher's theorem in [42] is new even in the case $X = \mathbb{R}^n$. This is carefully explained in the introduction to [42]. One of the points is that, as already mentioned in the text, the generalized linear functions in \mathbb{R}^n can be shown to be linear without using the fact that linear functions are differentiable or smooth; cf. [42, Remark 8.14]. A similar intrinsic characterization of linearity was used earlier by Cheeger and Gromoll in their celebrated splitting theorem [49]. In this connection, one should also mention works by Mostow [137], [138], and Pansu [143], where differentiability theorems in non-Riemannian settings were proved.

As mentioned in subsection 12.2, the proof (in [42, Theorem 3.7, p. 447]) for the asymptotic p -harmonicity of Lipschitz functions works in every metric measure

space where a Vitali type covering theorem holds (as formulated in [81, Theorem 1.6, p. 3] for example). Thus, it holds in \mathbb{R}^n for every Radon measure and in many infinite dimensional cases as well. See [133, p. 34 and p. 40].

Cheeger has conjectured that in Theorem 12.2 the Lebesgue measure in $\mathbb{R}^{N(\alpha)}$ of the image $X_\alpha(U_\alpha)$ of each “chart” U_α is positive; see [42, Conjecture 4.63, p. 463]. Cheeger’s paper contains more information about doubling p -Poincaré spaces, especially about their fine scale structure, than what is mentioned here. For applications of Cheeger’s work to Riemannian geometry, see [47].

The proof that is sketched in subsection 12.3 is from Keith’s paper [100]. The argument moreover allows for underlying measures that are not necessarily doubling but satisfy a condition called *chunkiness* in [100]. See [100] for details.

The density of Lipschitz functions in the present context, as explained in 12.4, was proved both by Cheeger [42] and by Shanmugalingam [163]. Recall that Cheeger used a different definition for a Sobolev space, as explained in subsection 10.6. See also [65].

The reflexivity of $N^{1,p}(X)$ for doubling p -Poincaré spaces, $p > 1$, is due to Cheeger; as explained in the text, this fact essentially follows from Theorem 12.3 [42]. For similar reasons, $N^{1,p}(X)$, $p > 1$, is reflexive under the weaker hypotheses on X appearing in Theorem 12.5; see [100]. The example in 12.5 of a nonreflexive Sobolev space $N^{1,p}(X)$, $p > 1$, seems to be the first such known. (See also Section 14.)

13. LIPSCHITZ ALGEBRAS AND DERIVATIONS

There is a purely algebraic construct, discovered by Weaver, that leads to a canonical cotangent structure on every metric measure space. In the present section, we will discuss this construct and its produce, a measurable exterior derivative defined on an algebra of bounded Lipschitz continuous functions. We point out how the measurable exterior derivative can be used to associate a Sobolev space with every metric measure space X in a canonical way.

13.1. Measurable derivations and the exterior differential. Let $X = (X, d)$ be a metric space. Denote by $\text{Lip}^\infty(X)$ the Banach space of all bounded Lipschitz functions $f : X \rightarrow \mathbb{R}$ with norm

$$\|f\|_L := \max\{\text{Lip}(f), \|f\|_\infty\},$$

where $\text{Lip}(f)$ is the Lipschitz number of f defined in (10.2).

The space $\text{Lip}^\infty(X)$ can be isometrically realized as a weak*-closed subspace of the Banach space

$$L^\infty(X) \oplus L^\infty(X \times X \setminus \{\text{diagonal}\})$$

by the embedding

$$f \mapsto (f, \text{difference quotient map})$$

with self-explanatory terminology. Here $L^\infty(A)$ refers to the Banach space of bounded functions on a set A . Note that $L^\infty(A)$ is the dual of $L^1(A)$, the Banach space of absolutely summable functions on A . The preceding embedding understood, it follows from standard functional analytic arguments that $\text{Lip}^\infty(X)$ is itself a dual Banach space and thus owns a weak*-topology. One can show that the weak*-topology on norm bounded sets in $\text{Lip}^\infty(X)$ agrees with the weak*-topology of $L^\infty(X)$ and that weak*-convergence is the same as pointwise convergence of functions with uniformly bounded norm.

Assume now that $X = (X, d, \mu)$ is a metric measure space. Associated with this triple there is a Banach space $\Upsilon(X)$ of all *derivations*,

$$\delta : \text{Lip}^\infty(X) \rightarrow L^\infty(X, \mu),$$

which are by definition bounded and weak* continuous linear mappings that satisfy

$$\delta(fg) = f \delta(g) + g \delta(f)$$

for all $f, g \in \text{Lip}^\infty(X)$. The boundedness means the usual boundedness of the Banach space theory,

$$(13.1) \quad \sup_{\|f\|_L \leq 1} \|\delta(f)\|_{L^\infty(X, \mu)} =: \|\delta\| < \infty,$$

and the weak*-continuity in this case means that

$$\int_X \delta(f_i) g \, d\mu \rightarrow \int_X \delta(f) g \, d\mu$$

for all $g \in L^1(X, \mu)$ whenever (f_i) is a net in $\text{Lip}^\infty(X)$ such that

$$\sup_i \|f_i\|_L < \infty \quad \text{and} \quad f_i \rightarrow f \text{ pointwise.}$$

The space of derivations is a module over $L^\infty(X, \mu)$ with the obvious action:

$$(b\delta)(f) := b\delta(f), \quad b \in L^\infty(X, \mu).$$

The dual module

$$\Omega(X) := \Upsilon(X)'$$

consists of all bounded module maps

$$\alpha : \Upsilon(X) \rightarrow L^\infty(X, \mu)$$

with the norm as in (13.1),

$$\sup_{\|\delta\| \leq 1} \|\alpha(\delta)\|_{L^\infty(X, \mu)} =: \|\alpha\|.$$

It can be shown that $\Omega(X)$ is a dual Banach space and so equipped with a weak*-topology.

The members of $\Upsilon(X)$ are called the *measurable vector fields* on X , and the members of $\Omega(X)$ are called the *measurable 1-forms* on X .

The two algebraic constructs, $\Upsilon(X)$ and $\Omega(X)$, have both the structure of a Banach space and a module over $L^\infty(X, \mu)$. They are canonically attached to each given metric measure space, and in fact they depend only on the given measure class; in other words, two mutually absolutely continuous measures on a metric space X give rise to the same modules of derivations and 1-forms.

Next, we consider a weak*-continuous bounded linear map

$$d : \text{Lip}^\infty(X) \rightarrow \Omega(X)$$

defined by

$$df(\delta) := \delta(f).$$

The measurable 1-form $df \in \Omega(X)$ is called the *measurable exterior differential* of a Lipschitz function f .

It can be shown that the exterior derivative operator d , as defined above, is in a sense a universal derivation on the algebra $\text{Lip}^\infty(X)$.

Perhaps surprisingly, the abstract constructs $\Upsilon(X)$ and $\Omega(X)$ harbour nontrivial geometric information about the underlying metric measure space. We will discuss some examples in subsection 13.2.

Theorem 13.1. *There is a unique map*

$$(13.2) \quad |\cdot| : \Omega(X) \rightarrow L^\infty(X, \mu)$$

with the following properties:

$$|\alpha| \geq 0, \quad \|\alpha\| = \|\alpha\|_\infty, \quad |f\alpha| = |f||\alpha|$$

for every $\alpha \in \Omega(X)$ and $f \in L^\infty(X, \mu)$. Here $\|\alpha\|$ denotes the (dual) norm on $\Omega(X)$.

The function $|\cdot|$ in (13.2) should be considered as a pointwise defined norm on measurable sections of a measurable Banach bundle over X .

Theorem 13.2. *If $\Omega(X)$ as a module over $L^\infty(X, \mu)$ is finite dimensional, then there exists a measurable partition $X = \cup_{i=1}^N X_n$ such that associated with almost every $x \in X_n$ there exists an n -dimensional Banach space $(V_x, |\cdot|_x)$ and an isometric isomorphism from $\Omega(X)$ to bounded measurable functions (sections)*

$$s : X \rightarrow \cup_x V_x, \quad s(x) \in V_x,$$

where

$$\|s\| := \text{ess sup} |s(x)|_x.$$

It seems to be unknown when exactly one can realize $\Omega(X)$ as the set of sections on a measurable Banach bundle where fibers are allowed to be infinite dimensional [188, p. 74].

The key property emerging from the definitions is the following *locality*:

Proposition 13.3. *For every Borel set $A \subset X$ of positive measure we have that*

$$(13.3) \quad \Upsilon(A) = \chi_A \cdot \Upsilon(X).$$

It is instructive to see how (13.3) follows from abstract principles. The proof relies on two facts. The first fact is that every Lipschitz function $f : A \rightarrow \mathbb{R}$ can be extended to a Lipschitz function $F : X \rightarrow \mathbb{R}$ such that $\|F\|_L = \|f\|_L$. (This is the well known *McShane-Whitney extension lemma*.)

The second fact is that every derivation $\delta \in \Upsilon(X)$ can be localized as in the following lemma. (Compare Lemma 10.3.)

Lemma 13.4. *If $f \in \text{Lip}^\infty(X)$ satisfies $f|_A = c$ for some constant $c \in \mathbb{R}$, then $\delta(f)|_A = 0$ almost everywhere.*

Proof. Because $\delta(1) = 0$, we can assume that $c = 0$. Let $f \in \text{Lip}^\infty(X)$ be such that $f|_A = 0$. Fix $\epsilon > 0$ and consider $\varphi_\epsilon(t) := \sqrt{t}$ for $t \geq \epsilon$, $\varphi_\epsilon(t) := \epsilon^{-1/2}t$ for $0 \leq t < \epsilon$, and extend φ_ϵ to be an odd function $\varphi_\epsilon : \mathbb{R} \rightarrow \mathbb{R}$. Define $f_\epsilon := \varphi_\epsilon \circ f$ and set

$$g_\epsilon := |f_\epsilon| \cdot f_\epsilon, \quad 0 < \epsilon < 1.$$

It is straightforward to check that

$$\|g_\epsilon\|_L \leq 10 \max\{\|f\|_L, 1\}.$$

Because also $g_\epsilon \rightarrow f$ pointwise, as $\epsilon \rightarrow 0$, we have that $g_\epsilon \rightarrow f$ weak* in $\text{Lip}^\infty(X)$. Therefore, by the definition of the derivation δ , we have that $\delta(f)$ is the weak*-limit of the functions $2f_\epsilon \cdot \delta(f_\epsilon)$, as $\epsilon \rightarrow 0$, in $L^\infty(X, \mu)$. Because $f_\epsilon|_A = 0$ for every $\epsilon > 0$, we have that $\delta(f)|_A = 0$. The lemma follows. \square

With these remarks, the proof for the locality (13.3) is straightforward.

13.2. A Sobolev space. By the preceding discussion, we can associate with each metric measure space X a Banach space $\Omega(X)$ whose members are called measurable 1-forms. A pointwise defined “fiber norm” $|\alpha| \in L^\infty(X, \mu)$ can be defined for elements $\alpha \in \Omega(X)$. Moreover, there is an exterior derivative that associates with each bounded Lipschitz function f on X an element df of $\Omega(X)$.

It is now easy to define a Sobolev space $O^{1,p}(X)$, $1 \leq p < \infty$, as the closure of $\text{Lip}^\infty(X)$ under the norm

$$(13.4) \quad \|f\|_{1,p} := \left(\int_X |f|^p d\mu \right)^{1/p} + \left(\int_X |df|^p d\mu \right)^{1/p}.$$

More precisely, we consider those functions in $\text{Lip}^\infty(X)$ for which the norm in (13.4) is finite and then form a Banach space completion.

In addition to the underlying metric, the construction of $\Omega(X)$ depends only on the measure class of μ . Thus, if μ is a measure on \mathbb{R}^n of the form $d\mu(x) = w(x) dx$ for some almost everywhere positive locally integrable function w , then $\Omega(\mathbb{R}^n, \mu) = \Omega(\mathbb{R}^n)$ is the collection of differential 1-forms with bounded measurable coefficients, and we have

$$(13.5) \quad O^{1,p}(\mathbb{R}^n; \mu) = H^{1,p}(\mathbb{R}^n; \mu)$$

as defined in subsection 5.1. In particular, we cannot always expect that the members of $O^{1,p}(X)$ are represented by L^p -functions.

The relationship between the spaces $O^{1,p}(X)$ and $N^{1,p}(X)$ has not been studied much. Equality (13.5) and the remark after it shows that the equality $N^{1,p}(X) = O^{1,p}(X)$ does not hold in general.

The Sobolev space $N^{1,p}(X)$ may reduce to $L^p(X)$ if X is poor in rectifiable curves (Proposition 10.9). Similarly, in the present case it may happen that

$$(13.6) \quad \Omega(X) \equiv 0.$$

In particular, $df \equiv 0$ for all $f \in \text{Lip}^\infty(X)$ in this case, and the corresponding Sobolev space $O^{1,p}(X)$ reduces to $L^p(X)$.

Let us say that X is *differentiably trivial* if (13.6) holds. By the Hahn-Banach theorem (for modules) we have that X is differentiably trivial if and only if $\Upsilon(X) \equiv 0$.

Weaver has computed some examples suggesting that the triviality of $\Upsilon(X)$ seems to be related to the lack of curves much as in the case of $N^{1,p}(X)$. For example, if X is the ternary Cantor set or if X is a snowflake space as in (9.2), then $\Upsilon(X) \equiv 0$. The first example can be seen by using the stipulated weak* continuity and density of simple functions in $\text{Lip}^\infty(X)$ for uniformly disconnected spaces; the second example follows essentially from the fact that

$$|df|(x) \leq \text{Lip}f(x)$$

for almost every $x \in X$, which easily follows from Theorem 13.1 and the locality Proposition 13.3. (Recall the definition for the Lipschitz number $\text{Lip}f(x)$ from (10.3).)

More interestingly, we have that some of the standard rectifiably connected fractals, such as the Sierpinski carpet, are differentiably trivial. One also knows that $N^{1,p}(X) = L^p(X)$ for the Sierpinski carpet. (This follows from Proposition 10.9 and the arguments in [160, pp. 29–33]; see also [102].)

Finally, it can be shown that if X is complete and connected and if $\|df\| = \text{Lip}(f)$ for all $f \in \text{Lip}^\infty(X)$, then X is a geodesic metric space.

In conclusion, out of abstract principles we have arrived at concepts of a measurable (co-)tangent bundle and a first order Sobolev space. As earlier, these concepts canonically arise from the underlying metric measure space structure. The non-triviality of these concepts carries definite information about the underlying metric space, akin to the nontriviality of $N^{1,p}(X)$. Moreover, the following fact is known.

Theorem 13.5. *If X is a doubling p -Poincaré space, then $\Omega(X)$ is isomorphic to the Cheeger cotangent bundle. In particular,*

$$(13.7) \quad N^{1,p}(X) = O^{1,p}(X)$$

isomorphically as Banach spaces.

Theorem 13.5 follows because the exterior differential is uniquely determined by certain characteristic properties satisfied by the Cheeger differential as in Theorem 12.3. In fact, for the same reason, the conclusion of Theorem 13.5 remains valid if X satisfies the weaker hypotheses as discussed in subsection 12.3.

As mentioned earlier, equality (13.7) does not hold in general. In the definition for $\Omega(X)$ only the measure class structure of a metric measure space was used, while the modulus of curve families is more sensitive to the underlying fixed measure. Consequently, the space $O^{1,p}(X)$ depends more on the infinitesimal structure of the metric measure space than that of the space $N^{1,p}(X)$. (This suggests a study of weak tangent spaces in this connection.) For example, if X is a totally disconnected measurable set in \mathbb{R}^n of positive Lebesgue measure, then $N^{1,p}(X)$ reduces to $L^p(X)$, while X is not differentially trivial by (13.3). In fact, we have in this case that df is the *approximate differential* which exists almost everywhere on X for every Lipschitz function $f : X \rightarrow \mathbb{R}$.

It would be interesting to further clarify the relationship between the Sobolev spaces $N^{1,p}(X)$ and $O^{1,p}(X)$ in general spaces.

13.3. Notes. Most of the material in this section is from Weaver's paper [188]. See also [187]. Weaver introduced the Banach modules of measurable vector fields and forms, and virtually all results and facts mentioned here are due to him (Lemma 13.4 has a more high-brow proof in [188, p. 83 ff.]). The fact that the Cheeger differential agrees with the differential constructed in [188] was jointly proved by Cheeger and Weaver [42, Remark 4.66, p. 463], [188, pp. 94–95]. Theorem 13.5 follows from this fact.

The Sobolev spaces $O^{1,p}(X)$ seem not to have been considered before in the literature.

14. OTHER SOBOLEV SPACES

A variety of definitions for Sobolev spaces on metric measure spaces have been suggested in recent years. Some of these agree with the space $N^{1,p}(X)$ introduced earlier in this article, but others are different, yet a new one was introduced in the previous section. Recall also that with Cheeger's definition, we have $N^{1,p}(X) = C^{1,p}(X)$ for $p > 1$, but not in general if $p = 1$; see subsection 10.6.

Here we review an important predecessor for the space $N^{1,p}(X)$.

14.1. Hajłasz-Sobolev space. The *Hajłasz-Sobolev* space $M^{1,p}(X)$, $1 \leq p < \infty$, on a metric measure space $X = (X, d, \mu)$ is defined as the vector space of all $L^p(X)$ -functions $u : X \rightarrow \mathbb{R}$ for which there exists an $L^p(X)$ -function $g : X \rightarrow [0, \infty)$ such that

$$(14.1) \quad |u(x) - u(y)| \leq d(x, y) (g(x) + g(y))$$

for x, y outside an exceptional set of measure zero in X . The norm in $M^{1,p}(X)$ is given by

$$\|u\|_{1,p} := \|u\|_p + \inf \|g\|_p,$$

where the infimum is taken over all g that satisfy the preceding requirement. Then $M^{1,p}(X)$ is a Banach space. Moreover, for $1 < p < \infty$ there exists, for each given $u \in M^{1,p}(X)$, an essentially smallest function g_u that satisfies the requirement in (14.1).

In many situations, the Hajłasz-Sobolev space is isomorphic (as a Banach space) to the standard Sobolev space. For example, we have that $M^{1,p}(\mathbb{R}^n) = W^{1,p}(\mathbb{R}^n)$ for $1 < p < \infty$. One can also show that $M^{1,p}(X)$ and $N^{1,p}(X)$ are isomorphic as Banach spaces if $1 < p < \infty$ and if X is a doubling p -Poincaré space. (The proof for this requires the self-improving Theorem 11.3.) When $p = 1$, the Hajłasz definition (14.1) gives a new function space; we have a strict inclusion

$$M^{1,1}(\mathbb{R}^n) \subsetneq W^{1,1}(\mathbb{R}^n) = N^{1,1}(\mathbb{R}^n).$$

The Hajłasz-Sobolev space $M^{1,p}(X)$ hardly ever degenerates into $L^p(X)$. Therefore it is possible to study Sobolev functions on disconnected fractals, for example. In fact, interesting Banach space structures emerge in such situations. One can show that for many self similar Cantor type sets, equipped with the Hausdorff measure, the Hajłasz-Sobolev space $M^{1,p}(X)$, $1 < p < \infty$, contains an isomorphic copy of the Banach space ℓ^∞ . In particular, $M^{1,p}(X)$ is not reflexive in these cases.

The main difference between the definition (14.1) and the definitions based on upper gradients or measurable derivations is that the “derivative” g_u cannot be localized (compare Lemmas 10.3 and 13.4). Indeed, essentially the optimal function g_u to use in (14.1) is the maximal function of $|\nabla u|$, if $X = \mathbb{R}^n$ and $p > 1$. In this sense, the Hajłasz-Sobolev space cannot be used for calculus as defined in the beginning of this article. On the other hand, even in classical settings, it is useful to have the simple metric characterization (14.1) for Sobolev functions.

The Hajłasz-Sobolev spaces were introduced by Hajłasz in [77]. These spaces precede the theory of Sobolev spaces $C^{1,p}(X)$ or $N^{1,p}(X)$, and they have been extensively studied during the past decade. For applications of Hajłasz-Sobolev spaces to classical problems, see for example [80]. The nonreflexivity of general Hajłasz-Sobolev spaces, mentioned in the text, was proved by Rissanen [149].

For an excellent survey of the Hajłasz-Sobolev spaces, as well as other related function spaces, see [78]. For a full technical treatment and references, see [79]. Equivalent definitions for various Sobolev spaces appear also in [103].

One can also develop a theory of Sobolev mappings with values in a Banach space. Ambrosio [7] was probably the first person to systematically study such mappings (in the context of mappings of bounded variation). For Riemannian domains, Sobolev mappings were studied in [112], [147]. For the general theory, see [88]. These studies have applications to harmonic mappings with singular targets; see for example [112], [61], [119], [141], [140].

15. POTENTIAL THEORY ON SINGULAR SPACES

Cheeger’s theory, as explained in Section 12, gives the possibility to develop a theory of elliptic partial differential equations on singular spaces satisfying a Poincaré inequality. In this section, we briefly discuss such developments.

Let $X = (X, d, \mu)$ be a doubling p -Poincaré space. Then we can choose a measurable bundle of inner products $\langle \cdot, \cdot \rangle_x$ on X so that

$$(15.1) \quad \langle du(x), du(x) \rangle_x \approx |du(x)|_x^2 = \text{Lip } u(x)^2$$

for almost every $x \in X$ and for every Lipschitz function $u : X \rightarrow \mathbb{R}$, where d is the Cheeger differential and $|\cdot|_x$ is the Banach norm given in Theorem 12.3. The constants of comparability in (15.1) depend only on the data associated with the Poincaré inequality and the doubling measure.

The preceding understood, a function $u \in N_{loc}^{1,2}(\Omega)$, where $\Omega \subset X$ is open, can be declared *harmonic* if

$$(15.2) \quad \int_{\Omega} \langle d\varphi(x), du(x) \rangle_x d\mu(x) = 0$$

for each compactly supported Lipschitz function φ in Ω .

More generally, one can consider nonlinear second order equations such as the p -Laplace equation on X . Recall that the p -Laplace equation, $1 < p < \infty$, in the weak formulation and in our general context, reads

$$(15.3) \quad \int_{\Omega} \langle d\varphi(x), |du(x)|^{p-2} du \rangle_x d\mu(x) = 0$$

for $u \in N_{loc}^{1,p}(\Omega)$, for each compactly supported Lipschitz function φ in Ω .

Armed with the assumptions that the underlying measure is doubling and that an appropriate Poincaré inequality holds, one can develop much of the classical potential theory of \mathbb{R}^n in the setting of metric measure spaces. The said assumptions suffice to run the usual machinery of iteration of various Sobolev type inequalities in order to conclude (Hölder) regularity of the solutions. Broadly speaking, the standard techniques go over as such, but there are some surprises in the details.

In equations (15.2) and (15.3), a choice has been made for the bundle of inner products. Therefore, the associated (p -)harmonic functions depend on this choice, and we do not have just one potential theory but many such theories. (Naturally, different choices yield very similar theories.)

There is a more natural potential theory that is canonically attached to each doubling p -Poincaré space. In this theory, one considers (local) minimizers of the p -energy functional

$$(15.4) \quad \int |du(x)|_x^p d\mu(x),$$

where the fiberwise Banach norm $|\cdot|_x$ on X is (almost everywhere) canonically determined by the underlying metric measure space structure, as explained in Theorem 12.3. Equivalently, one considers local minimizers of

$$(15.5) \quad \int \rho_u^p d\mu$$

as in (12.5). In the Euclidean case, when $X = \mathbb{R}^n$ with the standard metric and Lebesgue measure, the two approaches are equivalent, the p -Laplace equation being the Euler equation for the functional (15.4). In general, there need not be an

associated Euler equation for the minimizers of (15.4). This happens already when $X = \mathbb{R}^n$ equipped with a nondifferentiable Banach norm. More generally, one can consider Finsler manifolds. This fact brings interesting aspects to the theory, for one has to work with the minimizing property of solutions only.

Under the uniform bounds as in (15.1), the minimizers of (15.5) are *quasiminimizers* of the energy associated with the chosen inner product bundle. The theory of quasiminimizers of the usual p -energy in \mathbb{R}^n goes back to the fundamental ideas of De Giorgi from the 1950s and has been particularly important in this general context. In this vein, Hölder continuity and a Harnack inequality for minimizers can be established in doubling Poincaré spaces.

It was already known to Cheeger that the basic elliptic theory applies to solutions of (12.5); see [42, Remark 4.60, p. 462]. The usefulness of the quasiminimizer approach to regularity questions in this general context was first realized by Kinnunen and Shanmugalingam [109].

There is a growing literature on potential theory in doubling p -Poincaré spaces. References include [164], [165], [107], [108], [24], [25], [117], [110]. Note that these studies contain as special cases potential theories on Carnot-Carathéodory spaces or on Euclidean spaces with weights (on which a large older literature exists).

16. APPLICATIONS

The theory of calculus on metric measure spaces has shown that the scope for much of the classical first order differential analysis is broader than anticipated. On the other hand, there have been direct applications to previously existing mathematical problems as well. In this last section, I review some such successes of the theory.

16.1. Quasiconformal mappings. Upper gradients and Poincaré inequalities, as defined earlier in this article, were first considered in connection with quasiconformal mappings between metric spaces. Quasiconformality is, at the heart of it, a metric property of a homeomorphism. It is customary to call a homeomorphism $f : X \rightarrow Y$ between metric spaces *quasiconformal* if there exists a constant $H \geq 1$ such that

$$(16.1) \quad \limsup_{r \rightarrow 0} \frac{\sup\{d_Y(f(x), f(y)) : d_X(x, y) \leq r\}}{\inf\{d_Y(f(x), f(y)) : d_X(x, y) \geq r\}} \leq H < \infty$$

for every $x \in X$. In general, this definition is too weak to give an interesting class of homeomorphisms; for example, every diffeomorphism $f : \mathbb{R} \rightarrow \mathbb{R}$ would then be quasiconformal. It is however a remarkable fact that in \mathbb{R}^n , $n \geq 2$, definition (16.1) contains plenty of nontrivial information and yields a rich class of homeomorphisms that we have learned to call quasiconformal. In particular, every homeomorphism $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $n \geq 2$, that satisfies (16.1) also satisfies the following stronger condition: there exists $K \geq 1$ such that

$$(16.2) \quad \frac{\sup\{d_Y(f(x), f(y)) : d_X(x, y) \leq r\}}{\inf\{d_Y(f(x), f(y)) : d_X(x, y) \geq r\}} \leq K$$

for every $x \in \mathbb{R}^n$ and $r > 0$. (Here we understand that $X = Y = \mathbb{R}^n$ with the usual metric.) Homeomorphisms $f : X \rightarrow Y$ between metric spaces that satisfy (16.2) are called *quasisymmetric*.

A natural question to ask is: on which situations does the infinitesimal condition (16.1) imply the global condition (16.2)? The following answer was given by Koskela and the author.

Theorem 16.1. *Let X and Y be Ahlfors Q -regular spaces that support a Q -Poincaré inequality for some $Q > 1$. If $f : X \rightarrow Y$ is a homeomorphism that satisfies (16.1), then f also satisfies (16.2).*

Moreover, one can show that the statement in Theorem 16.1 is quantitative; under appropriate normalization of the mapping, the constant K in (16.2) depends only on the constant H in (16.1) and on the data associated with the spaces X and Y .

The idea behind Theorem 16.1 is that the validity of a Q -Poincaré inequality in an Ahlfors Q -regular space can be shown to be equivalent to another condition, called a *Loewner condition*, which is expressed in terms of moduli of curve families. More precisely, a pathwise connected Ahlfors Q -regular space X , $Q > 1$, is called a *Loewner space* if

$$(16.3) \quad \varphi(t) := \inf \text{mod}_Q(E, F) > 0$$

for every $t > 0$, where the infimum is taken over all pairs of nondegenerate continua $E, F \subset X$ such that

$$\text{dist}(E, F) \leq t \min\{\text{diam}(E), \text{diam}(F)\},$$

and where $\text{mod}_Q(E, F)$ stands for the Q -modulus of the curve family consisting of all curves joining E and F in X . (Recall the definition for modulus from subsections 7.2, 10.1, and recall that in an Ahlfors regular space the measure is understood to be the Hausdorff measure; see (11.2) for the definition.)

In other words, a space is a Loewner space if it contains lots of curves, as measured by modulus, at all locations and scales and in all directions. The following fact links the Loewner condition and Poincaré inequality: *an Ahlfors Q -regular space for $Q > 1$ is a Loewner space if and only if it supports a Q -Poincaré inequality.*

The study of quasiconformal mappings in abstract settings has applications to geometric rigidity theories. Such applications go back to Mostow, and we will discuss them in more detail in the next subsection. There are also other, perhaps unexpected, applications of these ideas. Namely, in the search of new methods to study quasiconformal mappings, Koskela and the author observed that for homeomorphisms $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $n \geq 2$, one can replace \limsup by \liminf in definition (16.1) and still obtain (16.2). This fact has successfully been applied in holomorphic dynamics of one complex variable.

The work on quasiconformal mappings in Loewner spaces and the associated Poincaré inequalities began in [84], [85], [86]. In particular, the aforementioned results can be found in these papers. There is a growing literature on these topics. Tyson's work [182], [183] in particular has been important for rigidity studies. The Loewner condition has also been used in the study of Gromov hyperbolic spaces; see [27], [18], [92]. For further refinements on definitions for quasiconformal mappings in Euclidean spaces and beyond see [96], [17], [19]. For the applications to dynamics, see [127], [76].

There are also applications of the concepts of an upper gradient and Banach-valued Sobolev mappings to the theory of boundary behavior of quasiconformal mappings in half spaces [15].

16.2. Geometric rigidity. It has already been mentioned that one of the first applications of nonsmooth calculus was the proof of the Mostow rigidity theorem for lattices in rank one symmetric spaces. The boundaries at infinity of these spaces have (locally) a description as a Carnot group, and a detailed analysis of quasiconformal homeomorphisms were critical to Mostow's theory.

Mostow's ideas have been expanded in later years. In particular, in proving further rigidity results, Pansu also furthered quasiconformal analysis on general Carnot groups.

Recently, Bourdon and Pajot established a rigidity property of certain negatively curved spaces, called *Fuchsian buildings*. Such buildings themselves are singular spaces, where hyperbolic planes are glued together along boundaries of fixed hyperbolic polygons. The rigidity result of Bourdon and Pajot states that *every quasiisometry of a Fuchsian building is a finite distance from an isometry*. Thus, the rigidity found in Fuchsian buildings is similar to that found by Pansu in the quaternionic and Cayley hyperbolic spaces. The Bourdon-Pajot theorem is proved by a Mostow type analysis using quasiconformal mappings. The given quasiisometry between the buildings descends to a quasiconformal (quasisymmetric) mapping between the boundaries; these boundaries are shown to be certain exotic Ahlfors regular spaces of fractional dimension, with a Poincaré inequality, and the task is to show that the induced quasiconformal mapping is in fact “conformal”, hence the restriction of an isometry. The abstract analytic theory of quasiconformal mappings and calculus on singular spaces are invoked in the proof.

Other remarkable rigidity results were proved recently by Bonk and Kleiner. In their work, results and techniques of nonsmooth calculus as presented here are decisively used. To mention one concrete example, we recall the well known *Cannon's conjecture* which predicts that every negatively curved word (or Gromov) hyperbolic group with boundary homeomorphic to the two dimensional sphere admits a discrete and cocompact action on hyperbolic three space. An affirmative answer to this conjecture is known to have deep consequences for three manifold theory; in particular, it implies the *hyperbolization conjecture* asserting that every negatively curved closed three manifold admits a metric of constant curvature.²⁹ By using analysis based on the abstract Poincaré inequality (among several other things), Bonk and Kleiner have reduced Cannon's conjecture to the problem of finding a minimal Hausdorff dimension among all spheres that can occur as boundaries of such groups.

Mostow's fundamental work is [137]. For Pansu's work, see [143], [142]. An excellent survey on these matters is [74]. Works of Bourdon and Pajot related to the subject here include [34], [35], [36]. The cited work of Bonk and Kleiner is [31]. See also [26] and [111] for forthcoming surveys.

16.3. Manifolds with Ricci bounded below. Gromov-Hausdorff limits of Riemannian manifolds with uniform lower bounds on their Ricci curvature tensor are examples of doubling Poincaré spaces. This was alluded to earlier in subsection 8.8. In a series of remarkable papers, Cheeger and Colding have studied Riemannian manifolds with Ricci curvature bounded below by exploiting this fact (among many

²⁹Although Perelman's recent work may yield a positive solution to the hyperbolization conjecture, Cannon's conjecture remains open as of this writing.

other things). They have proved results that describe the structure of such manifolds at a small but definite scale. Some of the theorems of Cheeger and Colding are proved via a limit space analysis, although the hypotheses as well as the conclusions in these theorems are expressed purely in smooth, Riemannian terms. I refer to [45], [46], [47], [43] for a more detailed discussion of these results. See also [44, p. 64] and [48, p. 910] for specific instances, where the results of [42] were used.³⁰

In subsection 8.4, we discussed a synthetic treatment of sectional curvature bounds, going back to Alexandrov in the 1950s. Similar questions for spaces with Ricci curvature bounds have attracted much attention recently. For an informative discussion of this problem, see [45, Appendix 2]. Lott and Villani [128], and Sturm [173], [174] have independently proposed an approach for defining metric measure spaces with Ricci curvature bounded from below. Lott and Villani also prove a logarithmic Sobolev inequality for metric spaces with Ricci curvature bounded from below. A global Poincaré type inequality then follows for spaces with positive lower curvature bounds. As far as I know, it is an open question whether the spaces considered by Lott-Villani and Sturm support a Poincaré inequality as defined in subsection 11.1.³¹

16.4. Nonembedding results. An interesting question, also from the point of view of applications, asks when a given metric space admit a bi-Lipschitz embedding in some finite dimensional Euclidean space. For example, every subset of \mathbb{R}^n with the induced metric is clearly doubling as defined in subsection 8.7, but there are doubling metric spaces that do not embed in any \mathbb{R}^n , or even in any Hilbert space. Every known example to this effect is somewhat nontrivial, and prior to Cheeger's work [42], each such nonembeddability result was verified by some *ad hoc* method. For example, it is known that none of the Carnot-Carathéodory spaces (subsection 8.5) admit a bi-Lipschitz embedding in a Hilbert space. The Laakso spaces mentioned in subsection 8.11 also do not admit a bi-Lipschitz embedding in any Hilbert space.³²

Cheeger applied his differentiation theory (as partially explained in Section 12) to provide a uniform statement that covers many of the known nonembedding results of this type. More precisely, he proved that *if a doubling p -Poincaré space X admits a bi-Lipschitz embedding in some finite dimensional Euclidean space, then for almost every point $x \in X$, every tangent space of X at x is bi-Lipschitz equivalent to some \mathbb{R}^N* . The more precise result is that if $\{U_\alpha\}$ is a measurable decomposition of X as in Theorem 12.2, then for almost every $x \in U_\alpha$, every tangent space of X at x is bi-Lipschitz equivalent to $\mathbb{R}^{N(\alpha)}$. This is a strong structure theorem for doubling p -Poincaré spaces that are subsets of finite dimensional Euclidean spaces. One can easily deduce from it the previously known nonembeddability results both for the Carnot-Carathéodory spaces and the Laakso spaces. It also follows from it that the boundaries of the Fuchsian buildings mentioned in subsection 16.2 do not admit bi-Lipschitz embeddings in any Euclidean space.

³⁰I am grateful to Jeff Cheeger for these references.

³¹Added in October 2005: Lott and Villani [129], and also Renesse [186], have meanwhile addressed this issue.

³²There are even stronger statements known regarding nonembeddability of these spaces in nice Banach spaces, e.g. in uniformly convex spaces.

One can also view this structure theorem as an indication that the truly exotic geometries that allow for calculus cannot be found inside Euclidean spaces, where by “truly exotic” we mean spaces whose infinitesimal or tangent space structure is singular as defined in subsection 8.3.

Incidentally, the first proof of the nonembeddability of the Carnot-Carathéodory spaces was also based on a differentiability result, namely that of Pansu [143]. Pansu’s differentiability theorem takes into account the algebraic structure in Carnot groups, which typically appear as tangent spaces in this case [135]; when mapping into a commutative group, such as \mathbb{R}^N , the differential always has non-trivial kernel, which is impossible for bi-Lipschitz embeddings. This deduction of nonembeddability of Carnot groups in Euclidean spaces from Pansu’s result is due to Semmes [157, Section 7] (who also acknowledges that the fact was independently known to Assouad).

Cheeger’s aforementioned theorem (which can be found in [42, Section 14]) was in turn applied in a totally different context by Bonk and Kleiner [31]. See also [125] and the references there for recent results on embeddability of metric spaces.

16.5. Metric parametrization of sets. It was mentioned in subsection 8.3 that it is a difficult problem to decide whether a given metric space is locally bi-Lipschitz equivalent to an open subset of some finite dimensional Euclidean space. This *bi-Lipschitz parametrization problem* has attracted much attention in recent years with contributions by many people. To illustrate the difficulty of the problem, let us mention that there are, for each integer n at least five, finite n -dimensional polyhedra that are topological manifolds but not locally bi-Lipschitz equivalent to a ball in \mathbb{R}^n (with respect to a natural intrinsic metric in the polyhedron). This was first observed by Siebenmann and Sullivan [166] as a consequence of the Edwards double suspension theorem [60]. Remarkable positive parametrization results were achieved by Toro [178], [179] in the mid 1990s, and by now there are several nontrivial results known, both positive and negative. The general problem is nevertheless far from being understood.

The nonsmooth calculus as discussed in this article has played a particularly important role in recent works [90], [89] on the bi-Lipschitz parametrization problem. Keith and the author (work in progress) have recently obtained further sufficient conditions for such parametrizations to exist; inspired by Sullivan’s work [176], we have moreover applied these methods to establish new sufficient conditions for smoothability of topological manifolds.

Finally, instead of bi-Lipschitz parametrizations, one can ask for weaker parametrizations e.g. by quasiconformal or quasisymmetric homeomorphisms. Bonk and Kleiner [28] have proved a two dimensional quasisymmetric uniformization result by using (among many things) ideas connected with the Loewner condition and Poincaré inequalities.

For a more detailed discussion of parametrization problems and applications, see [154], [158], [161], [82], and the references there.

ADDED AFTER POSTING: ABOUT THE AUTHOR

Juha Heinonen is a professor of mathematics at the University of Michigan. He was an invited speaker at the ICM 2002 in Beijing and is a member of the Finnish Academy of Science and Letters.

REFERENCES

1. *Geometry. IV*, Encyclopaedia of Mathematical Sciences, vol. 70, Springer-Verlag, Berlin, 1993. Nonregular Riemannian geometry, A translation of *Geometry, 4* (Russian), Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1989. Translation by E. Primrose. MR1263963 (94i:53038)
2. *Problems in low-dimensional topology*, Geometric Topology (Athens, GA, 1993) (Rob Kirby, ed.), AMS/IP Stud. Adv. Math., vol. 2, Amer. Math. Soc., Providence, RI, 1997, pp. 35–473. MR1470751
3. David R. Adams and Lars Inge Hedberg, *Function spaces and potential theory*, Grundlehren der Mathematischen Wissenschaften, vol. 314, Springer-Verlag, Berlin, 1996. MR1411441 (97j:46024)
4. Robert A. Adams, *Sobolev spaces*, Academic Press, New York-London, 1975, Pure and Applied Mathematics, Vol. 65. MR0450957 (56:9247)
5. L. Ahlfors and A. Beurling, *Conformal invariants and function-theoretic null-sets*, Acta Math. **83** (1950), 101–129. MR0036841 (12:171c)
6. A. D. Aleksandrov and V. A. Zalgaller, *Intrinsic geometry of surfaces*, Translated from the Russian by J. M. Danskin. Translations of Mathematical Monographs, Vol. 15, American Mathematical Society, Providence, RI, 1967. MR0216434 (35:7267)
7. Luigi Ambrosio, *Metric space valued functions of bounded variation*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **17** (1990), no. 3, 439–478. MR1079985 (92d:26022)
8. ———, *Some fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces*, Adv. Math. **159** (2001), no. 1, 51–67. MR1823840 (2002b:31002)
9. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005. MR2129498 (2006k:49001)
10. Luigi Ambrosio and Bernd Kirchheim, *Currents in metric spaces*, Acta Math. **185** (2000), no. 1, 1–80. MR1794185 (2001k:49095)
11. L. Ambrosio, M. Miranda, Jr., and D. Pallara, *Special functions of bounded variation in doubling metric measure spaces*, Calculus of Variations: Topics from the Mathematical Heritage of E. De Giorgi, Quad. Mat., vol. 14, Dept. Math., Seconda Univ. Napoli, Caserta, 2004, pp. 1–45. MR2118414 (2005j:49036)
12. Luigi Ambrosio and Francesco Serra Cassano (eds.), *Lectures notes on analysis in metric spaces*, Appunti dei Corsi Tenuti da Docenti della Scuola. [Notes of Courses Given by Teachers at the School], Scuola Normale Superiore, Pisa, 2000. Papers from the International Summer School held in Trento, May 1999. MR2023120 (2004h:00022)
13. Luigi Ambrosio and Paolo Tilli, *Topics on analysis in metric spaces*, Oxford Lecture Series in Mathematics and Its Applications, vol. 25, Oxford University Press, Oxford, 2004. MR2039660 (2004k:28001)
14. P. Assouad, *Plongements lipschitziens dans \mathbf{R}^n* , Bull. Soc. Math. France **111** (1983), 429–448. MR0763553 (86f:54050)
15. Kari Astala, Mario Bonk, and Juha Heinonen, *Quasiconformal mappings with Sobolev boundary values*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **1** (2002), no. 3, 687–731. MR1990676 (2004k:30045)
16. Pascal Auscher, Thierry Coulhon, and Alexander Grigor’yan (eds.), *Heat kernels and analysis on manifolds, graphs, and metric spaces*, Contemporary Mathematics, vol. 338, American Mathematical Society, Providence, RI, 2003. Lecture notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs held in Paris, April 16–July 13, 2002. MR2041910 (2004k:00018)
17. Z. Balogh and P. Koskela, *Quasiconformality, quasisymmetry and removability in Loewner spaces*, Duke Math. J. **101** (2000), no. 3, 554–577. With an appendix by J. Väisälä. MR1740689 (2001d:30029)
18. Z. M. Balogh and S. M. Buckley, *Geometric characterizations of Gromov hyperbolicity*, Invent. Math. **153** (2003), no. 2, 261–301. MR1992014 (2004i:30042)
19. Z. M. Balogh, P. Koskela, and S. Rogovin, *Absolute continuity of quasiconformal mappings on curves*, Geometric and Functional Analysis (to appear).
20. Z. M. Balogh, K. Rogovin, and T. Zürcher, *The Stepanov differentiability theorem in metric measure spaces*, J. Geom. Anal. **14** (2004), no. 3, 405–422. MR2077159 (2005d:28008)

21. D. Bao, S.-S. Chern, and Z. Shen, *An introduction to Riemann-Finsler geometry*, Graduate Texts in Mathematics, vol. 200, Springer-Verlag, New York, 2000. MR1747675 (2001g:53130)
22. A. Bellaïche, *The tangent space in sub-Riemannian geometry*, Sub-Riemannian Geometry, Birkhäuser, Basel, 1996, pp. 1–78. MR1421822 (98a:53108)
23. Arne Beurling, *The collected works of Arne Beurling. Vol. 1*, Contemporary Mathematicians, Birkhäuser Boston Inc., Boston, MA, 1989. Complex Analysis, edited by L. Carleson, P. Malliavin, J. Neuberger and J. Wermer. MR1057613 (92k:01046a)
24. Anders Björn, Jana Björn, and Nageswari Shanmugalingam, *The Dirichlet problem for p -harmonic functions on metric spaces*, J. Reine Angew. Math. **556** (2003), 173–203. MR1971145 (2004e:31007)
25. ———, *The Perron method for p -harmonic functions in metric spaces*, J. Differential Equations **195** (2003), no. 2, 398–429. MR2016818 (2004i:31012)
26. M. Bonk, *Quasiconformal geometry of fractals*, ICM Proceedings, Madrid (2006).
27. M. Bonk, J. Heinonen, and P. Koskela, *Uniformizing Gromov hyperbolic spaces*, Astérisque **270** (2001). MR1829896 (2003b:30024)
28. Mario Bonk and Bruce Kleiner, *Quasisymmetric parametrizations of two-dimensional metric spheres*, Invent. Math. **150** (2002), no. 1, 127–183. MR1930885 (2004k:53057)
29. ———, *Rigidity for quasi-Möbius group actions*, J. Differential Geom. **61** (2002), no. 1, 81–106. MR1949785 (2004b:53059)
30. ———, *Rigidity for quasi-Fuchsian actions on negatively curved spaces*, Int. Math. Res. Not. (2004), no. 61, 3309–3316. MR2096259 (2006b:53051)
31. ———, *Conformal dimension and Gromov hyperbolic groups with 2-sphere boundary*, Geom. Topol. **9** (2005), 219–246 (electronic). MR2116315 (2005k:20102)
32. Mario Bonk and Urs Lang, *Bi-Lipschitz parameterization of surfaces*, Math. Ann. **327** (2003), no. 1, 135–169. MR2006006
33. Nicolas Bourbaki, *Elements of the history of mathematics*, Springer-Verlag, Berlin, 1994, Translated from the 1984 French original by John Meldrum. MR1290116 (95c:01001)
34. M. Bourdon, *Immeubles hyperboliques, dimension conforme et rigidité de Mostow*, Geom. Funct. Anal. **7** (1997), 245–268. MR1445387 (98c:20056)
35. M. Bourdon and H. Pajot, *Poincaré inequalities and quasiconformal structure on the boundaries of some hyperbolic buildings*, Proc. Amer. Math. Soc. **127** (1999), no. 8, 2315–2324. MR1610912 (99j:30024)
36. ———, *Cohomologie l_p et espaces de Besov*, J. Reine Angew. Math. **558** (2003), 85–108. MR1979183 (2004e:20073)
37. Martin R. Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999. MR1744486 (2000k:53038)
38. A. Bruckner, J. Bruckner, and B. Thomson, *Real analysis*, Prentice-Hall, NJ, 1997.
39. J. Bryant, S. Ferry, W. Mio, and S. Weinberger, *Topology of homology manifolds*, Ann. of Math. (2) **143** (1996), no. 3, 435–467. MR1394965 (97b:57017)
40. Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001. MR1835418 (2002e:53053)
41. P. Buser, *A note on the isoperimetric constant*, Ann. Sci. École Norm. Sup. (4) **15** (1982), 213–230. MR0683635 (84e:58076)
42. J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal. **9** (1999), 428–517. MR1708448 (2000g:53043)
43. ———, *Degeneration of Riemannian metrics under Ricci curvature bounds*, Scuola Normale Superiore, 2001. Lezioni Fermiane, Pisa, 2001. MR006642 (2004j:53049)
44. ———, *Integral bounds on curvature elliptic estimates and rectifiability of singular sets*, Geom. Funct. Anal. **13** (2003), no. 1, 20–72. MR1978491 (2004i:53041)
45. Jeff Cheeger and Tobias H. Colding, *On the structure of spaces with Ricci curvature bounded below. I*, J. Differential Geom. **46** (1997), no. 3, 406–480. MR1484888 (98k:53044)
46. ———, *On the structure of spaces with Ricci curvature bounded below. II*, J. Differential Geom. **54** (2000), no. 1, 13–35. MR1815410 (2003a:53043)
47. ———, *On the structure of spaces with Ricci curvature bounded below. III*, J. Differential Geom. **54** (2000), no. 1, 37–74. MR1815411 (2003a:53044)

48. J. Cheeger, T. H. Colding, and G. Tian, *On the singularities of spaces with bounded Ricci curvature*, *Geom. Funct. Anal.* **12** (2002), no. 5, 873–914. MR1937830 (2003m:53053)
49. Jeff Cheeger and Detlef Gromoll, *The splitting theorem for manifolds of nonnegative Ricci curvature*, *J. Differential Geometry* **6** (1971/72), 119–128. MR0303460 (46:2597)
50. R. R. Coifman and G. Weiss, *Analyse harmonique non-commutative sur certains espaces homogènes*, *Lecture Notes in Mathematics*, no. 242, Springer-Verlag, Berlin, 1971. MR0499948 (58:17690)
51. Ronald R. Coifman and Guido Weiss, *Extensions of Hardy spaces and their use in analysis*, *Bull. Amer. Math. Soc.* **83** (1977), no. 4, 569–645. MR0447954 (56:6264)
52. Tobias H. Colding and William P. Minicozzi II, *Harmonic functions on manifolds*, *Ann. of Math. (2)* **146** (1997), no. 3, 725–747. MR1491451 (98m:53052)
53. Alain Connes, Dennis Sullivan, and Nicolas Teleman, *Quasiconformal mappings, operators on Hilbert space, and local formulae for characteristic classes*, *Topology* **33** (1994), no. 4, 663–681. MR1293305 (95g:58232)
54. Robert J. Daverman, *Decompositions of manifolds*, *Pure and Applied Mathematics*, vol. 124, Academic Press Inc., Orlando, FL, 1986. MR872468 (88a:57001)
55. G. David and S. Semmes, *Strong A_∞ weights, Sobolev inequalities and quasiconformal mappings*, *Analysis and Partial Differential Equations, Lecture Notes in Pure and Appl. Math.*, vol. 122, Marcel Dekker, 1990, pp. 101–111. MR1044784 (91c:30037)
56. ———, *Fractured fractals and broken dreams: self-similar geometry through metric and measure*, *Oxford Lecture Series in Mathematics and its Applications*, vol. 7, Clarendon Press, Oxford University Press, 1997. MR1616732 (99h:28018)
57. S. K. Donaldson and D. P. Sullivan, *Quasiconformal 4-manifolds*, *Acta Math.* **163** (1989), 181–252. MR1032074 (91d:57012)
58. J. L. Doob, *Classical potential theory and its probabilistic counterpart*, *Grundlehren der Mathematischen Wissenschaften*, vol. 262, Springer-Verlag, New York, 1984. MR0731258 (85k:31001)
59. Nelson Dunford and Jacob T. Schwartz, *Linear operators. Part I*, *Wiley Classics Library*, John Wiley & Sons Inc., New York, 1988. General Theory, with the assistance of William G. Bade and Robert G. Bartle, reprint of the 1958 original, a Wiley-Interscience Publication. MR1009162 (90g:47001a)
60. Robert D. Edwards, *The topology of manifolds and cell-like maps*, *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, *Acad. Sci. Fennica, Helsinki*, 1980, pp. 111–127. MR0562601 (81g:57010)
61. J. Eells and B. Fuglede, *Harmonic maps between Riemannian polyhedra*, *Cambridge Tracts in Mathematics*, vol. 142, Cambridge University Press, Cambridge, 2001. With a preface by M. Gromov. MR1848068 (2002h:58017)
62. L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, *Studies in Advanced Mathematics*, CRC Press, Boca Raton, Florida, 1992. MR1158660 (93f:28001)
63. E. B. Fabes, C. E. Kenig, and R. Serapioni, *The local regularity of solutions to degenerate elliptic equations*, *Comm. PDE* **7** (1982), 77–116. MR0643158 (84i:35070)
64. H. Federer, *Geometric measure theory*, *Die Grundlehren der Mathematischen Wissenschaften*, vol. 153, Springer-Verlag, New York, 1969. MR0257325 (41:1976)
65. B. Franchi, P. Hajłasz, and P. Koskela, *Definitions of Sobolev classes on metric spaces*, *Ann. Inst. Fourier (Grenoble)* **49** (1999), no. 6, 1903–1924. MR1738070 (2001a:46033)
66. Michael Hartley Freedman, *The topology of four-dimensional manifolds*, *J. Differential Geom.* **17** (1982), no. 3, 357–453. MR0679066 (84b:57006)
67. M. H. Freedman and F. Quinn, *Topology of 4-manifolds*, *Princeton Mathematical Series*, vol. 39, Princeton University Press, Princeton, NJ, 1990. MR1201584 (94b:57021)
68. Joseph H. G. Fu, *Bi-Lipschitz rough normal coordinates for surfaces with an L^1 curvature bound*, *Indiana Univ. Math. J.* **47** (1998), no. 2, 439–453. MR1647908 (99k:49090)
69. B. Fuglede, *Extremal length and functional completion*, *Acta Math.* **98** (1957), 171–219. MR0097720 (20:4187)
70. David Gilbarg and Neil S. Trudinger, *Elliptic partial differential equations of second order*, second ed., *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 224, Springer-Verlag, Berlin, 1983. MR0737190 (86c:35035)
71. M. Gromov, *Hyperbolic groups*, *Essays in Group Theory*, S. Gersten, editor, *MSRI Publications*, Springer-Verlag, 1987, pp. 75–265. MR0919826 (88e:20004)

72. ———, *Carnot-Carathéodory spaces seen from within*, Sub-Riemannian Geometry, Progress in Mathematics, vol. 144, Birkhäuser, Basel, 1996, pp. 79–323. MR1421823 (2000f:53034)
73. ———, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999. Based on the 1981 French original, with appendices by M. Katz, P. Pansu and S. Semmes, translated from the French by Sean Michael Bates. MR1699320 (2000d:53065)
74. M. Gromov and P. Pansu, *Rigidity of lattices: an introduction*, Geometric Topology: Recent Developments (Montecatini Terme, 1990), Lecture Notes in Math., vol. 1504, Springer, Berlin, 1991, pp. 39–137. MR1168043 (93f:53036)
75. M. Gromov and W. Thurston, *Pinching constants for hyperbolic manifolds*, Invent. Math. **89** (1987), no. 1, 1–12. MR0892185 (88e:53058)
76. Peter Haïssinsky, *Rigidity and expansion for rational maps*, J. London Math. Soc. (2) **63** (2001), no. 1, 128–140. MR1802762 (2001m:37085)
77. P. Hajłasz, *Sobolev spaces on an arbitrary metric space*, Potential Anal. **5** (1996), 403–415. MR1401074 (97f:46050)
78. ———, *Sobolev spaces on metric-measure spaces*, Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002), Contemp. Math., vol. 338, Amer. Math. Soc., Providence, RI, 2003, pp. 173–218. MR2039955 (2005c:46039)
79. P. Hajłasz and P. Koskela, *Sobolev met Poincaré*, Memoirs Amer. Math. Soc. **145** (2000), no. 688. MR1683160 (2000j:46063)
80. P. Hajłasz and O. Martio, *Traces of Sobolev functions on fractal type sets and characterization of extension domains*, J. Funct. Anal. **143** (1997), 221–246. MR1428124 (98d:46034)
81. J. Heinonen, *Lectures on analysis on metric spaces*, Springer-Verlag, New York, 2001. MR1800917 (2002c:30028)
82. ———, *The branch set of a quasiregular mapping*, Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002), Higher Ed. Press, Beijing, 2002, pp. 691–700. MR1957076 (2003k:30034)
83. J. Heinonen, T. Kilpeläinen, and O. Martio, *Nonlinear potential theory of degenerate elliptic equations*, The Clarendon Press, Oxford University Press, New York, 1993, Oxford Science Publications. MR1207810 (94e:31003)
84. J. Heinonen and P. Koskela, *Definitions of quasiconformality*, Invent. Math. **120** (1995), 61–79. MR1323982 (96e:30051)
85. ———, *From local to global in quasiconformal structures*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), 554–556. MR1372507 (96m:30034)
86. ———, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math. **181** (1998), 1–61. MR1654771 (99j:30025)
87. ———, *A note on Lipschitz functions, upper gradients, and the Poincaré inequality*, New Zealand J. Math. **28** (1999), 37–42. MR1691958 (2000d:46041)
88. J. Heinonen, P. Koskela, N. Shanmugalingam, and J. T. Tyson, *Sobolev classes of Banach space-valued functions and quasiconformal mappings*, J. Anal. Math. **85** (2001), 87–139. MR1869604 (2002k:46090)
89. Juha Heinonen and Seppo Rickman, *Geometric branched covers between generalized manifolds*, Duke Math. J. **113** (2002), no. 3, 465–529. MR1909607 (2003h:57003)
90. Juha Heinonen and Dennis Sullivan, *On the locally branched Euclidean metric gauge*, Duke Math. J. **114** (2002), no. 1, 15–41. MR1915034 (2004b:30044)
91. Ernst Heintze, *On homogeneous manifolds of negative curvature*, Math. Ann. **211** (1974), 23–34. MR0353210 (50:5695)
92. David A. Herron, *Conformal deformations of uniform Loewner spaces*, Math. Proc. Cambridge Philos. Soc. **136** (2004), no. 2, 325–360. MR2040578 (2005i:30029)
93. Francis Hirsch and Gilles Lacombe, *Elements of functional analysis*, Graduate Texts in Mathematics, vol. 192, Springer-Verlag, New York, 1999. Translated from the 1997 French original by Silvio Levy. MR1678925 (99j:46001)
94. Tadeusz Iwaniec, *The Gehring lemma*, Quasiconformal Mappings and Analysis (Ann Arbor, MI, 1995), Springer, New York, 1998, pp. 181–204. MR1488451 (99e:30012)
95. D. Jerison, *The Poincaré inequality for vector fields satisfying Hörmander’s condition*, Duke Math. J. **53** (1986), 503–523. MR0850547 (87i:35027)
96. S. Kallunki and P. Koskela, *Exceptional sets for the definition of quasiconformality*, Amer. J. Math. **122** (2000), no. 4, 735–743. MR1771571 (2001h:37095)

97. S. Kallunki and N. Shanmugalingam, *Modulus and continuous capacity*, Ann. Acad. Sci. Fenn. Math. **26** (2001), no. 2, 455–464. MR1833251 (2002c:31008)
98. Michael Kapovich and Bruce Kleiner, *Hyperbolic groups with low-dimensional boundary*, Ann. Sci. École Norm. Sup. (4) **33** (2000), no. 5, 647–669. MR1834498 (2002j:20077)
99. Stephen Keith, *Modulus and the Poincaré inequality on metric measure spaces*, Math. Z. **245** (2003), no. 2, 255–292. MR2013501 (2004k:31019)
100. ———, *A differentiable structure for metric measure spaces*, Adv. Math. **183** (2004), no. 2, 271–315. MR2041901 (2005g:46070)
101. ———, *Measurable differentiable structures and the Poincaré inequality*, Indiana Univ. Math. J. **53** (2004), no. 4, 1127–1150. MR2095451 (2005g:53068)
102. S. Keith and T. Laakso, *Conformal Assouad dimension and modules*, Geom. Funct. Anal. **14** (2004), no. 6, 1278–1321. MR2135168 (2006g:30027)
103. Stephen Keith and Kai Rajala, *A remark on Poincaré inequalities on metric measure spaces*, Math. Scand. **95** (2004), no. 2, 299–304. MR2098359 (2005f:26057)
104. S. Keith and X. Zhong, *The Poincaré inequality is an open ended condition*, Annals of Mathematics (to appear).
105. Jun Kigami, *Analysis on fractals*, Cambridge Tracts in Mathematics, vol. 143, Cambridge University Press, Cambridge, 2001. MR1840042 (2002c:28015)
106. Tero Kilpeläinen, *A remark on the uniqueness of quasi continuous functions*, Ann. Acad. Sci. Fenn. Math. **23** (1998), no. 1, 261–262. MR1601887 (99j:31016)
107. Juha Kinnunen and Olli Martio, *Potential theory of quasiminimizers*, Ann. Acad. Sci. Fenn. Math. **28** (2003), no. 2, 459–490. MR1996447 (2004b:31010)
108. ———, *Sobolev space properties of superharmonic functions on metric spaces*, Results Math. **44** (2003), no. 1-2, 114–129. MR2011911 (2004i:31013)
109. Juha Kinnunen and Nageswari Shanmugalingam, *Regularity of quasi-minimizers on metric spaces*, Manuscripta Math. **105** (2001), no. 3, 401–423. MR1856619 (2002i:35054)
110. ———, *Polar sets on metric spaces*, Trans. Amer. Math. Soc. **358** (2006), no. 1, 11–37. MR2171221
111. B. Kleiner, *The asymptotic geometry of negatively curved spaces: uniformization, geometrization and rigidity*, ICM Proceedings, Madrid (2006).
112. N. J. Korevaar and R. M. Schoen, *Sobolev spaces and harmonic maps for metric space targets*, Comm. Anal. Geom. **1** (1993), 561–659. MR1266480 (95b:58043)
113. P. Koskela, *Removable sets for Sobolev spaces*, Ark. Mat. **37** (1999), no. 2, 291–304. MR1714767 (2001g:46077)
114. ———, *Upper gradients and Poincaré inequalities*, Lecture Notes on Analysis in Metric Spaces (Trento, 1999), Appunti Corsi Tenuti Docenti Sc., Scuola Norm. Sup., Pisa, 2000, pp. 55–69. MR2023123 (2004i:46048)
115. ———, *Sobolev spaces and quasiconformal mappings on metric spaces*, European Congress of Mathematics, Vol. I (Barcelona, 2000), Progr. Math., vol. 201, Birkhäuser, Basel, 2001, pp. 457–467. MR1905335 (2003m:30049)
116. P. Koskela and P. MacManus, *Quasiconformal mappings and Sobolev spaces*, Studia Math. **131** (1998), 1–17. MR1628655 (99e:46042)
117. P. Koskela, Kai Rajala, and Nageswari Shanmugalingam, *Lipschitz continuity of Cheeger-harmonic functions in metric measure spaces*, J. Funct. Anal. **202** (2003), no. 1, 147–173. MR1994768 (2004e:31006)
118. P. Koskela, N. Shanmugalingam, and H. Tuominen, *Removable sets for the Poincaré inequality on metric spaces*, Indiana Univ. Math. J. **49** (2000), no. 1, 333–352. MR1777027 (2001g:46076)
119. P. Koskela, Nageswari Shanmugalingam, and Jeremy T. Tyson, *Dirichlet forms, Poincaré inequalities, and the Sobolev spaces of Korevaar and Schoen*, Potential Anal. **21** (2004), no. 3, 241–262. MR2075670 (2005f:31015)
120. Kazuhiro Kuwae, Yoshiroh Machigashira, and Takashi Shioya, *Sobolev spaces, Laplacian, and heat kernel on Alexandrov spaces*, Math. Z. **238** (2001), no. 2, 269–316. MR1865418 (2002m:58052)
121. T. Laakso, *Ahlfors Q -regular spaces with arbitrary Q admitting weak Poincaré inequalities*, Geom. Funct. Anal. **10** (2000), 111–123. MR1748917 (2001m:30027)
122. Jean-François Lafont, *Rigidity result for certain three-dimensional singular spaces and their fundamental groups*, Geom. Dedicata **109** (2004), 197–219. MR2114076 (2006g:57032)

123. N. S. Landkof, *Foundations of modern potential theory*, Springer-Verlag, New York, 1972. Translated from the Russian by A. P. Doohovskoy, Die Grundlehren der Mathematischen Wissenschaften, Band 180. MR0350027 (50:2520)
124. D. G. Larman, *A new theory of dimension*, Proc. London Math. Soc. **17** (1967), 178–192. MR0203691 (34:3540)
125. James R. Lee, Manor Mendel, and Assaf Naor, *Metric structures in L_1 : dimension, snowflakes, and average distortion*, European J. Combin. **26** (2005), no. 8, 1180–1190. MR2163751 (2006g:46012)
126. B. Levi, *Sul principio di Dirichlet*, Rend. Circ. Mat. Palermo **22** (1906), 293–359.
127. Genadi Levin and Sebastian van Strien, *Bounds for maps of an interval with one critical point of inflection type. II*, Invent. Math. **141** (2000), no. 2, 399–465. MR1775218 (2001i:37061)
128. J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Annals of Mathematics (to appear).
129. ———, *Weak curvature conditions and functional inequalities*, Journal of Functional Analysis (to appear).
130. J. Luukkainen and E. Saksman, *Every complete doubling metric space carries a doubling measure*, Proc. Amer. Math. Soc. **126** (1998), 531–534. MR1443161 (99c:28009)
131. J. Luukkainen and J. Väisälä, *Elements of Lipschitz topology*, Ann. Acad. Sci. Fenn. Ser. A I Math. **3** (1977), 85–122. MR0515647 (80b:57015)
132. Jan Malý and William P. Ziemer, *Fine regularity of solutions of elliptic partial differential equations*, Mathematical Surveys and Monographs, vol. 51, American Mathematical Society, Providence, RI, 1997. MR1461542 (98h:35080)
133. P. Mattila, *Geometry of sets and measures in Euclidean spaces*, Cambridge Studies in Advanced Mathematics, vol. 44, Cambridge University Press, Cambridge, 1995. MR1333890 (96h:28006)
134. Vladimir G. Maz'ja, *Sobolev spaces*, Springer-Verlag, Berlin, 1985. Translated from the Russian by T. O. Shaposhnikova. MR0817985 (87g:46056)
135. J. Mitchell, *On Carnot-Carathéodory metrics*, J. Differential Geom. **21** (1985), 35–45. MR0806700 (87d:53086)
136. C. B. Morrey, *Multiple integrals in the calculus of variations*, Springer-Verlag, Berlin, 1966. MR0202511 (34:2380)
137. G. D. Mostow, *Strong rigidity of locally symmetric spaces*, Princeton University Press, Princeton, NJ, 1973, Annals of Mathematics Studies, No. 78. MR0385004 (52:5874)
138. ———, *A remark on quasiconformal mappings on Carnot groups*, Michigan Math. J. **41** (1994), no. 1, 31–37. MR1260606 (95c:22017)
139. F. Nazarov, S. Treil, and A. Volberg, *The Tb-theorem on non-homogeneous spaces*, Acta Math. **190** (2003), no. 2, 151–239. MR1998349 (2005d:30053)
140. Shin-ichi Ohta, *Cheeger type Sobolev spaces for metric space targets*, Potential Anal. **20** (2004), no. 2, 149–175. MR2032946 (2005h:58017)
141. ———, *Harmonic maps and totally geodesic maps between metric spaces*, Tohoku Mathematical Publications, vol. 28, Tohoku University Mathematical Institute, Sendai, 2004. Dissertation, Tohoku University, Sendai, 2004. MR2051353 (2005e:58023)
142. P. Pansu, *Dimension conforme et sphère à l'infini des variétés à courbure négative*, Ann. Acad. Sci. Fenn. Ser. A I Math. **14** (1989), 177–212. MR1024425 (90k:53079)
143. ———, *Métriques de Carnot-Carathéodory et quasimétries des espaces symétriques de rang un*, Ann. of Math. (2) **129** (1989), 1–60. MR0979599 (90e:53058)
144. Conrad Plaut, *Metric spaces of curvature $\geq k$* , Handbook of Geometric Topology, North-Holland, Amsterdam, 2002, pp. 819–898. MR1886682 (2002m:53063)
145. Hans Rademacher, *Über partielle und totale differenzierbarkeit von Funktionen mehrerer Variablen und über die Transformation der Doppelintegrale*, Math. Ann. **79** (1919), no. 4, 340–359. MR1511935
146. K. Rajala, *Alexandrov-avaruudet, ylägradientit ja Poincarén epäyhtälö (in Finnish)*, Pro gradu -tutkielma. Jyväskylän yliopisto (2000), 1–41.
147. Yu. G. Reshetnyak, *Sobolev classes of functions with values in a metric space*, Sibirsk. Mat. Zh. **38** (1997), 657–675. MR1457485 (98h:46031)

148. ———, *On the conformal representation of Alexandrov surfaces*, Papers on Analysis, Rep. Univ. Jyväskylä Dep. Math. Stat., vol. 83, Univ. Jyväskylä, Jyväskylä, 2001, pp. 287–304. MR1886629 (2003b:53075)
149. Juha Rissanen, *Wavelets on self-similar sets and the structure of the spaces $M^{1,p}(E, \mu)$* , Ann. Acad. Sci. Fenn. Math. Diss. (2002), no. 125, 46. Dissertation, University of Jyväskylä, Jyväskylä, 2002. MR1880640 (2002k:42081)
150. W. Rudin, *Functional analysis*, second ed., McGraw-Hill Inc., New York, 1991. MR1157815 (92k:46001)
151. Stanisław Saks, *Theory of the integral*, Second revised edition. English translation by L. C. Young, with two additional notes by Stefan Banach, Dover Publications Inc., New York, 1964. MR0167578 (29:4850)
152. L. Saloff-Coste, *Uniformly elliptic operators on Riemannian manifolds*, J. Differential Geom. **36** (1992), 417–450. MR1180389 (93m:58122)
153. L. Schwartz, *Théorie des distributions. Tome I*, Actualités Sci. Ind., no. 1091 = Publ. Inst. Math. Univ. Strasbourg 9, Hermann & Cie., Paris, 1950. MR0035918 (12:31d)
154. S. Semmes, *Finding structure in sets with little smoothness*, Proc. of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 875–885. MR1403987 (97j:28010)
155. ———, *Finding curves on general spaces through quantitative topology, with applications to Sobolev and Poincaré inequalities*, Selecta Math. **2** (1996), 155–295. MR1414889 (97j:46033)
156. ———, *Good metric spaces without good parameterizations*, Rev. Mat. Iberoamericana **12** (1996), 187–275. MR1387590 (97e:57025)
157. ———, *On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A_∞ -weights*, Rev. Mat. Iberoamericana **12** (1996), 337–410. MR1402671 (97e:30040)
158. ———, *Mappings and spaces*, Quasiconformal Mappings and Analysis (Ann Arbor, MI, 1995), Springer, New York, 1998, pp. 347–368. MR1488459 (99c:30045)
159. ———, *Metric spaces and mappings seen at many scales (appendix)*, Birkhäuser Boston Inc., Boston, MA, 1999. Based on the 1981 French original, translated from the French by Sean Michael Bates. MR1699320 (2000d:53065)
160. ———, *Some novel types of fractal geometry*, Oxford Mathematical Monographs, Oxford University Press, New York, 2001. MR1815356 (2002h:53073)
161. ———, *Some topics concerning homeomorphic parameterizations*, Publ. Mat. **45** (2001), no. 1, 3–67. MR1829576 (2002c:57039)
162. N. Shanmugalingam, *Newtonian spaces: an extension of Sobolev spaces to metric measure spaces*, Ph.D. thesis, University of Michigan, 1999.
163. ———, *Newtonian spaces: an extension of Sobolev spaces to metric measure spaces*, Rev. Mat. Iberoamericana **16** (2000), no. 2, 243–279. MR1809341 (2002b:46059)
164. ———, *Harmonic functions on metric spaces*, Illinois J. Math. **45** (2001), no. 3, 1021–1050. MR1879250 (2003c:31010)
165. ———, *Some convergence results for p -harmonic functions on metric measure spaces*, Proc. London Math. Soc. (3) **87** (2003), no. 1, 226–246. MR1978575 (2005f:31010)
166. L. Siebenmann and D. Sullivan, *On complexes that are Lipschitz manifolds*, Geometric Topology (Proc. Georgia Topology Conf., Athens, Ga., 1977), Academic Press, New York, 1979, pp. 503–525. MR0537747 (80h:57027)
167. S. L. Sobolev, *On some estimates relating to families of functions having derivatives that are square integrable*, Dokl. Akad. Nauk SSSR **1** (1936), 267–270 (in Russian).
168. ———, *On a theorem in functional analysis*, Math. Sb. **4** (1938), 471–497 (in Russian).
169. ———, *Some applications of functional analysis in mathematical physics*, Translations of Mathematical Monographs, vol. 90, American Mathematical Society, Providence, RI, 1991. Translated from the third Russian edition by Harold H. McFaden, with comments by V. P. Palamodov. MR1125990 (92e:46067)
170. Michael Spivak, *A comprehensive introduction to differential geometry. Vol. II*, second ed., Publish or Perish Inc., Wilmington, Del., 1979. MR0532831 (82g:53003b)
171. E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton University Press, Princeton, NJ, 1970, Princeton Mathematical Series, No. 30. MR0290095 (44:7280)

172. ———, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Univ. Press, Princeton, NJ, 1993, Monographs in Harmonic Analysis, III. MR1232192 (95c:42002)
173. K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Mathematica **196** (2006), no. 1, 65–131. MR2237206
174. ———, *On the geometry of metric measure spaces. II*, Acta Mathematica **196** (2006), no. 1, 133–177. MR2237207
175. D. Sullivan, *Hyperbolic geometry and homeomorphisms*, Geometric Topology (Proc. Georgia Topology Conf., Athens, Ga., 1977), Academic Press, New York, 1979, pp. 543–555. MR0537749 (81m:57012)
176. ———, *Exterior d , the local degree, and smoothability*, Prospects in Topology (Princeton, NJ, 1994), Ann. of Math. Stud., vol. 138, Princeton Univ. Press, Princeton, NJ, 1995, pp. 328–338. MR1368667 (97d:57034)
177. L. Tonelli, *Sulla quadratura delle superficie*, Atti Reale Accad. Lincei **3** (1926), 633–638.
178. Tatiana Toro, *Surfaces with generalized second fundamental form in L^2 are Lipschitz manifolds*, J. Differential Geom. **39** (1994), no. 1, 65–101. MR1258915 (95b:49066)
179. ———, *Geometric conditions and existence of bi-Lipschitz parameterizations*, Duke Math. J. **77** (1995), no. 1, 193–227. MR1317632 (96b:28006)
180. Hans Triebel, *The structure of functions*, Monographs in Mathematics, vol. 97, Birkhäuser Verlag, Basel, 2001. MR1851996 (2002k:46087)
181. P. Tukia and J. Väisälä, *Lipschitz and quasiconformal approximation and extension*, Ann. Acad. Sci. Fenn. Ser. A I Math. **6** (1981), no. 2, 303–342 (1982). MR0658932 (84a:57016)
182. J. T. Tyson, *Quasiconformality and quasimetry in metric measure spaces*, Ann. Acad. Sci. Fenn. Ser. A I Math. **23** (1998), 525–548. MR1642158 (99i:30038)
183. ———, *Metric and geometric quasiconformality in Ahlfors regular Loewner spaces*, Conf. Geom. Dynam. **5** (2001), 21–73. MR1872156 (2002m:30026)
184. N. Th. Varopoulos, *Fonctions harmoniques sur les groupes de Lie*, C. R. Acad. Sci. Paris Sér. I Math. **304** (1987), 519–521. MR0892879 (88f:22026)
185. A. L. Vol’berg and S. V. Konyagin, *On measures with the doubling condition*, Izv. Akad. Nauk SSSR Ser. Mat. **51** (1987), 666–675. English translation: *Math. USSR-Izv.*, 30:629–638, 1988. MR0903629 (88i:28006)
186. M.-K. von Renesse, *On local Poincaré via transportation*, Mathematische Zeitschrift (to appear).
187. Nik Weaver, *Lipschitz algebras*, World Scientific Publishing Co. Inc., River Edge, NJ, 1999. MR1832645 (2002g:46002)
188. ———, *Lipschitz algebras and derivations. II. Exterior differentiation*, J. Funct. Anal. **178** (2000), no. 1, 64–112. MR1800791 (2002g:46040a)
189. Hassler Whitney, *Geometric integration theory*, Princeton University Press, Princeton, NJ, 1957. MR0087148 (19:309c)
190. J.-M. Wu, *Hausdorff dimension and doubling measures on metric spaces*, Proc. Amer. Math. Soc. **126** (1998), 1453–1459. MR1443418 (99h:28016)
191. K. Yosida, *Functional analysis*, Springer-Verlag, New York, 1980. MR0617913 (82i:46002)
192. W. P. Ziemer, *Weakly differentiable functions*, Graduate Texts in Mathematics, vol. 120, Springer-Verlag, New York, 1989. MR1014685 (91e:46046)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN 48109
E-mail address: juha@umich.edu