Modern Birkhäuser Classics

Many of the original research and survey monographs in pure and applied mathematics published by Birkhäuser in recent decades have been groundbreaking and have come to be regarded as foundational to the subject. Through the MBC Series, a select number of these modern classics, entirely uncorrected, are being re-released in paperback (and as eBooks) to ensure that these treasures remain accessible to new generations of students, scholars, and researchers.

Metric Structures for Riemannian and Non-Riemannian Spaces

Misha Gromov

With Appendices by

M. Katz, P. Pansu, and S. Semmes

English translation

by Sean Michael Bates

Reprint of the 2001 Edition

Birkhäuser

Boston • Basel • Berlin

Mikhail Gromov
Institut des Hautes Études Scientifiques
Département de Mathématiques
F-91440 Bures-sur-Yvette
France

Originally published as volume 152 in the series *Progress in Mathematics*

Cover design by Alex Gerasev.

Mathematics Subject Classification (2000): 01A60, 01A75, 51K99, 53-02, 53C20, 53C23, 53C70, 57N65

Library of Congress Control Number: 2006937425

ISBN-10: 0-8176-4582-9 e-ISBN-10: 0-8176-4583-7
ISBN-13: 978-0-8176-4582-3 e-ISBN-13: 978-0-8176-4583-0

Printed on acid-free paper.

Birkhäuser

Based on *Structures Métriques des Variétés Riemanniennes*
Edited by J. LaFontaine and P. Pansu

9 8 7 6 5 4 3 2 1

www.birkhauser.com (IBT)

Misha Gromov

with Appendices by
M. Katz, P. Pansu, and S. Semmes

Metric Structures for Riemannian and Non-Riemannian Spaces

Based on *Structures Métriques des Variétés Riemanniennes*
Edited by J. LaFontaine and P. Pansu
English Translation by Sean Michael Bates

Birkhäuser
Boston • Basel • Berlin

Author
Mikhail Gromov
Département de Mathématiques
Institut des Hautes Etudes Scientifiques
Bures-sur-Yvette, France

Translator (English edition)
Sean Michael Bates
Department of Mathematics
Columbia University
New York, NY 10027, USA

Editors (French Edition)
Jacques LaFontaine
Département des Sciences Mathématiques
Université de Montpellier
2, Place E. Bataillon,
34095 Montpellier Cedex 5, France

Pierre Pansu
Département de Mathématiques
Université de Paris-Sud
91405 Orsay Cedex, France

Library of Congress Cataloging-in-Publication Data
Gromov, Mikhail
[Structures metriques des varietes riemanniennes. English]
Metric structures for Riemannian and non-Riemannian spaces / Mikhail Gromov ; translated by Sean Michael Bates.
p. cm. -- (Progress in mathematics ; v. 152)
Includes bibliographical references and index.
ISBN 0-8176-3898-9 (acid-free paper)
1. Riemannian manifolds. I. Title. II. Series: Progress in mathematics (Boston, Mass.) ; vol. 152.
QA629.G8313 1998 97-24633
516.3'73--DC21 CIP

AMS Subject Classifications: 53B21, 53C20, 53C21, 53C23, 58D17, 54E35, 51H20, 51H25, 54E35, 28A78

Printed on acid-free paper.

Birkhäuser

ISBN 0-8176-3898-9 SPIN 10835928
ISBN 3-7643-3898-9

Typeset in LaTeX by Sean Michael Bates, New York, NY
Printed and bound by Hamilton Printing, Rensselaer, NY
Printed in the United States in America

9 8 7 6 5 4 3 2

...Meme ceux qui furent favorables à ma perception des verités que je voulais ensuite graver dans le temple, me félicitèrent de les avoir découvertes au microscope, quand je m'étais au contraire servi d'un télescope pour apercevoir des choses, très petites en effet, mais parce qu'elles étaient situées à une grande distance, et qui étaient chacune un monde.

Marcel Proust, *Le temps retrouvé*
(Pléiade, Paris, 1954, p. 1041)

Contents

Preface to the French Edition

This book arose from a course given at the University of Paris VII during the third trimester of 1979. My purpose was to describe some of the connections between the curvature of a Riemannian manifold V and some of its global properties. Here, the word *global* refers not only to the topology of V but also to a family of metric invariants, defined for Riemannian manifolds and mappings between these spaces. The simplest metric invariants of such V are, for example, its volume and diameter; an important invariant of a mapping from V_1 into V_2 is its dilatation. In fact, such invariants also appear in a purely topological context and provide an important link between the given infinitesimal information about V (usually expressed as some restriction on the curvature) and the topology of V.

For example, the classical Gauss–Bonnett theorem gives an upper bound for the diameter of a positively curved manifold V, from which one can deduce the finitude of the fundamental group of V. For a deeper topological study of Riemannian manifolds, we need more subtle invariants than diameter or volume; I have attempted to present a systematic treatment of these invariants, but this treatment is far from exhaustive.

Messrs. J. Lafontaine and P. Pansu have successfully completed the almost insurmountable task of transforming into a rigorous mathematical text the chaos of my course, which was scattered with imprecise statements and incomplete proofs. I thank them as well as M. Berger, without whose assistance and encouragement this book would never have come into being. I also thank the *Editions Cedic* for the liberty it afforded the authors at the time that the book was edited.

M. Gromov
Paris, June 1980

Preface to the English Edition

The metric theory described in this book covers a domain stretching somewhere between the fields of topology and global Riemannian geometry. The boundary of this domain has dramatically exploded since 1979 and accordingly, in the course of its translation from the 1979 French version into English, the book has approximately quadrupled in size, even though I tried not to maim the original text with unnecessary incisions, insertions, and corrections, but rather to add several new sections indicated by the $+$ subscript. The most voluminous additions are Chapter $3\frac{1}{2}{}_+$, which attempts to link geometry and probability theory, and Appendix B_+, where an analyst lays down his view on metric spaces. Here, the reader can painlessly learn several key ideas of real analysis made accessible to us geometers by the masterful exposition of Stephen Semmes, who has adapted his material to our non-analytic minds.

Additionally, Appendix C_+ reproduces my 1980-rendition of Paul Levy's inequality, while Misha Katz gives an overview of systolic freedom in Appendix D_+.

Acknowledgements: The initiative to publish an English translation of the book with Birkhäuser is due to Alan Weinstein. It was my pleasure to cooperate with Sean Bates, who translated the original version of the book and helped me edit the new sections. I am also grateful to Marcel Berger, Keith Burns, and Richard Montgomery for calling my attention to errors in the first version of the book.

M. Gromov
Bures-sur-Yvette, May 1997

Introduction: Metrics Everywhere

The conception of "distance" is already present in everyday language where it refers to two physical objects or two abstract ideas being mutually close or far apart. The most common (but by no means most general) mathematical incarnation of this vague idea is the notion of *metric space*, that is, an abstract set X where the distance between its elements, called *points* $x \in X$, is measured by *positive real numbers.* Thus a metric space is a set X with a given function in two variables, say $d : X \times X \to \mathbb{R}_+$ satisfying the famous *triangle inequality*

$$d(x, x'') \leq d(x, x') + d(x', x'')$$

for all triples of points x, x' and x'' in X.

Besides, one insists that the distance function be *symmetric*, that is, $d(x, x') = d(x', x)$. (This unpleasantly limits many applications: the effort of climbing up to the top of a mountain, in real life as well as in mathematics, is not at all the same as descending back to the starting point).

Finally, one assumes $d(x, x) = 0$ for all $x \in X$ and add the following *separation axiom.* If $x \neq x'$, then $d(x, x') = 0$. This seems to be an innocuous restriction, as one can always pass to the quotient space by identifying x and x' whenever $d(x, x') = 0$. But sometimes the separation becomes a central issue, e.g., for *Kobayashi* and *Hofer metrics*, where such identification may reduce X to a single point, for instance).

The archtypical example of a metric space is the ordinary Euclidean space $\mathbb{R}^n$ with the pythagorean distance between the points $x = (x_1, \dots , x_n)$ and $x' = (x'_1, \dots , x'_n)$ defined by

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + \cdots + (x_n - x'_n)^2}.$$

Next come subsets in $\mathbb{R}^n$ with this metric providing many appealing examples, such as *the sphere* $S^{n-1} = \{x \in \mathbb{R}^n | \sum_{i=1}^n x_i^2 = 1\}$ and the set of vertices of the unit cube, i.e., $\{0,1\}^n \subset \mathbb{R}^n$ with the *induced Euclidean* (pythagorean) distance. If X is a smooth connected submanifold in $\mathbb{R}^n$ (as the above sphere) then, besides the induced Euclidean distance $\mathrm{dist}_{\mathbb{R}^n}$ on X, one has the *induced Riemannian distance*, $\mathrm{dist}_X(x, x')$ defined as the infimum of lengths of curves *contained in* X and joining x and x'. (One may be tempted to use the above as a quick definition of a Riemannian metric. Indeed, every Riemannian manifold admits a smooth embedding into some $\mathbb{R}^n$ preserving the length of the curves according to the Nash theorem. But Euclidean embeddings hide rather than reveal the true metric structure of Riemannian manifolds due to uncontrolled distortion $\mathrm{dist}_X \mid \mathrm{dist}_{\mathbb{R}^n}$. Besides the full beauty and power of Riemannian geometry depend not only on the metric but also on the associated elliptic Riemannian equations, such as Laplace–Hodge, Dirac, Yang–Mills, and so on. These naturally come along with the Riemannian tensor but are nearly invisible on X embedded to $\mathbb{R}^n$.

Our approach to general metric spaces bears the undeniable imprint of early exposure to Euclidean geometry. We just love spaces sharing a common feature with $\mathbb{R}^n$. Thus there is a long tradition of the study of *homogeneous* spaces X where the isometry group acts *transitively* on X. (In the Riemannian case the metric on such X is fully determined by prescribing a positive definite quadratic form g_o on a single tangent space $T_{x_o}(X)$. But the simplicity of this description is illusory; it is quite hard to evaluate metric invariants of X in terms of g_o. For example, one has a very limited idea of how *systoles* (see below) behave as one varies a left invariant metric on a Lie group $SO(n)$ or $U(n)$ for instance.) Besides isometries, $\mathbb{R}^n$ possesses many nontrivial *self-similarities*, i.e., transformations f, such that $f^*(\mathrm{dist}) = \lambda\ \mathrm{dist}$ for some constant $\lambda \neq 0, 1$. There are no self-similar spaces besides $\mathbb{R}^n$ in the Riemannian category — this is obvious — but there are many such non-Riemannian examples such as p-adic vector spaces (these are totally disconnected) and some connected nilpotent Lie groups (e.g., the *Heisenberg group*) with Carnot–Carathéodory metrics (see 1.4, 1.18 and 2.6 in Appendix B).

Switching the mental wavelength, one introduces spaces with curvature $K \leq 0$ and $K \geq 0$ by requiring their small geodesic triangles to be "thinner" (correspondingly, "thicker") than the Euclidean ones (see 1.19_+). Here one is guided by the geometry of *symmetric spaces* that are distinguished homogeneous spaces, such as S^n and $\mathbb{C}P^n$ where $K \geq 0$ and $SL_n\mathbb{R}/SO(n)$ with $K \leq 0$.

Apart from direct Euclidean descendants there are many instances of metrics associated to various structures, sometimes in a rather unexpected and subtle way. Here are a few examples.

Complex manifolds. The complex space $\mathbb{C}^n, n \geq 1$, carries no metric invariant under holomorphic automorphisms. There are just too many of them! Yet, many complex (and almost complex) manifolds, e.g., *bounded* domains in $\mathbb{C}^n$, do possess such natural metrics, for example, the *Kobayashi metric* (see 1.8 bis$_+$).

Symplectic manifolds. No such manifold X of positive dimension carries an invariant metric, again because the group of symplectomorphisms is too large. Yet, the infinite dimensional space of closed *lagrangian* submanifolds in X (or rather each "hamiltonian component" of this space) does admit such a metric. (The construction of the metric is easy but the proof of the separation property, due to Hofer, is quite profound. Alas, we have no room for Hofer's metric in our book).

Homotopy category. Once can functorially associate an infinite dimensional metric polyhedron to the homotopy class of each topological space X, such that continuous maps between spaces transform to distance decreasing maps between these polyhedra. Amazingly, the metric invariants of such polyhedra (e.g., its systoles, the volumes of minimal subvarieties realizing prescribed homology classes) lead to new homotopy invariants which are most useful for (aspherical) spaces X with large fundamental groups (see Ch. 5H$_+$).

Discrete groups. A group with a distinguished finite set of generators carries a natural discrete metric which only moderately, i.e., bi-Lipschitzly, changes with a change of generators. Then, by adopting ideas from the geometry of noncompact Riemannian manifolds, one defines a variety of asymptotic invariants of infinite groups that shed new light on the whole body of group theory (see 3C, 5B, and 6B, C).

Lipschitz and bi-Lipschitz. What are essential maps between metric spaces? The answer "isometric" leads to a poor and rather boring category. The most generous response "continuous" takes us out of geometry to the realm of pure topology. We mediate between the two extremes by emphasizing *distance decreasing maps* $f : X \to Y$ as well as general λ-*Lipschitz* maps f which enlarge distances at most by a factor λ for some $\lambda \geq 0$.

Isomorphisms in this categeory are λ-*bi-Lipschitz homeomorphisms* and most metric invariants defined in our book transform in a λ-controlled way under λ-Lipschitz maps, as does for example, the *diameter* of a space, $\mathrm{Diam} X = \sup_{x,x'} \mathrm{dist}(x, x')$. We study several classes of such invariants with a special treatment of *systoles* measuring the minimal volumes of homology classes in X (see Ch. 4 and App. D) and of *isoperimetric profiles* of complete Riemannian manifolds and infinite groups which are linked in

Ch. 6 to quasiconformal and quasiregular mappings.

Asymptotic viewpoint. Since every diffeomorphism between *compact* Riemannian manifolds is λ-bi-Lipschitz for some $\lambda < \infty$, our invariants tell us preciously little if we look at a *fixed compact* manifold X. What is truly interesting in the λ-Lipschitz environment is the metric behavior of *sequences* of compact spaces. This ideology applies, for example, to an individual *noncompact* space X, where the asymptotic geometry may be seen as X is exhausted by a growing sequence of compact subspaces. We also study sequences of maps and homotopy classes of maps between fixed compact spaces, say $f_i : X \to Y$ (see Ch. 2, 7) and relate this asymptotic metric behavior to the homotopy structure of X and Y (with many fundamental questions remaining open).

Metric sociology. As our perspective shifts from individual spaces X to families (e.g., sequences) of these, we start looking at all metric spaces simultaneously and observe that there are several satisfactory notions of distance *between* metric spaces (see Ch. 3). Thus we may speak of various kinds of metric convergence of a sequence X_i to a space X and study the asymptotics of particular sequences. For example, if X_i, $i = 1, 2, \dots$, are Riemannian manifolds of dimension n with a fixed bound on their sectional curvatures, then there is a subsequence that converges (or collapses) to a mildly singular space of dimension $m \leq n$ (see Ch. 8).

Metric, Measure and Probability. Suppose our metric spaces are additionally given some measures, e.g., the standard Riemannian measures if we deal with compact Riemannian manifolds. Here one has several notions of metric convergence of spaces modulo subsets of measure $\to 0$ (see Ch. $3\frac{1}{2}{}_+$). Then there is a weaker convergence most suitable for sequences X_i with $\dim X_i \to \infty$. According to this, unit spheres $S^i \subset \mathbb{R}^{i+1}$ with normalized Riemannian measures converge (or concentrate, see $3\frac{1}{2}{}_+$) to a single atom of unit mass! This can be regarded as a geometric version of the *law of large numbers* that is derived in the present case from the *spherical isoperimetric inequality* (see $3\frac{1}{2}{}_+$E and Appendix C).

From local to global. This is a guiding principle in geometry as well as in much of analysis, and the reader will find it in all corners of our book. It appears most clearly in Ch. 5 where we explain how the lower bound on the *Ricci curvature of a manifold* X implies the *measure doubling property*, saying that the volume of each not very large $2r$-ball $B(2r) \subset X$ does not exceed const Vol $B(r)$ for the concentric ball $B(r) \subset B(2r)$. This leads to several topological consequences concerning the fundamental group of X (see Ch. 5).

Besides the volumes of balls, the Ricci curvature controls the isoperimetric profile of X. For example, the spherical isoperimetric inequality generalizes to the manifolds with Ricci $\geq -$ const (see Appendix C).

Analysis on metric spaces. Smooth manifolds and maps, being infinitesimally linear, appear plain and uneventful when looked upon through a microscope. But singular fractal spaces and maps display a kaleidoscopic variety of patterns on all local scales. Some of these spaces and maps are sufficiently regular, e.g., they may possess the doubling property, and provide a fertile soil for developing rich geometric analysis. This is exposed by Stephen Semmes in Appendix B.

I have completed describing what is in the book. It would take another volume just to indicate the full range of applications of the metric idea in various domains of mathematics.

Misha Gromov

April 1999

Chapter 1

Length Structures: Path Metric Spaces

Introduction

In classical Riemannian geometry, one begins with a C^∞ manifold X and then studies smooth, positive-definite sections g of the bundle S^2T^*X. In order to introduce the fundamental notions of covariant derivative and curvature (cf. [Grl–Kl–Mey] or [Milnor]$_{\mathrm{MT}}$, Ch. 2), use is made only of the differentiability of g and not of its positivity, as illustrated by Lorentzian geometry in general relativity. By contrast, the concepts of the length of curves in X and of the geodesic distance associated with the metric g rely only on the fact that g gives rise to a family of continuous norms on the tangent spaces T_xX of X. We will study the associated notions of length and distance for their own sake.

A. Length structures

1.1. Definition: The *dilatation* of a mapping f between metric spaces X, Y is the (possibly infinite) number

$$\mathrm{dil}(f) = \sup_{\substack{x,x' \in X \\ x \neq x'}} \frac{d(f(x), f(x'))}{d(x, x')},$$

where "d" stands for the metrics (distances) dist_X in X and dist_Y in Y. The *local dilatation* of f at x is the number

$$\mathrm{dil}_x(f) = \lim_{\varepsilon \to 0} \mathrm{dil}\left(f|_{B(x,\varepsilon)}\right).$$

A map f is called *Lipschitz* if $\mathrm{dil}(f) < \infty$; it is called λ-*Lipschitz* if $\mathrm{dil}(f) \leq \lambda$, in which case, the infimal such λ is called *the Lipschitz constant* of f.

If f is a Lipschitz mapping of an interval $[a, b]$ into X, then the function $t \mapsto \mathrm{dil}_t(f)$ is measurable.

1.2. Definition: The *length* of a Lipschitz map $f : [a, b] \to X$ is the number

$$\ell(f) = \int_a^b \mathrm{dil}_t(f)\, dt.$$

If f is merely continuous, we can define $\ell(f)$ as the supremum of all sums of the form $\sum_{i=0}^{n} d(f(t_i), f(t_{i+1}))$ where $a = t_0 \leq t_1 \leq \cdots \leq t_{n+1} = b$ is a finite partition of $[a, b]$.

If φ is a homeomorphism of a closed interval I' onto $I = [a, b]$, then ℓ satisfies $\ell(f \circ \varphi) = \ell(f)$, as follows from the fact that φ is strictly monotone (invariance under change of parameter).

The two definitions of $\ell(f)$ stated above are equivalent when f is absolutely continuous (cf. [Rinow], p. 106). This fact permits us to define $\ell(f)$ as the integral of the local dilatation when f is Lipschitz and to set $\ell(f \circ \varphi) = \ell(f)$ for each homeomorphism φ of I onto I'. More generally,

1.3. Definition: A *length structure* on a set X consists of a family $\mathcal{C}(I)$ of mappings $f : I \to X$ for each interval I and a map ℓ of $\mathcal{C} = \bigcup \mathcal{C}(I)$ into $\mathbb{R}$ having the following properties:

- (a) **Positivity:** We have $\ell(f) \geq 0$ for each $f \in \mathcal{C}$, and $\ell(f) = 0$ if and only if f is constant (we assume of course that the constant functions belong to $\mathcal{C}$).

- (b) **Restriction, juxtaposition:** If $I \subset J$, then the restriction to I of any member of $\mathcal{C}(J)$ is contained in $\mathcal{C}(I)$. If $f \in \mathcal{C}([a, b])$ and $g \in \mathcal{C}([b, c])$, then the function h obtained by juxtaposition of f and g lies in $\mathcal{C}([a, c])$ and $\ell(h) = \ell(f) + \ell(g)$.

- (c) **Invariance under change of parameter:** If φ is a homeomorphism from I onto J and if $f \in \mathcal{C}(J)$, then $f \circ \varphi \in \mathcal{C}(I)$ and $\ell(f \circ \varphi) = \ell(f)$.

- (d) **Continuity:** For each $I = [a, b]$, the map $t \mapsto \ell(f|_{[a,t]})$ is continuous.

Using conditions $(a), (b)$, and (c), we can define a pseudo-metric d_ℓ on X called the *length metric* by setting

$$d_\ell(x, y) = \inf\{\ell(f) : f \in \mathcal{C}, x, y \in \mathrm{im}(f)\}.$$

As usual, this pseudo-metric induces a topology on X.

It is common to define $\ell(f) = \infty$ when the map $f : I \to X$ is not contained in $C(I)$.

1.4. Examples:

(a) A metric space (X, d) is equipped with a canonical length structure: The set C consists of all continuous mappings from intervals into X, and the function ℓ is defined as in 1.2 above. The resulting structure is called the *metric length structure* of (X, d); in general, however, the length metric d_ℓ differs from d, and their corresponding topologies may also be distinct.

(b_+) *Tits-like metrics and snowflakes*: Consider $\mathbb{R}^n$ equipped with polar coordinates (r, s), where $r \in [0, \infty)$ and $s \in S^{n-1}$, the unit sphere in $\mathbb{R}^n$. Define

$$d(x_1, x_2) = |r_1 - r_2| + r\|s_1 - s_2\|^{1/2},$$

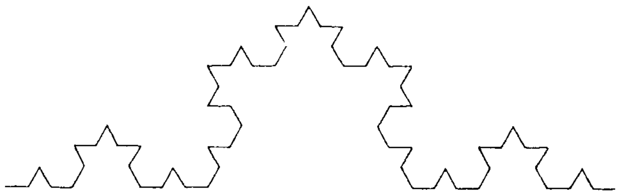
where $x_i = (r_i, s_i) \in \mathbb{R}^n$, $i = 1, 2$, $\|s_1 - s_2\|$ denotes the Euclidean distance on $S^{n-1} \subset \mathbb{R}^n$, and $r = \min\{r_1, r_2\}$. This d gives rise to the usual topology on $\mathbb{R}^n$, but

$$d_\ell(x_1, x_2) = \begin{cases} |r_1 - r_2| & \text{for } s_1 = s_2 \\ r_1 + r_2 & \text{for } s_1 \neq s_2, \end{cases}$$

and so $(\mathbb{R}^n, d_\ell)$ becomes the *disjoint* union of the Euclidean rays $[0, \infty) \times s$ for all $s \in S^{n-1}$, all glued together at the origin only. In particular, the unit sphere $S^{n-1} \subset \mathbb{R}^n$ is discrete with respect to d_ℓ. Metrics of this type naturally appear on (the ideal boundaries of) manifolds with nonpositive sectional curvatures and are collectively referred to as *Tits metrics* (cf. [Ba–Gr–Sch]).

An analog of the metric d_ℓ can be constructed on the subset of Euclidean 3-space consisting of the straight cone $X \subset \mathbb{R}^3$ over the Koch snowflake $S \subset \mathbb{R}^2$. (Here, the snowflake is the base of the cone and plays the role of the sphere S^{n-1} in the Tits-like example above). The only curves in X having finite Euclidean length are those contained in the (straight) generating lines of the cone, and so these lines are disjoint with respect to d_ℓ away from the vertex (compare [Rinow], p. 117, and Appendix B_+ of this book).

In general, the metrics d, d_ℓ always satisfy the inequality $d \leq d_\ell$, so that their corresponding topologies coincide if and only if for each $x \in X$ and $\varepsilon > 0$, there exists a d-neighborhood of x in which each point is connected to x by a curve of length at most ε.

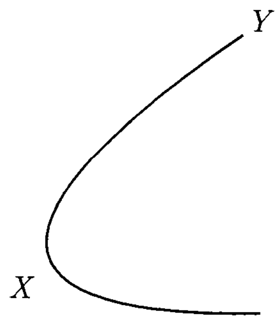


(c) If X is a manifold, then any Riemannian or Finslerian structure on X naturally gives rise to a length structure: One proceeds as in 1.2, noting that when f is differentiable, its local dilatation at a point x equals the norm of its derivative at x.

(d) *Induced length structures*: If X is equipped with a length structure and φ is a map from a set Y into X, then we obtain a length structure on Y by setting

$$\ell_Y(f) = \ell_X(\varphi \circ f)$$

for each $f\colon I \to Y$.



(e) *First exposure to Carnot–Caratheodory spaces.* We can associate a length structure on a Riemannian manifold (V, g) with any tangent subbundle $E \subset TV$ by defining the length of a curve c to be its usual Riemannian length if c is absolutely continuous and its tangent vector lies within E at a.e. point, and by setting $\ell(c) = \infty$ otherwise. If E is integrable, then the topology defined by d_ℓ is none other than the leaf topology. The case of nonintegrable E is of great interest.

A basic example of the latter structure is provided by the 3-dimensional Heisenberg group $\mathbb{H}^3$ of matrices of the form

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix},$$

equipped with a left-invariant metric. The quotient of $\mathbb{H}^3$ by its center C (isomorphic to $\mathbb{R}$) defines a Riemannian fibration (see [Ber–Gau–Maz], Ch. 1) of $\mathbb{H}^3$ over the Euclidean plane $\mathbb{H}^3/C \simeq \mathbb{R}^2$. The subbundle E then consists of the horizontal subbundle of this fibration, which coincides with the kernel of the 1-form $dz - x\,dy$.

1.5. Suppose X is equipped with a length structure ℓ, and let $\tilde{\ell}$ be the length structure defined by the metric d_ℓ. The following criterion, which is nothing more than an axiomatic version of the classical properties of the lengths of curves in metric spaces, describes when these two structures are identical.

1.6. Proposition: *If, for each interval I, the function ℓ is lower semicontinuous on $\mathcal{C}(I)$ with respect to the compact-open topology, then $\ell = \tilde{\ell}$.*

Proof. By 1.3(d), the function $t \mapsto \ell(f|_{[a,t]})$ is uniformly continuous on $I = [a, b]$. For each $\varepsilon > 0$, there exists $\eta > 0$ such that if $|t - t'| < \eta$, then $d_\ell(f(t), f(t')) < \varepsilon$.

Let $a = t_0 \leq t_1 \leq \cdots \leq t_{n+1} = b$ be a partition of I having increments no larger than η. For each integer i between 0 and n, there exists a map g_i in $\mathcal{C}([t_i, t_{i+1}])$ having the same values as f at t_i, t_{i+1} such that

$$\ell(g_i) \leq d_\ell(f(t_i), f(t_{i+1})) + \varepsilon/n.$$

By juxtaposing the g_i, we obtain a curve h_ε satisfying

$$\ell(h_\varepsilon) = \sum_{i=0}^{n} \ell(g_i) \leq \sum_{i=0}^{n} d_\ell(f(t_i), f(t_{i+1})) + \varepsilon \leq \tilde{\ell}(f) + \varepsilon$$

and such that for each $t \in I$, we have $d_\ell(h_\varepsilon(t), f(t)) \leq 3\varepsilon$.

From the hypothesis that ℓ is lower semicontinuous, it follows that

$$\ell(f) \leq \liminf_{\varepsilon \to 0} \ell(h_\varepsilon) \leq \tilde{\ell}(f),$$

whereas the opposite inequality is an immediate consequence of the definition of ℓ.

Remark: If ℓ is the length structure associated with a metric d, then the same argument as above shows that $\ell = \tilde{\ell}$, using the semicontinuity of length with respect to d (cf. 1.2 and [Choq], p. 137). In other words, by following the sequence of constructions

(X, d), a metric space	$\longrightarrow$	metric length structure ℓ on X	$\longrightarrow$	(X, d_ℓ), a new metric associated with the length structure,

we obtain the same length structure. Nevertheless, we again emphasize that $\ell \neq \tilde{\ell}$ in general.

$1.6\frac{1}{2}$+ Locality of the length structure. If two length structures agree on some open subsets covering X, then they are obviously equal. Conversely, if we are given a covering of X by open subsets X_i for $i \in I$, together with length structures ℓ_i on the X_i which are compatible on the intersections $X_i \cap X_j$ for all $i, j \in I$, then there obviously exists a (unique) length structure on X that restricts to ℓ_i on each X_i. (In other words, the length structures comprise a *sheaf* over X.) On the other hand, metrics on X are not local (they form only a presheaf over X), but they can be localized as follows: Given a metric d on X, we consider all metrics d' that are *locally majorized* by d. This means that for each $x \in X$, there exists a neighborhood $Y_x \subset X$ of x such that $d|_{Y_x} \geq d'|_{Y_x}$. Now take the supremum of all these d' and call it d_m. (Note that the supremum of a bounded family of metrics is again a metric. In general, this supremum may be infinite at some pairs of points in X, but otherwise it looks like a metric.) Clearly $d_m \leq d_\ell$ in any metric space (X, d); if (X, d) is *complete*, then $d_m = d_\ell$, as a trivial argument shows (see Section 1.8 below).

B. Path metric spaces

1.7. Definition: A metric space (X, d) is a *path metric space* if the distance between each pair of points equals the infimum of the lengths of curves joining the points (i.e., if $d = d_\ell$).

Examples: Note that, according to this definition the Euclidean plane is a path metric space, but the plane with a segment removed is *not*.

The n-sphere S^n is not a path metric space when equipped with the metric induced by that of $\mathbb{R}^{n+1}$, but it *is* a path metric space for the geodesic metric by Proposition 1.6.

Path metric spaces admit the following simple characterization.

1.8. Theorem: *The following properties of a metric space (X, d) are equiv-*

alent:

1. *For arbitrary points $x, y \in X$ and $\varepsilon > 0$, there is a z such that*
$$\sup(d(x,z), d(z,y)) \leq \frac{1}{2}\, d(x,y) + \varepsilon.$$

2. *For arbitrary $x, y \in X$ and $r_1, r_2 > 0$ with $r_1 + r_2 \leq d(x,y)$, we have*
$$d(B(x,r_1), B(y,r_2)) \leq d(x,y) - r_1 - r_2,$$
for
$$d(B_1, B_2) = \inf_{\substack{x' \in B_1 \\ y' \in B_2}} d(x', y').$$

Every path metric space has these properties, and conversely, if (X,d) is complete and satisfies (1) or (2), then it is a path metric space.

Proof. Let (X,d) be a complete metric space satisfying condition (1), and set $\delta = d(x,y)$. Given a sequence (ε_k) of positive numbers, there is a point $z_{1/2}$ such that $\max(d(x,z_{1/2}), d(z_{1/2},y)) \leq \delta/2 + \varepsilon_1\delta/2$, and points $z_{1/4}, z_{3/4}$ for which each of the distances $d(x,z_{1/4}), d(z_{1/4},z_{1/2}), d(z_{1/2},z_{3/4}), d(z_{3/4},y)$ are less than
$$1/2(\delta/2 + \varepsilon_1\delta/2) + \varepsilon_2(\delta/2 + \varepsilon_1\delta/2), \text{ etc.}$$
By choosing the sequence (ε_k) so that $\sum_k \varepsilon_k < \infty$, we can define a map f from the dyadic rationals in $[0,1]$ into X satisfying
$$d\left(f\left(\frac{p}{2^n}\right), f\left(\frac{p+1}{2^n}\right)\right) \leq \frac{\delta}{2^n} \prod_{k=1}^{\infty} (1 + \varepsilon_k).$$
If (X,d) is complete, then this map extends to the entire interval $[0,1]$. Since we can choose the ε_k so that the product $\prod(1+\varepsilon_k)$ is arbitrarily close to 1, we obtain curves whose lengths tend to $\delta = d(x,y)$, which proves the last assertion.

The implication (1) $\Rightarrow$ (2) is proven in the same way, whereas (2) $\Rightarrow$ (1) and the assertion that a path metric space satisfies (1),(2) are trivial.

Path metric spaces enjoy some of the same geometric properties as Riemannian manifolds.

1.8bis. Property: If (X,d) is a path metric space, and if f is a map of X into a metric space Y, then the dilatation of f obviously equals the supremum of its local dilatation, i.e., $\operatorname{dil}(f) = \sup_{x \in X} \operatorname{dil}_x(f)$. Note that if X and Y are Riemannian manifolds, and if f is differentiable, then the differential $Df_x \colon T_xX \to T_{f(x)}Y$ satisfies $\mathrm{dil}_x(f) = \|Df_x\|$.

1.8bis$_+$ Kobayashi metrics. Let $\triangle$ be a path metric space and let X be an arbitrary (say, topological) space with a distinguished set of maps $f \colon \triangle \to X$. Consider all metrics d' on X for which these f have $\mathrm{dil}(f) \leq 1$, i.e., for which the mappings are (nonstrictly) distance decreasing, and define d_K as the supremum of the metrics d' on X. Here, it is convenient to admit *degenerate* metrics d' (in the sense that $d'(x,y) = 0$ for perhaps some $x \neq y$), so that d_K may itself be degenerate. In fact, this d_K is a (possibly degenerate) path metric by the property above.

In the classical example, due to Kobayashi, $\triangle$ is the unit open disk equipped with the Poincaré metric (i.e., the hyperbolic plane), X is a complex analytic space, and the collection of distinguished maps consists of all holomorphic mappings $\triangle \to X$. The usefulness of this metric is based on the *Schwarz lemma* (and its various generalizations), which implies that d_K is *nondegenerate* for many X. Such X are said to be (Kobayashi) *hyperbolic*. For example, the disk $\triangle$ is itself hyperbolic since d_K in this case equals the Poincaré metric (following from the fact that every holomorphic map $\triangle \to \triangle$ is distance decreasing with respect to the Poincaré metric, a consequence of the classical Schwarz-Ahlfors lemma). The basic features of the Kobayashi metric and hyperbolicity do not depend on the integrability of the implied (almost) complex structure of X and therefore extend to all *almost complex* manifolds X (via the theory of pseudo-holomorphic curves in X, cf. [McD–Sal]). For example, hyperbolicity is stable under small (possibly singular) perturbations of almost complex structures on compact manifolds and (suitably defined) singular almost complex spaces (compare [Kobay], [Brody], [Krug–Over]).

There is also a real analog of Kobayashi hyperbolicity, in which X is a Riemannian manifold, $\triangle$ is as above, and the set of distinguished maps consists of all conformal, globally area minimizing mappings $\triangle \to X$. In this case, hyperbolicity of X is equivalent to δ-hyperbolicity (see (e) in 1.19$_+$ below) under mild restrictions on X, which are satisfied, for example, if X is the universal cover of a compact manifold. (In fact, X does not have to be a manifold here — it can be a rather singular space, e.g., a simplicial polyhedron as in 1.15$_+$, see [Gro]$_{\mathrm{HG}}$, [Gro]$_{\mathrm{HMGA}}$.)

1.9. Definition: A *minimizing geodesic* in a path metric space (X,d) is any curve $f \colon I \to X$ such that $d(f(t), f(t')) = |t - t'|$ for each $t, t' \in I$. A *geodesic* in X is any curve $f : I \to X$ whose restriction to any sufficiently small subinterval in I is a minimizing geodesic.

In this connection, we have the following :

Hopf–Rinow theorem. *If (X, d) is a complete, locally compact path metric space, then*

1. *Closed balls are compact, or, equivalently, each bounded, closed domain is compact.*

2. *Each pair of points can be joined by a minimizing geodesic.*

Before turning to the proof of the theorem, we observe that if (X, d) is a complete, locally compact metric space, then there are many noncompact balls for the metric $d' = \inf(1, d)$.

1.11 Compactness of closed balls. Note that if a is a point in X, then the ball $B(a, r)$ is by hypothesis compact if r is sufficiently small. We will first show that if $B(a, r)$ is compact for all r in an interval $[0, \rho)$, then $B(a, \rho)$ is compact as well.

Let (x_n) be a sequence of points in $B(a, \rho)$. We may suppose that the distances $d(a, x_n)$ tend to ρ; otherwise, there is a ball $B(a, r)$ with $r < \rho$ containing infinitely many of the x_n and thus a limit point of the sequence. Let (ε_p) be a sequence of positive real numbers tending to zero. By applying property (2) of Theorem 1.8, we find that for each p, there exists an integer $n(p)$ such that for each $n > n(p)$, there is a point y_n^p satisfying

$$y_n^p \in B(a, \rho - 2\varepsilon_p) \qquad \text{and} \qquad d(x_n, y_n^p) \leq \varepsilon_p.$$

For each p, the sequence (y_n^p) lies within a compact set; by a diagonal argument (or since the product of compact sets is compact), it follows that there is a sequence of integers (n_k) such that the subsequence $(y_{n_k}^p)$ converges for all p. The sequence (x_{n_k}), which is the uniform limit of the $(y_{n_k}^p)$, is a Cauchy sequence and therefore converges by completeness of X.

By the preceding remarks, the supremum of the r for which $B(a, r)$ is compact is infinite: if instead it equalled $\rho < \infty$, then we could find $\rho' > \rho$ such that $B(a, \rho)$ would be compact, by using a finite covering of the sphere $S(a, \rho)$ by compact balls.

1.12. Existence of a minimizing geodesic joining two arbitrary points

We first consider the case when X is compact.

Lemma: *If (X, d) is a compact path metric space and $a, b \in X$, then there exists a curve of length $d(a, b)$ joining a and b.*

Proof. It suffices to consider curves $f\colon [0,1] \to X$ which are parametrized by arc length. From the definition of path metric spaces, if follows that for each positive integer n, there exists such a curve f_n joining a to b and having length less than $d(a,b) + 1/n$. The set of f_n is therefore equicontinuous, and by Ascoli's theorem, there exists a subsequence f_{n_k} that converges uniformly to a curve $f : [0,1] \to X$. Since the length function ℓ is lower semicontinuous, we have

$$\ell(f) \leq \liminf_{k\to\infty} \ell(f_{n_k}) = d(a,b).$$

In the case of a complete, locally compact, but noncompact path metric space, it suffices to note that the images of the curves f_n chosen in the preceding paragraph lie within the compact ball $B(a, 2d(a,b))$.

1.13. Remarks: (a) In the case of Riemannian manifolds, this proof fulfills the promise made in the introduction that use would be made only of the associated length structure.

(b) The equicontinuity argument of Lemma 1.12 also shows that *in a compact path metric space, every free homotopy class is represented by a length-minimizing curve*, and that the minimizing curves are geodesics. Moreover, if X is a manifold, then for each real r, *there is only a finite number of homotopy classes represented by curves of length less than* r (again, it suffices to use Ascoli's theorem and the fact that the homotopy classes are open subsets of $C^0(S^1, X)$; cf. [Dieu], p. 188). These results also hold for homotopy classes of curves based at a point $x \in X$ and geodesics based at x (but not necessarily smooth at x) and will play a key role, particularly in Chapter 5.

C. Examples of path metric spaces

1.14. Riemannian manifolds with boundary and subsets of $\mathbb{R}^n$ with smooth boundary. Let X be a domain in $\mathbb{R}^n$ with smooth boundary, equipped with the metric and length structure induced by that of $\mathbb{R}^n$, and let f be the identity map

$$(X, \text{induced metric}) \to (X, \text{induced length metric}).$$

It is easy to see that if the boundary of X is smooth, then $\mathrm{dil}_x(f) = 1$ for each $x \in X$, and that $\mathrm{dil}(f) = 1$ if and only if X is convex.

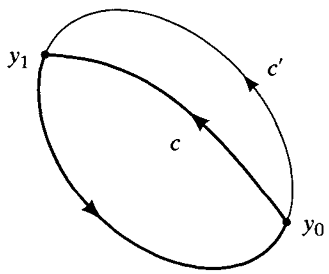
Distortion$_+$: More generally, let X be a subset of a path metric space A and let $\mathrm{distort}(X)$ denote the dilatation of the identity map $f\colon X \to X$ with respect to the two induced metrics, i.e.,

$$\mathrm{distort}(X) = \sup \frac{(\text{length dist})|_X}{\text{dist}|_X}.$$

Our first observation is the following:

(*a*) *Let X be a compact subset of $\mathbb{R}^n$. If* $\mathrm{distort}(X) < \frac{\pi}{2}$ *(which means that every two points in X that lie within a Euclidean distance of d from one another can be joined by a curve in X of length $< d\pi/2$), then X is simply connected.*

Proof. To prove this assertion, we argue by contradiction. Suppose $\pi_1(X) \neq 0$ and let α be a nontrivial homotopy class in which there exists a curve of minimal length among all homotopically nontrivial loops (the existence of such α is guaranteed by Remark 1.13(b)). Let Y be the image of c and $g: Y \to Y$ the identity map of the space Y with its induced length structure. We claim that $\mathrm{dil}(g) = \mathrm{dil}(f|_Y)$.



To prove the claim, let y_1, y_0 be two points of Y and fix a parametrization of Y by arc length, i.e., a map $c\colon [0,\ell] \to Y$ such that $c(0) = y_0 = c(\ell)$ and $y_1 = c(d)$ for some $d \in [0,\ell]$ such that $d \leq \ell - d$. Then $c|_{[0,d]}$ is the shortest path joining y_0 to y_1 in X. Indeed, if there were a strictly shorter path c' from y_0 to y_1, then the two loops obtained by adjoining c' and the two parts of c defined by the parameters 0 and d would be strictly shorter than c. Since their product is homotopic to c, however, we could conclude that one of the two is not homotopic to 0 in X, which contradicts the minimality of c. Since the path $c|_{[0,d]}$ lies within Y, it follows that d is the distance from y_0 to y_1 for the length metrics of X and Y, and that

$$\mathrm{dil}_{(y_0,y_1)}(g) = \mathrm{dil}_{(y_0,y_1)}(f).$$

Thus, we have $\mathrm{dil}(g) < \pi/2$. Extend c to a periodic function on $\mathbb{R}$ and set $r(s) = d(c(s), c(s + \ell/2))$, so that the inequality $r(s) \geq \ell/2\,\mathrm{dil}(g)$ holds. Set $u(s) = (c(s + \ell/2) - c(s))/r(s)$. The curve u is differentiable almost everywhere, and its image lies within the unit sphere of $\mathbb{R}^n$. Moreover, $u(s + \ell/2) = -u(s)$, so that the length of u is at least 2π. Thus,

$$\left\|\frac{du}{ds}\right\|^2 \leq \frac{1}{r(s)^2}\left(4 - \left(\frac{dr}{ds}\right)^2\right) \leq \frac{4}{r(s)^2} \leq \left(\frac{4\,\mathrm{dil}(g)}{\ell}\right)^2,$$

and so $\ell(u) \leq 4\,\mathrm{dil}(g) < 2\pi$, which is the desired contradiction.

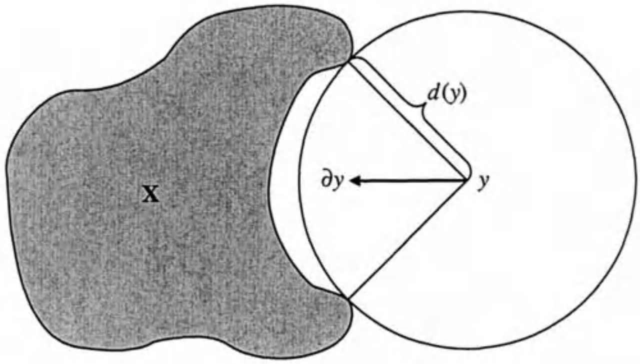
Remark: If $\mathrm{dil}(d) = \pi/2$, and if X is not simply connected, then X contains a round circle.

(*b*) *If* $\mathrm{distort}(X) < \pi/2\sqrt{2}$, *then* X *is contractible.*

Proof$_+$. The idea is to homotopy-retract $\mathbb{R}^n$ to X by following the flow of a suitable vector field ∂ which plays the role of $-\operatorname{grad} d(y)$ for the distance function $d\colon y \mapsto \mathrm{dist}(y, X) = \inf_{x\in X} \|y - x\|$. In general, the function d is nonsmooth, even on the complement $\mathbb{R}^n \setminus X$. Nonsmoothness at a point $y \in \mathbb{R}^n \setminus X$ is due to the fact that the sphere S_y^{n-1} at y of radius $d(y)$ can meet X at several points. These points essentially realize the above infimum, since the *open* ball bounded by S_y^{n-1} does not intersect X, while the set $X \cap S_y^{n-1}$ is nonempty and $d(y) = \|x - y\|$ for all $x \in X \cap S_y^{n-1}$. Now we observe that the normal projection $X \to S_y^{n-1}$ is distance-decreasing and thus $(\text{length dist})_X(x_1, x_2) \geq (\text{length dist})_{S_y^{n-1}}(x_1, x_2)$ for all pairs of points in the intersection $X \cap S_y^{n-1}$. It follows that the latter distance is bounded by $\pi d(y)/2$, and since we assume the *strict* inequality $\mathrm{distort}(X) < \pi/2\sqrt{2}$, the distance above is bounded by $\delta\pi d(y)/2$ for some $\delta < 1$ independent of y. Consequently, the intersection $X \cap S_y^{n-1}$ is *strictly* contained in a hemisphere, or, in other words, *there exists a unit vector* ∂_y *at every* $y \in \mathbb{R}^n \setminus X$ *such that*

$$\partial_y \|x - y\| \leq -\varepsilon < 0 \qquad \text{for all} \qquad x \in X \cap S_y^{n-1}, \tag{$*$}$$

where $\partial_y\|x - y\|$ *denotes the* ∂_y*-derivative of the (distance) function* $y \mapsto \|x - y\|$. In fact, one can take ∂_y to be the vector which points towards center of the minimal spherical cap in S_y^{n-1} containing $X \cap S_y^{n-1}$, thus obtaining a (Borel) measurable vector field $y \mapsto \partial_y$ satisfying $(*)$.
Finally, we can easily smooth this vector field, so that the resulting (now smooth!) unit vector field, say $y \mapsto \overline{\partial}_y$ on $\mathbb{R}^n \setminus X$, satisfies $(*)$ (possibly with a smaller $\varepsilon > 0$) as well. Clearly, every forward orbit of such a field



converges to a point in X, and so the flow generated by $\overline{\partial}$ eventually retracts $\mathbb{R}^n$ to X.

Remark$_+$: A sharper result is proved in Appendix A, where one allows $\operatorname{distort}(X) < 1 + \alpha_n$ for some specific $\alpha_n > 1 - \pi/2\sqrt{2}$. Furthermore, there are many examples of (necessarily contractible) subsets X with arbitrarily large (even infinite) distortion for which $X \cap S_y^{n-1}$ is still strictly contained in an open hemisphere for each $y \in \mathbb{R}^n \setminus X$ and to which our argument applies. On the other hand, we do not know the precise value of α_n for which $\operatorname{distort}(X) < 1 + \alpha_n$ necessarily implies that X is contractible.

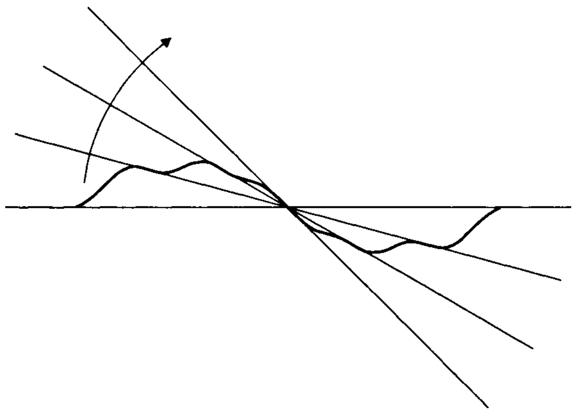
Exercise$_+$: Construct a closed, convex surface X in $\mathbb{R}^3$ with $\operatorname{distort}(X) < \pi/2$. (Compare Appendix A.)

Problem$_+$: Given a topological space X, evaluate the infimum of all distortions induced by embeddings $X \to \mathbb{R}^n$ or of those distortions induced by embeddings which lie in a fixed isotopy class. The first interesting case arises when X is the circle and we minimize the distortion for X knotted in $\mathbb{R}^3$ in a prescribed way (compare [Gro]$_{\text{HED}}$, [O'Hara]$_{\text{EK}}$).

Remark$_+$: The geometry of subsets $X \subset \mathbb{R}^n$ satisfying $\operatorname{distort}(X) \leq 1+\alpha_n$ can be rather complicated, even for small $\alpha_n > 0$. For example, there are simple smooth arcs in $\mathbb{R}^2$ with arbitrarily small (i.e. close to 1) distortion which have an arbitrarily large turn of the tangent direction. To see this, consider diffeomorphisms $T_i : \mathbb{R}^2 \to \mathbb{R}^2$ with the following properties: Each T_i fixes the complement of the disk of radius 2^{-i} around the origin and isometrically maps the disk of radius 2^{-i-1} into itself by rotating it by a small angle $\alpha > 0$.

Clearly, one can choose the T_i so that they and their inverses are $(1+\varepsilon)$-

Lipschitz for $\varepsilon < 10\alpha$. Since the distortion of such a T_i is located in the annulus between 2^{-i} and 2^{-i+1}, the diffeomorphism given by the composition $U_i = T_1 \circ T_2 \circ \cdots \circ T_i$ are also $(1+\varepsilon)$ Lipschitz along with their inverses U_i, since the "distortion supports" of the T_i are mutually disjoint. Note, however, that for large i, the map U_i is the identity outside of the unit disk and rotates the 2^{-i-1} disk by the angle $i\alpha$, which can be quite large for large i. Also, the U_i-image of the unit segment in $\mathbb{R}^2$ passing through the origin may have arbitrarily large rotation of the tangent direction, while its distortion does not exceed $1 + \varepsilon$.



The full richness of this picture becomes apparent in Appendix B_+.

Asymptotic distortion$_+$. Define

$$\operatorname{distort}(X; D) = \sup \operatorname{dist}(x_1, x_2),$$

where the supremum ranges over all $x_1, x_2 \in X \subset A$ with $\operatorname{dist}_A(x_1, x_2) \geq D$ with respect to the induced length metric, and study the asymptotic behavior of this distortion as $D \to \infty$. For example, if X is a properly embedded, noncompact submanifold in *hyperbolic space* $\mathbb{H}^n$, then either $\operatorname{distort}(X) < \infty$ or $\operatorname{distort}(X; D)$ grows *exponentially* as $D \to \infty$. On the other hand, the distortion of every connected subgroup X of a Lie group A is *at most* exponential. This is easy to see for $A = GL_n$ and for algebraic groups in general (see [Gro]$_{\mathrm{AI}}$); the (more difficult) nonalgebraic case was recently settled by Varopoulos (see [Var]).

1.15$_+$ Polyhedral metrics. Assign a length structure to the standard n-simplex

$$\triangle^n = \{x_0, x_1, \ldots, x_n \geq 0 : \sum_{i=0}^{n} x_i = 1\} \subset \mathbb{R}^{n+1}$$

that is invariant under the automorphisms of $\triangle^n$ corresponding to permutations of the vertices of $\triangle^n$ (labelled by the indices $0, 1, \dots, n$). Suppose this is done *coherently* for all $\triangle^n$, for $n = 1, 2, \dots$, i.e., so that the chosen length structure on $\triangle^{n-1}$ coincides with the length structure induced by the standard (face) embedding $\triangle^{n-1} \subset \triangle^n$ from the chosen length structure on $\triangle^n$. For example, we can take the Euclidean structure on $\triangle^n \subset \mathbb{R}^{n+1}$, or the spherical structure induced by the radial projection $\triangle^n \to S^n$. Furthermore, we can modify the ambient metric in $\mathbb{R}^{n+1}$, e.g., by using the ℓ_1-norm $\|x\|_{\ell_1} = \sum_{i=0}^n |x_i|$, or we can identify $\triangle^n$ with the regular simplex of diameter d in hyperbolic n-space for a given $d \in [0, \infty]$.

Once the length structure on $\triangle^n$ has been chosen, every simplicial polyhedron K acquires a unique length structure which restricts to this structure on all of its simplices (compare $1.6\frac{1}{2}{}_+$). The resulting path metric spaces exhibit an unexpected richness of geometry. Even in the simplest case where K is a *tree* (i.e., a contractible 1-complex), the picture is not quite trivial (just try to figure out when a 4 point metric space admits an isometric embedding into such a tree). And if we allow infinite-dimensional polyhedra K, then we will be able to apply metric treatment to arbitrary homotopy types (of spaces) via their (semi-) simplicial models (see 5.42).

$1.15\frac{1}{2}{}_+$ Semialgebraic sets. A subset $X \subset \mathbb{R}^n$ is called (real) *algebraic* if it is defined via some polynomials P_i, $i = 1, \dots, I$, on $\mathbb{R}^n$ as follows:

$$X = \{x \in \mathbb{R}^n : P_i(x) = 0 \text{ for } i = 1, \dots, I\}.$$

We call X *special semialgebraic* if, in addition to the polynomial equations above, we allow finitely many strict and nonstrict polynomial inequalities:

$$X = \{x \in \mathbb{R}^n : P_i(x) = 0; Q_j(x) \geq 0; R_k(x) > 0\}.$$

The basic measure of the geometric complexity of such X is the sum of the degrees of these polynomials:

$$D(P_i, Q_j, R_k) = \sum_i \deg(P_i) + \sum_j \deg(Q_j) + \sum_k \deg(R_k),$$

or, more precisely, the infimum $\operatorname{algdeg}(X)$ of the numbers $D(P_i, Q_j, R_k)$, where the P_i, Q_j, R_k range over all polynomials which represent a given X as above (the representation of X by P_i, Q_j, R_k is highly nonunique. Note, for instance, that X remains the same when we replace the given polynomials by their cubes P_i^3, Q_j^3, R_k^3). Finally, *semialgebraic* X are defined as a finite union of special semialgebraic subsets.

It is not hard to see that the induced length structure on $X \subset \mathbb{R}^n$ is rather regular (albeit far from fully understood). In particular, the length metric induces the same topology on (each connected component of) X as the induced metric. In fact, if X is an algebraic curve of degree d, then it meets each hyperplane in at most d points, and so the length of the intersection of X with a ball of radius R does not exceed $2Rd$ by the Crofton formula. This gives a rough idea of the length metric on such X when we recall (the well known fact) that for each $x \in X$, there exists $R > 0$ such that the intersection of X with the Euclidean r-ball around x is connected for all $r < R$. (In fact, $X \cap B(x,r)$ is *contractible for all semialgebraic* X and $r \leq R = R(x) > 0$.) This implies that

$$\operatorname{lengthdist}(x,y) \leq 2d \operatorname{dist}(x,y)$$

for every $x \in X$ and all $y \in X \cap B(x,r)$. Moreover, by projecting a semialgebraic curve X onto lines in $\mathbb{R}^n$ and then isolating "branches" of X, one can improve the above estimate to

$$\operatorname{lengthdist}(x,y) \leq (1+\varepsilon) \operatorname{dist}(x,y), \tag{$(\leq)_\varepsilon$}$$

where $\varepsilon = \varepsilon(x,r) \to 0$ as $r \to 0$. (Probably $(\leq)_\varepsilon$ holds for all semialgebraic sets X.) Note that the convergence $\varepsilon \to 0$ might be nonuniform in x, as is seen in the case when X consists of two smooth curves which are tangent at a single point. But, one can show that every bounded, connected semialgebraic set X satisfies

$$\operatorname{lengthdist}(x,y) \leq C \operatorname{dist}^\alpha(x,y), \tag{$(\leq)_\alpha$}$$

for some $C = C(X)$ and $\alpha = \alpha(\operatorname{algdeg}(X)) > 0$. This is weaker than $(\leq)_\varepsilon$ for individual x, but it has the advantage of being satisfied by all x and y, with C independent of these points. (See $[\text{Gro}]_{\text{Yom}}$, $[\text{Gro}]_{\text{SGSS}}$ for more information and references about the geometry of semialgebraic sets)

1.16_+ Quotient spaces X/Γ. If (X, d_X) is a metric space and $f\colon X \to Y$ is a surjective map, then Y can be equipped with the *quotient metric* d, equal to the supremum of those d' on Y for which the map f is (nonstrictly) distance decreasing (i.e., with $\operatorname{dil}(f) \leq 1$). Notice that the metric d may degenerate (as did the Kobayashi metrics), but if $Y = X/\Gamma$ for a group Γ that acts *isometrically* on X and that has *closed* orbits, then d is nondegenerate. (Here, the distance between the orbits $y_1 = \Gamma(x_1)$ and $y_2 = \Gamma(x_2)$ clearly satisfies $d(y_1,y_2) = \inf_{\gamma\in\Gamma} d(x_1, \gamma(x_2))$.) If d_X is a *path* metric, then so is d, since its construction is local. In fact, the length structure in Y can sometimes be recaptured (but not always!) by lifting curves c from Y to X and then defining $\operatorname{length}_Y(c)$ as the infimum of the X-lengths of all such lifts. Ideally, we would like to have *horizontal* lifts $\tilde{c}$ of curves c in Y to X

so that $\text{length}_X(\tilde{c}) = \text{length}_Y(c)$, and this is possible, for example, when X is a *Riemannian manifold* and $Y = X/\Gamma$. Here, we *declare* a smooth curve $\bar{c}$ in X to be *horizontal* if it is normal to the orbits of Γ (which is no restriction at all if Γ is discrete) and introduce the length structure in Y using the projection c of these horizontal $\bar{c}$ to X, where we define

$$\text{length}_Y(c) = \text{length}_X(\bar{c}).$$

Then the metric in Y associated with this length structure equals the quotient metric d in Y, as easily follows from the slice theorem for isometry groups. Note that this Y is singular unless the action of Γ is free. This remark will play an important role in Chapters 3 and 8.

1.17$_+$ Covering metrics. If $\tilde{X} \to X$ is a covering map, then length structures, being local, lift from X to $\tilde{X}$, as do path metrics. In fact, every path metric d on X lifts to a unique path metric $\tilde{d}$ for which the covering map is a *local* isometry.

Although the definition of $\tilde{d}$ is obvious, the actual determination of its essential properties can become insurmountably difficult. For example, if X is a compact Riemannian manifold or a finite simplicial polyhedron, then all we may care to ask about d stands clearly before our eyes. But if we pose some seemingly inocuous question about $\tilde{d}$ on the *universal cover* $\tilde{X} \to X$, e.g., whether $\tilde{X}$ has finite or infinite diameter, then we encounter an (algorithmically) unsolvable decision problem of finiteness of the fundamental group of X. Here, the logical complexity of the problem is ingrained in the very definition of the path metric, the infimum of length over *all* curves between given points. This "all" makes the passage length→distance highly non-effective! Thus, complete knowledge about "length" does not immediately translate into comparable information about "distance," even in such simple cases as a Lie group X equipped with a left invariant Riemannian metric d. Here $d = d_g$ is uniquely determined by the prescription of a positive definite quadratic form g on a single tangent space, T_eX, but evaluating $d_g(x_1, x_2)$ for given x_1 and x_2 may be difficult. (To see the point, take $X = GL_n\mathbb{R} \subset \mathbb{R}^{n^2}$ and measure the d_g-distance between the matrices id and $-\text{id}$ in terms of g on $T_{\text{id}}X = \mathbb{R}^{n^2}$, or try to estimate the diameter of $X = (SO(n), d_g)$ in terms of g.)

1.18. Carnot–Caratheodory metrics associated with a tangent subbundle of a Riemannian manifold. One can show that the length structure associated with a completely nonintegrable subbundle (cf. 1.4(e)) — as in the case of the Heisenberg group — gives rise to a distance on V that induces the same topology but radically alters the metric properties (e.g., Hausdorff dimension, etc.) of V.

1.19$_+$ Curvature. It is hardly possible to find a convincing definition of the curvature (tensor) for an arbitrary metric space X, but one can distinguish certain classes of metric spaces corresponding to Riemannian manifolds with curvatures of a given type. This can be done, for example, by imposing inequalities between mutual distances of finite configurations of points in X. More precisely, let M_r denote the space of positive symmetric $r \times r$ matrices, and let $K_r(X) \subset M_r$ be the subset realizable by the distances among r-tuples of points in X. Thus, $(m_{ij}) \in K_r(X)$ if and only if there exist points $x_1, \dots, x_r$ in X, such that $\mathrm{dist}(x_i, x_j) = m_{ij}$, for $i, j = 1, \dots, r$. Then every subset $\mathcal{K} \subset M_r$ defines the (global) *$\mathcal{K}$-curvature class*, which consists of the spaces X with $K_r(X) \subset \mathcal{K}$, and the *local $\mathcal{K}$-curvature class*, where each point $x \in X$ is required to admit a neighborhood U with $K_r(U) \subset \mathcal{K}$. In this setting, we have the basic

Curvature problem: Given $\mathcal{K} \subset M_r$, describe the spaces X in the (local or global) $\mathcal{K}$-curvature class.

In general, a $\mathcal{K}$-condition on X does not seem to lead to an interesting theory unless one exercises good judgement in choosing the set $\mathcal{K}$. Geometers usually start with some standard (model) space Y (or with a family of such spaces) and take $\mathcal{K} = K_r(Y)$ (or the union of these for a given family of spaces Y). If one is willing to sacrifice the combinatorial aspect of the problem, then the phenomena associated with discrete spaces can be ruled out by imposing various connectivity restrictions on X, such as local path connectedness, local contractability, *controlled contractability*, etc. In particular, one can require X to be a *path metric* space, in which case one can go amazingly far with a judiciously chosen $\mathcal{K}$. But it is difficult to decide by looking at a given $\mathcal{K}$ whether the corresponding $\mathcal{K}$-class is "interesting." For instance, when does there exist a *path* metric space X with $K_r(X) = \mathcal{K}$ for a given $\mathcal{K} \subset M_r$?

Examples: (*a*) If X is compact and connected, then (obviously) the only information contained in $K_2(X)$ is the *diameter* of X, i.e., $\sup \mathrm{dist}(x_1, x_2)$, where the supremum ranges over all $x_1, x_2 \in X$.

(*b*) If $r = 3$, then for any metric space X, the *triangle inequality* requires $K_3(X)$ to lie within the subset $\mathrm{Tri} \subset M_3$ consisting of matrices (m_{ij}) that satisfy $m_{ij} \leq m_{ik} + m_{kj}$ for all i, j, k running through $1, 2, 3$. It seems that most *path* metric spaces X with *infinite* diameter satisfy $K_3(X) = \mathrm{Tri}$ (with some notable exceptions such as $X = \mathbb{R}$ or $\mathbb{R} \times$ compact), but I do not know a good criterion that guarantees this property. For example, all *uniformly contractible* manifolds X of dimension ≥ 2 have large $K_3(X)$, probably always equal to Tri.

(c) The curvature story starts (and essentially ends, in the present state of the art) with $r = 4$. For example, K_4 characterizes Euclidean spaces as follows:

If a complete path metric space X has $K_4(X) \subset K_4(\mathbb{R}^2)$, then it is isometric to a convex subset in $\mathbb{R}^2$. If $K_4(X) \subset K_4(\mathbb{R}^3)$, then X is isometric to a convex subset in a Hilbert space.

The proof is straightforward and easily extends to complete simply connected spaces Y with constant curvature in place of $\mathbb{R}^3$ (see [Rinow] §41). However, a similar characterization has not been worked out for other Riemannian *symmetric spaces* Y, where one expects every complete path metric space X with $K_r(X) \subset K_r(Y)$ for some (sufficiently large) $r = r(Y)$ to be (isometric to) a convex subset in another (possibly infinite dimensional) symmetric space Y' that naturally extends the geometry of Y (as, for example, the Grassmannian manifold $Gr_k(\mathbb{R}^\infty)$ extends $Gr_k(\mathbb{R}^n)$). On the other hand, it is easy to see that (convex subsets in) normed vector spaces are characterized by K_5, since the homogeneity of a norm is a five point property.

(*d*) *Alexandrov spaces with* $K \geq 0$: One could define this class by taking the union of $K_4(Y) \subset M_4$ for all *convex* complete surfaces $Y \subset \mathbb{R}^3$ equipped with their induced path metrics. Customarily, however, one uses a somewhat larger $\mathcal{K}(\geq 0) \subset M_4$, where the $\mathcal{K}$-positivity condition is felt only by those quadruples of points $\{x_i\} \subset X$, $i = 1, \ldots, 4$, with the property that x_4 lies *between* x_2 and x_3, i.e., x_4 satisfies

$$\operatorname{dist}(x_4, x_2) + \operatorname{dist}(x_4, x_3) = \operatorname{dist}(x_2, x_3). \tag{$\triangle$}$$

In other words, one defines $K \geq 0$ by requiring that every 1-Lipschitz map $f : \{x_1, x_2, x_3\} \to \mathbb{R}^2$ extend to a 1-Lipschitz map $\{x_1, x_2, x_3, x_4\} \to \mathbb{R}^2$ for all quadruples $\{x_i\}$ in X satisfying ($\triangle$), where "1-Lipschitz" means $\operatorname{dil}(f) \leq 1$ (see [Alex], [Rinow]). The resulting relation $\mathcal{K}(\geq 0) \subset M_4$, called the *Alexandrov–Topogonov comparison inequality* for $K \geq 0$, is quite resilient, e.g., *locally* $K \geq 0$ *implies globally* $K \geq 0$ (the proof is nontrivial, see [Top], [Bu–Gr–Per]), isometric quotients X/Γ inherit $K \geq 0$ from X (this is straightforward), etc., and thus leads to a full-fledged geometric theory without any regularity assumptions on X (see [Per]$_{\text{SCBB}}$ and the references therein).

The Alexandrov–Topogonov inequality can be thought of as a kind of concavity condition on the distance function $x \mapsto \operatorname{dist}(x_0, x)$ that rules out, for example, all non-Euclidean normed spaces (see [Rinow], p. 33). These, however, reside within a larger class of path metric spaces X having a less stringent concavity condition on dist. This condition can be described, for example, by requiring that *inward equidistant deformations preserve convexity of hypersurfaces in X.* However, the corresponding theory (starting from the K_5 or K_6 characterization of this property) has not yet been developed.

The $K \geq 0$ theory generalizes to $K \geq \kappa$ for given $\kappa \in \mathbb{R}$, but other differential geometric curvature positivity conditions have not been extended to singular spaces so far. One might expect that the positivity of the curvature operator and that of the complexified sectional curvature (especially on the isotropic planes) admit K_r renditions with some not very large r. On the other hand, the curvature classes Ricci ≥ 0 and Sc ≥ 0 lie beyond the reach of the K_r language. (See [Gro]$_{\text{sign}}$, [Gro]$_{\text{PCMD}}$ for a survey of various positive curvatures; also see Chapters 5 and 8 of this book.)

(*e*) *Alexandrov spaces with $K \leq 0$*: These spaces are defined via a suitable $\mathcal{K}(\leq 0) \subset M_4$ that expresses the Cartan–Alexandrov–Topogonov comparison inequality for $K \leq 0$. (Here, the test spaces Y are convex space-like surfaces in the light cone of the Lorentz $(++-)$-space.) The singular path spaces with $K \leq 0$ (also called CAT(0)-spaces) are by far more numerous than those with $K \geq 0$. For example, simplicial (as well as nonsimplicial) trees equipped with length metrics and cartesian products of these spaces have $K \leq 0$. In fact, there are plenty of polyhedra with $K \leq 0$ locally at each point, and in practice, the inequality $K \leq 0$ can be confirmed by an elementary combinatorial argument. On the other hand, the local condition $K \leq 0$ implies strong global properties of X, such as its *asphericity* (a nice property that is hard to prove using traditional topology). In fact, those polyhedra with $K \leq 0$ are the core of the geometro-combinatorial approach to infinite groups initiated by Dehn at the turn of the century.

Following Busemann [Busemann], one can extend the notion $K \leq 0$ to encompass normed linear spaces. The enlarged class is distinguished by the (local) convexity of the distance function, $x \mapsto \text{dist}(x_0, x)$ on every geodesic segment in X, or by the property that the convexity of hypersurfaces is preserved under outward equidistant deformations (compare [Gro]$_{\text{HMGA}}$, [E–H–S]).

Another notion (generalizing $K \leq -\kappa < 0$, rather than $K \leq 0$) is as follows: Consider all simplicial trees Y and let Tre $\subset M_4$ be the union of $K_4(Y)$ over all trees Y. Then take the δ-neighborhood Tre$_\delta \subset M_4$ of Tre for some $\delta > 0$ (defined with respect to the obvious Euclidean metric in M_4) and introduce *δ-hyperbolic* spaces X as those with $K_4(X) \subset$ Tre$_\delta$. Here, the essential geometry appears at large scales $>> \delta$, where the behavior of δ-hyperbolic spaces is indistinguishable from those with $K < 0$ (see [Gro]$_{\mathrm{HG}}$).

The traditional definition of $\mathcal{K}(\leq 0) \subset M_4$ is dual to that for given $\mathcal{K}(\geq 0)$ since it involves the extendability of 1-Lipschitz maps *from* $\mathbb{R}^2$ *into* X. Namely, for each quadruple $\{x_i\} \subset \mathbb{R}^2$ satisfying equation $(\triangle)$ in (d) above, every 1-Lipschitz map $\{x_1, x_2, x_3\} \to X$ should extend to a 1-Lipschitz map $\{x_i\} \to X$, $i = 1, \dots, 4$.

Exercise: Let X_+ satisfy the Alexandrov–Topogonov condition $K \leq 0$, and let $Z \subset X_+$ be the union of the minimizing geodesic segments whose ends comprise a three point subset $\{x_1, x_2, x_3\} \subset X_+$. Show that every 1-Lipschitz map $\{x_1, x_2, x_3\} \to X$ admits a 1-Lipschitz extension to Z.

Kirszbraun–Lang–Schroeder[1] Theorem: *Let X_+ and X_- be path metric spaces having $K \geq 0$ and $K \leq 0$ respectively in the Alexandrov sense. Then, every* 1*-Lipschitz map $X'_+ \to X_-$ for $X'_+ \subset X_+$ admits a 1-Lipschitz extension to all of X_+.*

If $X_+ = X_- = \mathbb{R}^n$, this is a classical result (see [Kirsz], [Fed]$_{\mathrm{GMT}}$, [Gro]$_{\mathrm{MV}}$), while the general case is quite recent (see [Lan–Sch], where the authors treat the case when $K(X_+) \geq \kappa \geq K(X_-)$ for all $\kappa \in \mathbb{R}$). Notice that the Euclidean (and the spherical) case of this theorem follows from monotonicity of the volume of intersections of balls in $\mathbb{R}^n$, i.e., $\mathrm{vol}(B(x_0, r_0) \cap B(x_1, r_1) \cap \cdots \cap B(x_n, r_n))$ is monotone decreasing in the distances $\|x_i - x_j\|$ between the centers of the balls (which was probably known in antiquity). Also observe the following equivalent version of the Kirszbraun theorem.

($\star$) *Let $f\colon A \to S^n$ be a λ-Lipschitz map from an arbitrary subset $A \subset S^n$ where $\lambda < 1$. Then $f(A)$ is strictly contained in a hemisphere of diameter $\leq \pi - \varepsilon < \pi$. (If $\lambda = 1$ and $f(A)$ is* not *contained in a hemisphere, then f is isometric.)*

Exercises: Prove the monotonicity claim above and derive Kirszbraun's theorem from it. Derive ($\star$) from Kirszbraun's theorem, and conversely show that ($\star$) implies Kirszbraun's theorem by studying optimal λ-Lipschitz extensions of 1-Lipschitz (partially defined) maps with minimal possible λ. (A reader versed in $K \geq, \leq 0$ geometry may notice that the Lang-Schroeder

[1]When I met Petrunin in Spring 96, I asked him whether the Kirszbraun theorem extended to Alexandrov spaces. He responded with "yes, for dim $< \infty$," hardly hiding his surprise at my inability to furnish a proof by myself. The next day, I found the preprint by Lang and Schroeder with a complete proof of the most general case, and similar ideas were communicated to me somewhat earlier by Branka Pavlovic.

theorem for (smooth!) Riemannian manifolds also follows from ($\star$) via the study of optimal λ-Lipschitz maps).

Remark: It is unclear which spaces Y (say with $K(Y) \le 0$) admit Lipschitz extensions of maps from $A \to Y$ to X for arbitrary metric spaces X and subsets $A \subset X$, where the implied Lipschitz constant of the extension $X \to Y$ is allowed to be bigger than that for the original map $A \to Y$. It is rather obvious that $Y = \mathbb{R}$ has this property (with $\mathrm{Lip}\,|_{\mathbb{R}} = \mathrm{Lip}\,|_A$), and consequently this is true for $\mathbb{R}^n$ (with $\mathrm{Lip}\,|_{\mathbb{R}^n} \le \delta_n \,\mathrm{Lip}\,|_A$). The same applies to trees (in place of $\mathbb{R}$) and their products. The question is open, however, in the case of the hyperbolic plane,[2] for example (see [Ball] for some results in this direction).

(f) Finally, we mention the possibility of having $\mathcal{K}_x \subset M_r$ depend on $x \in X$ and then consider the (variable) local $\mathcal{K}_x$-curvature class at each $x \in X$. This notion captures, for example, the idea of *local pinching*, which requires that the sectional curvatures of X at each point x be pinched between $K(x)$ and $\kappa K(x)$, for some positive (or negative) functions $K(x)$ on X and a local pinching constant $\kappa > 0$. (This also seems to be the appropriate framework for the K_r-rendition of the positivity of the curvature operator and the complexified sectional curvatures mentioned earlier).

D. Arc-wise isometries

1.20. The requirements for a map to be an isometry or even local isometry are too stringent to provide a sufficiently rich class of morphisms for path metric spaces. For example, any Riemannian n-manifold that is locally isometric to $\mathbb{R}^n$ must be flat. A more flexible notion is the following.

1.21. Definition: An *arc-wise isometry* of path metric spaces X, Y is a map $f\colon X \to Y$ such that $\ell(f \circ c) = \ell(c)$ for each Lipschitz curve $c\colon I \to X$ of X.

Examples:

(1) Every closed, piecewise C^1 curve admits an arc-wise isometric mapping into $\mathbb{R}$.

[2] This was recently settled by Branka Pavlovic.

(2) Every flat manifold of dimension $n < 5$ admits an arc-wise isometry into $\mathbb{R}^n$ that is additionally piecewise linear (see [Zalg]). The question is open for $n \geq 5$.

(3) An intuitively plausible property, such as the nonexistence of arc-wise isometries from X to Y when $\dim(X) > \dim(Y)$, is trivial to prove for C^1 maps but less transparent in general. The proof requires Rademacher's theorem (cf. $[\mathrm{Fed}]_{\mathrm{GMT}}$ 3.1.6), which states that Lipschitz mappings are differentiable almost everywhere.

We end this chapter with a few results obtained using the methods of Nash and Kuiper.

1.22. *If X, Y are Riemannian manifolds with* $\dim(Y) \geq \dim(X)$, *then there exists an arc-wise isometry from X into Y.*

In particular, incredible as it seems, every n-dimensional Riemannian manifold X admits an arc-wise isometry $X \to \mathbb{R}^n$. Of course, such mappings will not be C^1 in general!

1.23 An approximation problem. Given a Lipschitz mapping $f_0 \colon X \to Y$ and $\varepsilon > 0$, does there exist an arc-wise isometry $f_\varepsilon \colon X \to Y$ such that

$$d(f_0, f_\varepsilon) = \sup_{x \in X} d(f_0(x), f_\varepsilon(x)) \leq \varepsilon?$$

Evidently, we must have $\operatorname{dil}(f_0) \leq 1$.

1.24. Definition: A mapping f between path metric spaces is called *short* if $\operatorname{dil}(f) \leq 1$ and *strictly short* if $\operatorname{dil}(f) < 1$.

1.25. Theorem (cf. $[\mathbf{Gro}]_{\mathrm{PDR}}$): *If X, Y are Riemannian manifolds with* $\dim(Y) \leq \dim(X)$, *and if f is a strictly short mapping from X into Y, then the approximation problem has a positive solution.*

This brings us back to the basic problem of the existence of a homotopy of a given continuous map to a short map which is addressed in the following chapter.

$1.25\frac{1}{2}{}_{+}$. If one feels disgusted by the spineless flexibility of arc-wise isometric maps, then some rigidifying conditions can be added so that the remaining mappings f in question have no *folds*. (The construction underlying the proof of 1.25 uses multiply folded maps $X \to Y$.) For example, one may insist that $f \colon X \to Y$ be *open* (i.e., that f map open subsets of X onto open subsets of Y), which is quite a strong assumption for Lipschitz (and especially arc-wise isometric) maps. Secondly, if X and Y are oriented manifolds of the same dimension, then one may require the *local degree* of f to be positive. In the present setting, this means (usually one uses a stronger condition $\text{locdeg}_x > 0$, $x \in X$) that whenever f properly maps some open subset $U \subset X$ onto a subset $V \subset Y$, the degree of $f: U \to V$ as defined in 2.A is > 0. Thirdly, one may express the same idea by requiring that the inequality $\text{Jacobian}(f)_x > 0$ holds for almost all points $x \in X$ (which makes sense for Lipschitz maps). Finally, one can introduce *co-Lipschitz* maps (compare [B–J–L–P–S] as follows: Let

$$\text{codil}(f) = \sup_{y_1, y_2} \frac{\text{dist}_H(f^{-1}(y_1), f^{-1}(y_2))}{\text{dist}(y_1, y_2)},$$

where (y_1, y_2) runs over all pairs of distinct points in Y and dist_H denotes the Hausdorff distance (see Section 3.B). If $\text{codil}(f) \leq \lambda$, then f is called *λ-co Lipschitz.* The quotient map $X \to Y = X/\Gamma$ for a proper isometry group Γ acting on X is 1-co-Lipschitz (as well as 1-Lipschitz), for example.

Next, consider the R-balls $B_x(R)$ in X around some point $x \in X$ and let r be the radius of the maximal ball in Y centered at $f(x)$ and contained in the image $f(B_x(R)) \subset Y$. We define

$$\text{codil}_x(f) = \limsup_{R \to 0} \frac{R}{r}$$

and observe that if Y is a *path*-metric space and X is *locally compact* (which can often be relaxed), then

$$\text{codil}(f) = \sup_{x \in X} \text{codil}_x(f).$$

In fact, if $\text{codil}_x(f) \leq c < \infty$, then every path p: $[0,1] \to Y$ can be covered by a path $\tilde{p}$: $[0,1] \to Y$ issuing from a given point $y \in Y$ over $p(0) \in X$. Furthermore, if p is 1-Lipschitz, then $\tilde{p}$ can be chosen to be c-Lipschitz.

Exercises:

(a) Show that if X is complete and $f: X \to Y$ is co-Lipschitz, then Y is also complete.

(b) Study the following implications between the preceding four conditions for Lipschitz maps between suitably oriented equidimensional Riemannian manifolds:

$$\text{co-Lipschitz} \Rightarrow \text{Jacob} \geq 0 \Leftrightarrow \text{locdeg} \geq 0 \Rightarrow \text{open}.$$

(*c*) Decide which of these four conditions ensure that *arc-wise isometric* maps are local isometries. (I did not solve (*b*) and (*c*) myself, but I have prepared a good excuse for this.)

(*d*) Compare "Lipschitz + co-Lipschitz" with quasiregular maps (see Section 6.A).

1.26_+ Historical and terminological remarks. The idea of length structures and path metric spaces goes back at least as far as Gauss, who studied how the bending of surfaces in $\mathbb{R}^3$ distorts the induced metric but preserves the length structure. In classical geometry, one distinguishes the induced path metric with the name *inner metric*, and the word "inner" is often used for general length spaces as well. Traditionally, geometers were most interested in those "inner metric spaces" that are similar to Riemannian manifolds, and much effort was spent trying to identify and study such spaces with inner (i.e., path) metrics (see [Busemann], [Rinow], [Alex], [Alex–Zalg]).

In this book, we emphasize a different point of view, in which the Riemannian idea is stretched to its limits in order to capture such non-Riemannian beasts as Carnot–Caratheodory spaces (also called "nonholonomic spaces" and "sub-Riemannian" or "subelliptic" manifolds) and trees with infinite branching at every (!) point. In fact, one should not even restrict attention to the length structure, especially if one is attracted by *fractal spaces* (one can also play around with area instead of length, as inspired by symplectic geometry, string theory, and/or the area isoperimetric inequalities in geometric group theory).

Finally, we call the reader's attention to the paper [Bing], where it is shown that *every compact, connected and locally connected metrizable space admits a path metric.*

Chapter 2

Degree and Dilatation

A. Topological review

Throughout this chapter, M and N will denote connected, oriented, C^∞ manifolds having the same dimension n. Additionally, M is assumed to be compact and without boundary.

2.1. Proposition: *Let $f: M \to N$ be a C^∞ map, so that there exist regular values of f, i.e., $y \in N$ such that for each $x \in f^{-1}(y)$ the differential Df_x has rank n. For each such $y \in N$, the set $f^{-1}(y)$ is finite, and if we set $o(x) = 1$ if Df_x preserves orientation and $o(x) = -1$ otherwise, then the number*

$$\deg(f, y) = \sum_{x \in f^{-1}(y)} o(x)$$

does not depend on the regular value y.

(See [Milnor]$_{\text{TDV}}$ or [Berger]$_{\text{Cours}}$, Ch. 7).

2.2. Definition: The *degree* $\deg(f)$ of a smooth map $f: M \to N$ is the number $\deg(f, y)$, for any regular value y of f.

2.3. Proposition (see [Milnor]$_{\text{TDV}}$): *If $f, g: M \to N$ are homotopic, then* $\deg(f) = \deg(g)$.

2.4. Definition: If $f: M \to N$ is continuous, then the degree of f is defined as the degree of any smooth map homotopic to f.

2.5. Remark: If some $y \in N$ does not lie in the image of a continuous map $f: M \to N$, then $\deg(f) = 0$. Note that if g is smooth and sufficiently approximates f, then g is homotopic to f, and y does not lie in the image of g. Thus, y is a regular value of g, and $\deg(f) = \deg(g) = \deg(g, y) = 0$.

2.6. Proposition: *The integration of n-forms defines an isomorphism $H^n(M;\mathbb{R}) \to \mathbb{R}$. Moreover, the condition that $\int_M \omega = 1$ determines a unique class $\omega_M \in H^n(M;\mathbb{R})$; similarly, if N is compact, there exists a corresponding class $\omega_N \in H^n(N;\mathbb{R})$. If $f : M \to N$ is continuous, then $f^*\omega_N$ is proportional to ω_M, and the proportionality constant equals the degree of f defined above. (See [Massey] and [Godb], p. 212).*

2.7. Remark: If M is a compact, oriented Riemannian manifold, then its *volume form* is defined as the unique differential n-form ν_M, which for every $x \in M$ has value 1 on each oriented, orthonormal frame in T_xM. The *volume* of M equals the integral $\mathrm{vol}(M) = \int_M \nu_M$. If N is another compact, oriented Riemannian manifold with volume form ν_N, and $f : M \to N$ is smooth, then $\deg(f) = (\mathrm{vol}(N))^{-1} \int_M f^*\nu_N$. At each point $m \in M$, the skew-symmetric n-form $f^*\nu_N$ is proportional to ν_M by a factor known as the *Jacobian* of f at the point m, denoted $J(f, m)$. If f is a diffeomorphism from a subset A of M onto a subset B of N, then by the change-of-variables formula, we have

$$\int_A f^*\nu_N = \int_A J(f,m)\,\nu_M = \int_B \nu_N = \mathrm{vol}(B),$$

so that we are justified in setting

$$\mathrm{vol}(f|_A) \stackrel{\mathrm{def}}{=} \int_A f^*\nu_N,$$

even if f is not a diffeomorphism. In this notation, we have

$$\deg(f) = \mathrm{vol}(f|_M)/\,\mathrm{vol}(N),$$

where

$$\mathrm{vol}(f) \stackrel{\mathrm{def}}{=} \mathrm{vol}(f|_M).$$

In general, the degree of a map from M to N does not completely determine its homotopy class. For example, the mapping $S^1 \times S^1 \to S^1 \times S^1$ given by $(x, y) \mapsto (1, y)$ is not surjective and therefore has zero degree, although the map is not homotopic to 0. Homotopy classes of mappings into spheres, however, *are* determined by their degree (see $[\text{Milnor}]_{\text{TDV}}$, p. 51).

2.8. Theorem (H. Hopf): *If M is a compact, oriented n-manifold, then two mappings $M \to S^n$ are homotopic if and only if they have the same degree.*

$2.8\frac{1}{2}_+$ Remarks about degree.

(*a*) The pointwise degree $\deg(f, y)$ can also be defined for open and/or disconnected manifolds M and N; all one needs is the finiteness of the pullback $f^{-1}(y) \subset M$. This finiteness condition is (obviously) satisfied for regular values y of *proper* maps f, i.e., those f for which $f^{-1}(Y) \subset M$ is *compact for all compact subsets* $Y \subset N$ (when M and N are allowed to have boundaries $\partial M, \partial N$, respectively, the definition of properness usually includes the requirement that $f(\partial M) \subset \partial N$). If f is proper, then $\deg(f, y)$ is locally constant on each connected component of N. In particular, $\deg(f, y)$ is independent of y if N is connected, in which case it is regarded as $\deg(f)$. This number does not change under homotopies of proper maps and therefore extends to all *continuous proper maps* $f: M \to N$ (compare [Milnor]$_{\mathrm{TDV}}$). If M is an open connected manifold without boundary, then *the degree of proper maps* $f: M \to \mathbb{R}^n$, *denoted* $f \mapsto \deg(f) \in \mathbb{Z}$, *establishes a bijection between the homotopy classes of such maps and* $\mathbb{Z}$, provided that $n \geq 2$, by a (trivial) modification of Hopf's theorem. (If $n = 1$, then every proper map $M \to \mathbb{R}$ has degree ± 1 for connected M, which must be homeomorphic to $\mathbb{R}$.)

(*b*) The notion of degree extends to certain nonproper maps $f: M \to N$ such as:

(*i*) Maps f that are constant at infinity and on ∂M, i.e., constant outside a compact subset $K \subset \mathrm{Int}(M)$.

(*ii*) If $\dim(M) = \dim(N) \geq 2$, one can also allow maps $M \to N$ which are *locally constant* outside some compact subset $K \subset \mathrm{Int}(M)$.

(*iii*) Maps f which send $\mathrm{Int}(M) \backslash K$ onto a subset of topological codimension ≥ 2 in N, i.e., those having $\mathrm{rank}(f|_{\partial M}) < \dim \partial M$ for compact M.

In all of these cases, $\deg(f)$ is homotopy invariant and is defined for the continuous maps satisfying (*i*), (*ii*), or (*iii*). Such maps also satisfy the conclusion of Hopf's theorem. For example, if M is compact and connected, then the homotopy classes of maps $f : (M, \partial M) \to (S^n, s_0)$ bijectively correspond with $\mathbb{Z}$ via $f \mapsto \deg(f)$.

(*c*) Let N be a Riemannian manifold with finite total volume, $\mathrm{vol}(N) < \infty$, and let $f: M \to N$ be a smooth map with

$$|\mathrm{vol}|(f) \overset{\mathrm{def}}{=} \int_M f^*(|\nu_N|) < \infty,$$

(where $f^*(|\nu_N|) = |Jf|\nu_M$ for *Riemannian manifolds* M, and where ν_M, ν_N denote the oriented Riemannian volume forms). Then one can define

$$\deg^*(f) = \frac{\mathrm{vol}(f)}{\mathrm{vol}(N)},$$

where we recall that

$$\mathrm{vol}(f) \overset{\mathrm{def}}{=} \mathrm{vol}(f|_M) \overset{\mathrm{def}}{=} \int_M f^*\nu_N.$$

It is easy to show that this degree can be obtained by integrating the point degree over N, i.e.,

$$\deg^*(f) = \int_N \deg(f,y)\,\nu_N,$$

where in fact the integration is taken over the set of regular values of f, which has full measure in N. Thus, this $\deg^*(f)$ coincides with the topological (homotopy invariant) degree $\deg(f)$ whenever both make sense.

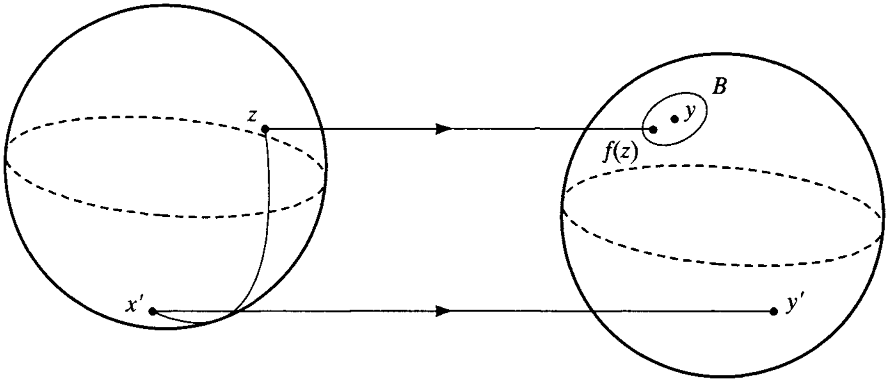
B. Elementary properties of dilatations for spheres

Given two path metric spaces V, W, what can be said about the space of mappings from V into W having dilatation less than or equal to a given number D? In particular, does there exist such a mapping in each homotopy class? This is a considerable question, which will occupy us in this chapter as well as in Chapter 7. When D is small, few homotopy classes are represented by such maps; for example, if V, W are spheres equipped with their standard metrics (i.e., that of the unit sphere in Euclidean space), then we have

2.9. Proposition: *If a map $f : S^n \to S^n$ has dilatation strictly less than 2, then the degree of f equals 1, 0, or -1 (cf. [Osser], and for further results on spheres [Heft],[Oliv]).*

Proof. If f is not surjective, then the degree of f is zero by Remark 2.5. Thus, we may assume that f is surjective and we endeavor to construct a homotopy inverse to f. Set $\varepsilon = 2 - \mathrm{dil}(f) > 0$ and note that if B is a ball of radius ε in S^n, then its inverse image $f^{-1}(B)$ is contained in an open hemisphere. Indeed, let y be the center of B, let y' be the point of S^n diametrically opposed to y, and fix an element x' of $f^{-1}(y')$. If $z \in f^{-1}(B)$, then

$$d(x',z) \geq \frac{1}{2-\varepsilon}\, d(y', f(z)) > \frac{\pi - \varepsilon}{2-\varepsilon} \geq \pi/2,$$



which shows that $f^{-1}(B)$ is contained in the open hemisphere opposite x'.

First suppose $n = 2$. By the preceding paragraph, the geodesic simplex formed by three points of $f^{-1}(B)$ is therefore well-defined. Given a triangulation of S^2, each of whose simplices are geodesic and contained in a ball of radius ε, we associate with each vertex v a point $g(v) \in f^{-1}(v)$. Next, we extend g to a continuous map $S^2 \to S^2$ by requiring that on each triangle with vertices v_1, v_2, v_3, the extension g equals the unique linear map whose image is the geodesic simplex having vertices $g(v_1), g(v_2), g(v_3)$. To complete the construction, we now show that $g \circ f$ is homotopic to the identity. Given $x \in S^2$, let T be a triangle containing $f(x)$ and let H be an open hemisphere containing $f^{-1}(T)$. By definition, each vertex of $g(T)$ lies in $f^{-1}(T) \subset H$, and, since g is linear and H is convex, $g(T) \subset H$ so that $g \circ f(x) \in H$. This shows that $d(x, g \circ f(x)) < \pi$. Since this inequality holds for each $x \in S^2$, a standard argument shows that $g \circ f$ is homotopic to the identity.

For $n > 2$, we proceed similarly, replacing the three points above by $n + 1$ points and by considering their corresponding spherical n simplex.

2.10. Remark: The hypothesis that $\operatorname{dil}(f) < 2$ cannot be relaxed to $\operatorname{dil}(f) \leq 2$ since, for example, there exists a map $f : S^3 \to S^3$ having dilatation 2 and degree 4. Indeed, let $z = (re^{i\theta}, \rho e^{i\varphi})$, where $r^2 + \rho^2 = 1$, and denote a point of S^3. Define $f(z) = (re^{2i\theta}, \rho e^{2i\varphi})$, so that f is the composition of the maps $z \mapsto (re^{2i\theta}, \rho e^{i\varphi})$ and $z \mapsto (re^{i\theta}, \rho e^{2i\varphi})$, which, as suspensions of the map $z \mapsto z^2$ of $S^1 \to S^1$, both have degree 2. Thus, f has degree 4. Moreover, we have $dz^2 = dr^2 + r^2 d\theta^2 + d\rho^2 + \rho^2 d\varphi^2$, and $d(f(z))^2 = dr^2 + 4r^2 d\theta^2 + d\rho^2 + 4\rho^2 d\varphi^2 \leq 4dz^2$, so that $\operatorname{dil}_z(f) \leq 2$ for each point z. Since S^3 is a path metric space, we conclude by Remark 1.8.bis that $\operatorname{dil}(f) \leq 2$; Proposition 2.9 therefore implies that $\operatorname{dil}(f) = 2$.

$2.10\frac{1}{2}_+$ Remarks and Exercises: (*a*) Probably it is not hard to classify maps $f : S^n \to S^n$ of maximal possible degree with $\mathrm{dil}(f) = 2$, and also with $\mathrm{dil}(f) = 2 + \varepsilon$.

(*b*) Let $f : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be a 1-Lipschitz λ-co-Lipschitz map with $\lambda < 2$. Show that f is one-to-one. Study 1-Lipschitz 2-co-Lipschitz maps $f : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ of degree ≥ 2.

In general, a map $f : S^n \to S^n$ with a given dilatation cannot have an arbitrary degree.

2.11. Proposition:

1. *For any* $f : S^n \to S^n$, *we have* $|\deg(f)| \leq (\mathrm{dil}(f))^n$.

2. *For each* n, *there exists a constant* c_n *with* $0 < c_n < 1$ *and mappings* $f : S^n \to S^n$ *with arbitrarily large degree satisfying the inequality* $|\deg(f)| > c_n(\mathrm{dil}(f))^n$, *i.e.,*

$$0 < c_n \leq \limsup_{\deg(f) \to \infty} \frac{|\deg(f)|}{(\mathrm{dil}(f))^n} \leq 1.$$

Proof. We first prove (1) in the case when f is C^1. Let ω be a normalized volume form on S^n, i.e., $\int_{S^n} \omega = 1$. By our earlier remarks, we then have $\deg(f) = \int_{S^n} f^*\omega$ (see [Godb], p. 219), and

$$|\deg(f)| \leq \int_{S^n} \|f^*\omega\| \leq \left(\sup_{x \in S^n} \|Df_x\| \right)^n \int_{S^n} \omega \leq (\mathrm{dil}(f))^n.$$

In the general case, we have that if $\mathrm{dil}(f) < \infty$, then there exists for each $\varepsilon > 0$, a C^1 map $f_\varepsilon : S^n \to S^n$ (obtained from f by convolution, for example) such that for each $x \in S^n$, $d(f(x), f_\varepsilon(x)) < \varepsilon$ and $\mathrm{dil}(f_\varepsilon) \leq \mathrm{dil}(f) + \varepsilon$. If $\varepsilon \leq \pi$, then f is homotopic to f_ε, so that $\deg(f) = \deg(f_\varepsilon) \leq (\mathrm{dil}(f_\varepsilon))^n \leq (\mathrm{dil}(f) + \varepsilon)^n$. Since this inequality holds for all $\varepsilon > 0$, assertion (1) follows.

(2) We first construct a map f_0 from the n-ball into S^n having degree 1 at each point. Fix a point $p \in S^n$ and let $B^n(r)$ be the ball of radius r centered at the origin of the tangent space to S^n at p. The mapping $f_0 : B^n(\pi) \to S^n$ is then taken to be the restriction of the exponential mapping. By composing with the dilation $B^n(r) \to B^n(\pi)$ on the ball of radius r, we obtain a map $f_0(r) : B^n(r) \to S^n$ with dilatation π/r, degree 1 at each point, and which maps $\partial B^n(r)$ onto p', the point of S^n diametrically opposed to p.

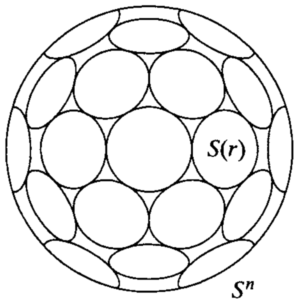
Recall that, even if one cannot pack all of S^n by small balls of the same radius, there exists a *filling ratio* $c'_n > 0$. In other words, if $S(r)$ is the union of a maximal disjoint collection of radius r balls having cardinality $N(r)$, then

$$\lim_{r \to 0} \frac{\operatorname{vol}(S(r))}{\operatorname{vol}(S^n)} = c'_n,$$

so that

$$N(r) = \frac{\operatorname{vol}(S(r))}{\operatorname{vol}(B_{S^n}(r))},$$

which is asymptotic to $c'_n \operatorname{vol}(S^n)/\operatorname{vol}(B^n)r^n$ as $r \to 0$, where B^n denotes the unit ball in $\mathbb{R}^n$.



We define a map $f_r : S^n \to S^n$ by first isometrically mapping each small ball of $S(r)$ onto $B^n(r)$ and composing with $f_0(r) : B^n(r) \to S^n$, then extending by $f_r = p'$ on the remainder of S^n. Then $\deg(f_r) = N(r)$ tends to infinity as $r \to 0$ and $\operatorname{dil}(f) = \pi/r$, so that $\deg(f_r) \sim c_n(\operatorname{dil}(f_r))^n$ with

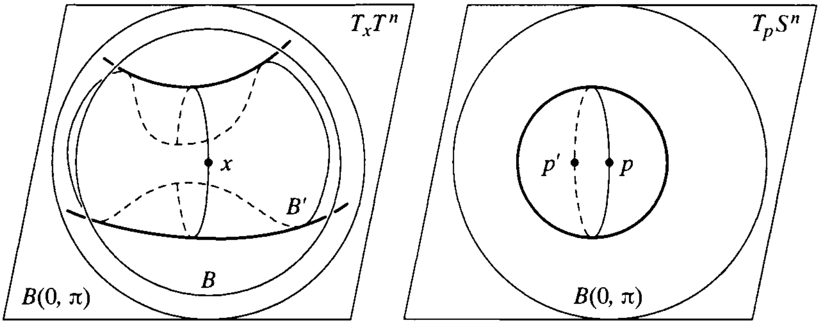
$$c_n = \frac{c'_n \operatorname{vol}(S^n)}{\pi^n \operatorname{vol}(B^n)}.$$

Remark: The preceding result is asymptotic, in that it gives a relationship between degree and dilatation when the degree is large, and it will later be improved in Theorem 2.18. The question as to the existence of mappings having a *fixed* degree and small dilatation is no doubt very complicated.

Maps of flat tori to S^n. Let T^n be a flat torus and let ℓ be the shortest length of a closed curve in T^n that is not null-homotopic (thus, $\ell/2$ is the injectivity radius of T^n).

2.12. Proposition: *There exists a map $f : T^n \to (S^n, can)$ with dilatation at most 1 and nonzero degree if and only if $\ell \geq 2\pi$.*

Proof. For each $x \in T^n$, the exponential map defines a diffeomorphism from the ball $B = B^{T_xT^n}(0, \ell/2)$ onto an open set B' in T^n having dilatation 1 (as does its inverse, since the mapping is a local isometry).



We now compose $\exp^{-1} : B' \to B$ with a dilation $B \to B^{T_xT^n}(0, \pi)$, the isometry $B^{T_xT^n}(0, \pi) \to B^{T_pS^n}(0, \pi)$, and the exponential map $\exp : B^{T_pS^n}(0, \pi) \to S^n$, for $p = f(x)$. The result is a map $f : B' \to S^n$ having degree 1 and dilatation $2\pi/\ell$, and which satisfies $\lim_{y \to \partial B'} f(y) = p'$. We can extend f to all of T^n by setting $f(T^n \setminus B') = p'$. Then $\deg(f) = 1$ and $\operatorname{dil}(f) = 2\pi/\ell \leq 1$. If $\ell \geq 2\pi$, then $\operatorname{dil}(f) \leq 1$, as desired.

Conversely, we now suppose that $\ell < 2\pi$. Let $f : T^n \to S^n$ be a mapping with dilatation 1; we will show that f is homotopic to 0. Since T^n is flat, we may assume that $T^n = T^{n-1} \times S^1$, where each $x \times S^1$ is a curve of length $\ell < 2\pi$. Let $f_0 : T^n \to S^n$ be the map given by $f_0(t, s) = f(t, 1)$. Now, the curve $S^1 \to S^n$ defined as $s \mapsto f(t, s)$ has length $< 2\pi$, and $d(f(t, s), f(t, 1)) < \pi$ for each $(t, s) \in T^n$ so that $d(f, f_0) < \pi$. Thus, f is homotopic to f_0, which is not surjective, and so f has zero degree by Remark 2.5. (Specifically, the homotopy can be constructed explicitly by connecting each pair $f(x), f_0(x)$ by the unique minimizing geodesic joining them).

$2.12\frac{1}{2}{}_+$ Remark: The latter argument applies to an arbitrary (say, compact for safety) metric space X admitting a topological fibration $p : X \to B$, all of whose fibers are circles of length $< 2\pi$. Namely, every map $f : X \to S^n$ with $\operatorname{dil}(f) \leq 1$ is homotopic to a map f_1 which factors through another map $f_0 : B \to S^n$, i.e., $f_1 = f_0 \circ p$. Now let us suppose that the fibers $p^{-1}(b) \subset X$ are k-dimensional and that their k-dimensional Hausdorff measures are strictly smaller than that of the unit sphere S^k. Here, one wants to homotope each map $f : X \to S^n$ with $\operatorname{dil}(f) \leq 1$ to f_1, sending each fiber $p^{-1}(b)$ onto a subset of topological (or Hausdorff) dimension $< k$ in S^n. Probably one can handle the case in which the fibers are (sufficiently smooth) surfaces using the argument of §3.6C–C′ of $[\mathrm{Gro}]_{\mathrm{CC}}$. For $k \geq 3$, some results are available for *smooth* fibrations $p: X \to B$. In these cases, our f can be viewed as a map of B into the space of k-cycles in S^n, namely $b \mapsto f(p^{-1}(b))$, which is continuous with respect to the so-called *flat norm.* Then, by Almgren–Morse theory, the map $B \to \mathrm{cycles}(S^n)$ is *contractible* for $\mathrm{mes}_k(f(p^{-1}(b))) < \mathrm{mes}_k(S^k)$, since every minimal k-dimensional subvariety in S^n has $\mathrm{mes}_k \geq \mathrm{mes}_k(S^k)$. In particular, if $\dim(X) = n$, then $\deg(f) = 0$, and so f is contractible (see Appendix 1 in $[\mathrm{Gro}]_{\mathrm{FRM}}$). It is unclear what happens to general smooth maps $p : X \to B$ with small $\mathrm{mes}_k(f(p^{-1}(b)))$ (where the difficulties disappear for sufficiently generic maps p for which $p^{-1}(b) \in \mathrm{cycles}(X)$ is continuous in b). Finally, one wishes to replace S^n by a more general Riemannian manifold or metric space Y and to evaluate the homotopy complexity of maps $f : X \to Y$ in terms of $\sup_{b \in B} \mathrm{mes}_k(f(p^{-1}(b)))$. An especially interesting Y in this regard is the unit sphere in ℓ_∞^n-space, i.e., $\mathbb{R}^n$ equipped with the norm

$$\|x\|_{\ell_\infty} = \sup_{1 \leq i \leq n} |x_i|$$

(compare $[\mathrm{Gro}]_{\mathrm{FRM}}$).

2.13. Problem: What condition on the metric of the torus T^n guarantees that there exist mappings $f: T^n \to S^n$ with degree d and dilatation 1? We can pose the same question in a slightly different form: Given a metric on T^n, for which values of d do there exist mappings $T^n \to S^n$ with degree d and dilatation less than D?

C. Homotopy counting Lipschitz maps

Given two *pointed* metric spaces $(V, v), (W, w)$, we denote by $\#D$ the number of homotopy classes of maps from (V, v) to (W, w) that contain at least one map having dilatation $\leq D$.

We will use the notation $[(V, v), (W, w)]$ for the set of equivalence classes of continuous maps from V to W under the relation of homotopy relative to v and w.

2.14. Definition: Let X be a precompact path metric space. For $\varepsilon > 0$, the *ε-capacity* of X, denoted $\mathrm{Cap}_\varepsilon(X)$, is the minimum number of radius-ε balls required to cover X. In other words, $\mathrm{Cap}_\varepsilon(X)$ is the minimum number of points in an ε-net in X. (An ε-net is a subset N of X such that for each $x \in X$, $d(x, N) < \varepsilon$.)

This definition enables us to state the following very general theorem.

2.15. Proposition: *Let $(X, x_0), (Y, y_0)$ be precompact, pointed path metric spaces such that Y has the property that for uniformly "close" mappings from X to Y are homotopic (this is the case if Y is a manifold or a finite polyhedron in the sense of [Span]). Then there exist two constants c, c' depending only on Y such that*

$$\#(D) \leq c^{\mathrm{Cap}_{1/c'D}(X)}.$$

Proof. By hypothesis, there exists $\delta > 0$ independent of X such that two maps $f_0, f_1 : X \to Y$ with $d(f_0(x), f_1(x)) < \delta$ for each $x \in X$ are necessarily homotopic. Let R_Y be a $(\delta/4)$-net in Y, set $\varepsilon = \delta/4D$, and let R_X be an ε-net in X. If $f : X \to Y$ satisfies $\mathrm{dil}(f) \leq D$, then f maps a ball of radius ε in X into a ball of radius $\delta/4$ in Y.

We will first prove that $\#(D)$ is finite. Suppose that $f : X \to Y$ satisfies $\mathrm{dil}(f) \leq D$ and let $\hat{f} : R_X \to \mathcal{P}(R_Y)$ be the mapping defined by $x \mapsto \{y \in R_Y : d(y, f(x)) < \delta/4\}$. Suppose that $g : X \to Y$ satisfies $\mathrm{dil}(g) \leq D$ and $\hat{f}(x) \cap \hat{g}(x) \neq \emptyset$ for each $x \in X$. Then for all $x \in R_X$, there is some $y \in R_Y$ such that $d(y, f(x)) \leq \delta/4$ and $d(y, g(x)) < \delta/4$. If $z \in B_X(x, \varepsilon)$, then $d(f(z), f(x)) < \delta/4$, and $d(g(z), g(x)) < \delta/4$, so that $d(f(z), g(z)) < \delta$. Since the balls $B_X(x, \varepsilon)$ for $x \in R_X$ cover X, we conclude that f is homotopic to g; in particular, we have just shown that if $\hat{f} = \hat{g}$, then f is homotopic to g. This proves that $\#(D) \leq (\mathrm{card}\, \mathcal{P}(R_Y))^{\mathrm{card}(R_X)}$.

Next, we choose a representative f_α for each homotopy class α, and for each $x \in R_X$ we fix an element of $f_\alpha(x)$ denoted by $\hat{\alpha}(x)$, so that $\hat{\alpha}$ is a mapping from R_X to R_Y. If $\hat{\alpha} = \hat{\beta}$, then $f_\alpha(x) \cap f_\beta(x) \neq \emptyset$ for all $x \in R_X$. Thus, f_α is homotopic to f_β and so $\alpha = \beta$. Thus the map $\alpha \mapsto \hat{\alpha}$ is injective, and $\#(D) \leq (\mathrm{card}(R_Y))^{\mathrm{card}(R_X)}$. By choosing the net R_X to have the minimum number of points, $\mathrm{Cap}_\varepsilon(X)$, we obtain

$$\#(D) \leq c^{\mathrm{Cap}_{1/c'D}(X)},$$

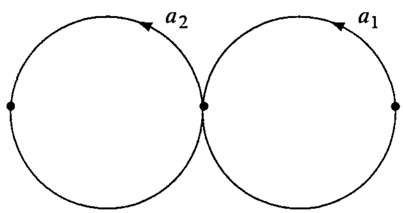
where $c = \mathrm{card}(R_Y)$ and $c' = 4/\delta$.

2.16. Corollary: *If X is a compact Riemannian n-manifold, then Proposition 2.15 implies that $\#(D) \leq c^{(c'D)^n}$.*

Proof. Here, the question is local, and we know that each point of X admits a neighborhood that is almost isometric to a Euclidean ball, for which one has the estimate $\mathrm{Cap}_\varepsilon \simeq C \cdot \varepsilon^{-n}$ as $\varepsilon \to 0$. Indeed, a ball of radius ε has volume equal to $\mathrm{vol}(B(0,1)) \cdot \varepsilon^n$; if N balls of radius ε cover the unit ball, then necessarily $N \geq \mathrm{vol}(B(0,1))/\mathrm{vol}(B(0,\varepsilon)) = \varepsilon^{-n}$. Conversely, it is easy to construct a collection of small balls that cover the unit ball: One simply takes their centers to be all points whose coordinates are multiples of $2\varepsilon/\delta n$, so that the number of balls is on the order of $\mathrm{vol}(B(0,1))\varepsilon^{-n}$.

2.17. Examples:

1. If X is a standard circle and Y is a bouquet (join) of two standard circles, then $\#(D) = 4 \cdot 3^{D-1}$. Indeed, $[(S^1, p), (S^1 \vee S^1)]$ is the free group on two generators, a_1, a_2. We obtain a word of length k in this group by attaching to a word x of length $k-1$ one of the letters $a_1, a_2, a_1^{-1}, a_2^{-1}$ distinct from the last letter of x. There are three choices at each stage, except for the first, and so there are $4 \cdot 3^{k-1}$ words of length $\leq k$. Thus, the smallest possible dilatation of a map $X \to Y$ represented by a word of length k is exactly k. Proposition 2.15 gives the bound 3^{8D} since $\delta = \pi$, $\mathrm{Cap}_\varepsilon(S^1) = 2\pi/\varepsilon$, and one can take R_Y to be the net consisting of three points indicated in the figure.



2_+. Let X be a closed orientable n-manifold and let Y be the bouquet of the standard n-sphere S^n and the k-torus $S^1 \times \cdots \times S^1$ (k factors). If $k \geq n$, then $\#(D) \geq 2^{c'D^n}$ for some $c' = c'_n(X) > 0$. In fact, for each $D \geq 0$, there (obviously) exists a map $\tilde{f}_0 : X \to \mathbb{R}^k$ serving as the universal covering of the k-torus, such that $\mathrm{dil}(\tilde{f}_0) \leq D$ and whose image $\tilde{f}_0(X) \subset \mathbb{R}^k$ covers at least cD^n standard unit cubes in $\mathbb{R}^k$, each of which is a fundamental domain for the $\mathbb{Z}^k$-action on $\mathbb{R}^k$ giving the k-torus $\mathbb{R}^k/\mathbb{Z}^k$. Since the universal covering $\tilde{Y}$ of Y equals $\mathbb{R}^k$ with n-spheres attached to the centers of these cubes, one can "bubble" $\tilde{f}_0$ independently in at least cD^n points, thus generating 2^{cD^n} mutually nonhomotopic maps $\tilde{f} : X \to \tilde{Y}$ of dilatation D' slightly greater than D. These project to the required $2^{c'D^n}$ maps $f : X \to Y$ where $c' = c(D/D')^n$.

For example, if $X = S^n$, then a rough estimate shows that $c' \geq n^{-10n}$; say for $n = 2$,

$$2^{0.0003D^2} \leq \#(D) \leq 4^{\pi^2 D^2}.$$

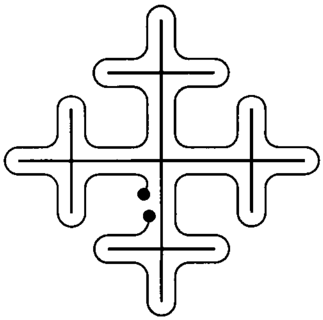
But, if $k < n$, then $\#(D)$ is on the order of D^{D^k}, namely

$$(c_0 D)^{(c_0 D)^k} \leq \#(D) \leq (c_1 D)^{(c_1 D)^k} \qquad (*)$$

for some positive c_0, c_1 depending on n and X. Indeed, our map $\tilde{f}_0\colon X \to \mathbb{R}^k$ covers at most $c_1' D^k$ cubes, and each bubbling at a given point $p \in \mathbb{R}^k$ covers the sphere S^n attached at p with degree at most $c_1 D^n$ if we do not allow dilatation $\geq D$. This gives the upper bound in $(*)$, while the lower bound follows by an actual display of independent bubblings at $\simeq c_0 D^k$ points in $\mathbb{R}^k$ of degrees $\leq c_0 D$.

Notice that the (approximate) upper bound $\#(D) \lesssim D^{D^k}$ remains valid (and obvious) for any compact manifold Y whose fundamental group $\pi_1(Y)$ has polynomial growth of degree $\leq k$ (see 5.B) and whose universal covering $\tilde{Y}$ is $(n-1)$-connected. (If $\tilde{Y}$ fails this connectivity requirement, then one can attempt to apply the minimal model argument of 7.B to large balls $B(D) \subset Y$ and hope for a nontrivial upper bound on $\#(D)$ if $\pi_i(Y)$ are finitely generated over $\pi_1(Y)$, for instance.)

3_+. Let $Y = S^n \vee S^1 \vee S^1$, so that $\tilde{Y}$ equals the regular 3-adic tree Tre with S^n attached at every vertex. It is not hard to produce maps of the unit n-ball $\tilde{f}_D : B^n \to$ Tre, with dilatations $D = 1, 2, \ldots$, such that $\tilde{f}_D(B) \subset$ Tre contains at least cD^n vertices of Tre as shown in the picture for $n = 1$.



Bubbling $\tilde{f}_D$ at these vertices generates 2^{D^n} mutually nonhomotopic maps $f\colon X \to Y$ of dilatation $\approx D$ for each closed oriented n-manifold X.

Another example worth looking at is $Y = S^n \vee N$, where N is a k-dimensional nilmanifold whose (nilpotent) fundamental group has (polynomial) growth of degree $m > k$ (i.e., is nonabelian). We want to evaluate the crudest characteristic of $\#(D)$, the exponent

$$\delta = \delta(D) = (\log\log \#(D))/\log(D)$$

as $D \to \infty$. It is clear from the above that $\limsup_{D\to\infty} \delta \leq \min(n, m)$, but we do not know whether this inequality is strict for some N, or whether $\delta(D) \to \min(m, n)$ as $D \to \infty$ for all N. It is clear that $\delta(D) \to n$ if the nilpotent group $\tilde{N}$ covering N contains an abelian subgroup A of rank n, since the unit ball B^n can be mapped to $A \subset \tilde{N}$ with dilatation D and having image of ε capacity $\simeq D^n$. On the other hand, such maps $B^n \to \tilde{N}$ probably do *not* exist unless the *homogeneous group* $\tilde{N}_\infty$, which is the *Hausdorff limit* $\lim_{\varepsilon \to 0} \varepsilon \tilde{N}$ (where $\varepsilon(\tilde{N}, \mathrm{dist}) = (\tilde{N}, \varepsilon\, \mathrm{dist})$, see $3.18\frac{1}{2}{}_+$) contains an abelian subgroup A of rank n.

The following sketchy proof of this fact in a special case emerged in conversations with S. Semmes: Scale $\tilde{N}$ by $\varepsilon = D^{-1}$ and pass to a (sub)-limit map $f_\infty : B^n \to \tilde{N}_\infty$ as $D \to \infty$. This f_∞ has dilatation ≤ 1 with respect to the limit (as $D \to \infty$) Carnot–Caratheodory metric dist_∞ in $\tilde{N}_\infty$. (Recall that $\tilde{N}_\infty$ admits a dilatation which scales dist_∞ by a constant, see [Gro]$_{CC}$). By the Pansu–Rademacher theorem (see [Pan]$_{CC}$), this f_∞ is a.e. differentiable with $\mathrm{rank}(\mathcal{D} f_\infty) \leq \max\mathrm{rank}(A)$ and so $\mathrm{mes}_n(f_\infty(B^n)) = 0$ if $\max\mathrm{rank}(A) < n$, where max rank is the maximum of the ranks of abelian subgroups $A \subset \tilde{N}_\infty$. Unfortunately, it does not imply (at least not in an obvious way) the desired relation

$$\frac{\mathrm{Cap}_\varepsilon(f_D(B^n))}{D^n} \to 0 \qquad (\star)$$

for a fixed $\varepsilon > 0$ and $D \to \infty$. Yet, if $\tilde{N}_\infty$ is a *nonabelian* group of polynomial growth of degree $\leq n$, then the rank inequality $\mathrm{rank}(\mathcal{D} f_\infty) < n$ does imply $(\star)$ by an argument similar to one used in 2.46.

Open questions$_+$. (*a*) Can one obtain a quantitative version of the preceding statement by finding an effective bound on $\mathrm{Cap}_\varepsilon(f_D(B^n))$? For example, does there exist $\nu > 0$ such that

$$\frac{\mathrm{Cap}_\varepsilon(f_D(B^n))}{D^{n-\nu}} \to 0$$

as $D \to \infty$. (If so, then a simple counting argument shows that $\delta(D) < n$ for these N.)

(*b*) More generally, let Y be a compact manifold (or finite polyhedron) with fundamental group Γ. We want to evaluate the ε-capacity (or the n-dimensional measure) of the images of maps $f_D : B^n \to \tilde{Y}$ with dilatation D for $D \to \infty$, where $\tilde{Y}$ denotes the universal covering of Y. In particular, we want to decide whether the ratio $\mathrm{Cap}_\varepsilon(f_D(B^n))/D^n$ necessarily tends to zero as $D \to \infty$.

The same question can be rephrased in terms of a single Lipschitz map $f : \mathbb{R}^n \to \tilde{Y}$ and the relation between the volumes of concentric R-balls in $\mathbb{R}^n$ and the volumes, or, better, the ε-capacities of their images. In particular, we ask whether $(\mathrm{Cap}_\varepsilon f(B(R)))/R^n \to 0$ as $R \to \infty$.

Alternatively, a combinatorial version of this question can be described by considering a Lipschitz map $f: \mathbb{Z}^n \to \Gamma$, where both groups are given their respective word metrics (see 3.24_+), and by then comparing the cardinalities of the balls in $\mathbb{Z}^n$ to those of their f-images. Specifically, we want to know when there exists a Lipschitz map $\mathbb{Z}^n \to \Gamma$ such that for all balls $B \subset \mathbb{Z}^n$,

$$\operatorname{card}(f(B)) \geq c \cdot \operatorname{card}(B)$$

for some $c = c(f) > 0$.

(c) The latter setup can be further generalized by replacing $\mathbb{Z}^n$ with an arbitrary finitely presented (or just finitely generated) group Γ_0 and then studying the rate of decrease of the cardinalities of balls in Γ_0 under Lip schitz mappings $\Gamma_0 \to \Gamma$. For example, we want to know if, for given Γ_0 and Γ, there exists a Lipschitz map such that all balls $B \subset \Gamma_0$ (or just the balls B around a fixed point $\gamma_0 \in \Gamma_0$) satisfy

$$\operatorname{card}(f(B)) \geq \operatorname{const} \operatorname{card}(B). \qquad (*)$$

(Probably the existence of f satisfying $(*)$ implies that $\operatorname{exprank}(\Gamma) \geq \operatorname{exprank}(\Gamma_0)$ by a modification of the corresponding argument for *uniform embeddings* in 7.E2 of $[\mathrm{Gro}]_{\mathrm{AI}}$. For example, the product of two free groups $F_2 \times F_2$ should not admit such a mapping into F_2.)

(c') Next, we can consider a discrete co-compact (i.e., with compact quotient) isometric action of Γ_0 on some metric space $\tilde{X}$ (playing the role of $\mathbb{R}^n$ for $\Gamma_0 = \mathbb{Z}^n$) and study Lipschitz maps $\tilde{X} \to \tilde{Y}$. Here, we are again interested in a possible decrease of the ε-capacities of balls under such maps.

(c'') Returning to our homotopy counting problem, we can consider concentric balls $B(R)$ in $\tilde{X}$ and study homotopy classes of maps $f_R: B(R) \to \tilde{Y}$, where the f_R are constant on the boundaries $\partial B(R)$ and have $\operatorname{dil}(f_R) \leq$ const for some constant independent of R. This looks hard in general, and one might simplify the situation by assuming that the $B(R)$ are homeomorphic to the standard ball B^n and that $\tilde{Y}$ is $(n-1)$-connected. (If not, one can count maps $B(R) \to \tilde{Y}$ up to *homological* equivalence, which is easier.)

If the group Γ_0 above is nilpotent, then the problem can be reduced to the study of Lipschitz maps between *compact* spaces, $X \to Y$, and where X is now given a Carnot–Caratheodory metric. In fact, much of what we have said in 1_+–3_+ extends to Carnot–Caratheodory metrics with only minor adjustments (compare $[\mathrm{Gro}]_{\mathrm{CC}}$). Here, one may be tempted to allow Carnot–Caratheodory metrics in Y as well, but this brings in serious local problems (see $[\mathrm{Gro}]_{\mathrm{CC}}$).

D. Dilatation of sphere-valued mappings

The goal of this section is to prove the following theorem:

2.18. Theorem: *Let V be a compact, connected, orientable n-manifold with (possibly empty) boundary ∂V and let S^n be the n-sphere equipped with an arbitrary Riemannian metric g. Then the number $\#(D)$ of homotopy classes of maps $(V, \partial V) \to (S^n, s_0)$ of dilatation $\leq D$ is asymptotic to $c_g D^n \operatorname{vol}(V)$, i.e.,*

$$\#(D)D^{-n} \to c_g \operatorname{vol}(V)$$

as $D \to \infty$, where $c_g > 0$ is some constant that depends on g but is independent of V and the choice of marking $s_0 \in S^n$.

2.19. Definition: Let (X, x_0) be a metric space, and let B^n be the standard unit ball in $\mathbb{R}^n$. For $\alpha \in \pi_n(X, x_0)$, considered as a homotopy class of mappings $(B^n, \partial B^n) \to (X, x_0)$, we define

$$\|\alpha\| = \left(\inf_{f \in \alpha} \operatorname{dil}(f)\right)^n \operatorname{vol}(B^n).$$

If (V, v_0) is a pointed, compact Riemannian manifold, we similarly define a function $[(V, v_0), (X, x_0)] \to \mathbb{R}_+$ by $\alpha \mapsto \|\alpha\|$, where

$$\|\alpha\| = \operatorname{vol}(V) \left(\inf_{f \in \alpha} \operatorname{dil}(f)\right)^n.$$

2.20. Remark 1: If $n = 1$ and $\alpha \in \pi_1(X, x_0)$, then $\|\alpha\|$ is the length of a shortest curve representing α, so that $\|\cdot\|$ defines a norm on the group $\pi_1(X, x_0)$.

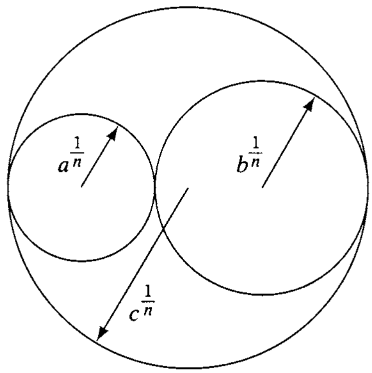
Indeed, by a change of parametrization (i.e., replacing a representative f of α by $f \circ h$, where h is a homeomorphism of $B^1 = [-1, 1]$ isotopic to the identity), we may assume that the local dilatation of f is constant and equal to $1/2 \operatorname{length}(f)$. This reduces the dilatation without changing the homotopy class α and thus proves the assertion.

The set $\pi_n(X, x_0)$ has a group structure defined by "juxtaposition" of maps defined on balls that satisfies

2.21. For $\gamma, \delta \in \pi_n(X, x_0)$,

$$\|\gamma\delta\|^{1/n} \leq \|\gamma\|^{1/n} + \|\delta\|^{1/n}.$$

Indeed, let a, b be fixed real numbers with $a > \|\gamma\|$ and $b > \|\delta\|$. Then the classes γ, δ can be represented by short maps f, g defined on balls of radius $a^{1/n}$ and $b^{1/n}$ whose boundaries are sent to the point x_0. By placing these balls side-to-side within a ball of radius $c^{1/n} = a^{1/n} + b^{1/n}$ and extending f, g to have the constant value x_0 in the large ball, we obtain a short map defined on a ball of radius $c^{1/n}$ that represents $\gamma\delta$. This proves $\|\gamma\delta\| > c$.



For $n = 1$, we can conclude that $\|\cdot\|$ is a seminorm on $\pi_1(X, x_0)$; if X is a locally compact, complete path metric space, then each homotopy class $\alpha \in \pi_1(X, x_0)$ admits a minimizing loop (Remark 1.12(b)), i.e., a loop whose length equals $\|\alpha\|$. It follows that $\alpha \neq 0$ implies $\|\alpha\| \neq 0$, and so $\|\cdot\|$ is a norm. If X admits a universal cover $\tilde{X}$, then the length structure induced by the projection $p\colon \tilde{X} \to X$ makes $\tilde{X}$ a locally compact, complete path metric space, and the deck transformations (see [Godb]) of $\tilde{X}$ are isometries. If $\tilde{x}_0 \in \tilde{X}$ lies in the fiber over x_0, then an element $\alpha \in \pi_1(X, x_0)$ induces an automorphism $\alpha\colon \tilde{X} \to \tilde{X}$ and $\|\alpha\| = d(\tilde{X}_0, \alpha\cdot\tilde{x}_0)$. On the other hand, there is no reason for $\|\cdot\|$ to be a seminorm on $\pi_n(X, x_0)$ for $n \geq 2$.

Remark 2_+: Take the unit sphere S^n marked at the north pole $s_+ \in S^n$ and define

$$\|\alpha\|' = (\inf \operatorname{dil}(f'))^n \operatorname{vol}(S^n)$$

where the infimum ranges over all maps $f'\colon (S^n, s_+) \to (X, x_0)$ representing α. *Then*

$$(\operatorname{vol}(B^n))\|\alpha\|' \leq (\operatorname{vol}(S^n))\|\alpha\| \leq \pi^n(\operatorname{vol}(B^n))\|\alpha\|'.$$

In fact, to bound $\|\alpha\|$ in terms of $\|\alpha\|'$, we first embed $B^n \subset \mathbb{R}^n_p = T_pS^n$ and then map it to S^n by $x \mapsto \exp(\pi x)$. The dilatation of this map, say $e : B^n \to S^n$, equals π, and so for each $f' : S^n \to X$, the composition

$f = f' \circ e \colon B \to X$ is homotopic to f' and satisfies $\operatorname{dil}(f) \le \pi \operatorname{dil}(f')$. Thus,

$$\|\alpha\| \le \pi^n \|\alpha'\| \frac{\operatorname{vol}(B^n)}{\operatorname{vol}(S^n)}.$$

Conversely, to bound $\|\alpha\|'$ by $\|\alpha\|$, we use the normal projection of the southern hemisphere $S^n_- \subset \mathbb{R}^{n+1}$ (defined by $x_{n+1} \le 0$) onto the unit ball B^n in $\mathbb{R}^n \subset \mathbb{R}^{n+1}$ (given by $x_{n+1} = 0$). This projection, say $\varphi \colon S^n_- \to B^n$, satisfies $\operatorname{dil}(\varphi) \le 1$, and so every $f : (B^n, \partial B^n) \to (X, x_0)$ gives rise to a map $f' \colon (S^n, s_+) \to (X, x_0)$ defined by

$$f'(s) = \begin{cases} x_0 & \text{for } s \in S^n_+ \\ (f \circ \varphi)(x) & \text{for } s \in S^n_- \end{cases}$$

and satisfying $\operatorname{dil}(f') \le \operatorname{dil}(f)$. Hence,

$$\|\alpha\|' \le \frac{\|\alpha\|(\operatorname{vol}(S^n))}{\operatorname{vol}(B^n)}.$$

$2.21\frac{1}{2}{}_+$ Idea of the proof of 2.18. Define

$$\|\alpha\|^{lim} = \liminf_{p \to \infty} \frac{\|\alpha^p\|}{p}$$

for all $\alpha \in \pi_n(X, x_0)$. If X is homeomorphic to S^n, then the group $\pi_n(X)$ is cyclic, generated by some α_0, and we claim that the number $\#(D)$ of homotopically distinct maps $V \to X \simeq S^n$ is asymptotic to $D^n \operatorname{vol}(V)/\|\alpha_0\|^{lim}$. The proof of this claim divides naturally into two parts:

(1) Showing that $\#(D) \succeq D^n \operatorname{vol}(V)/\|\alpha_0\|^{lim}$, where $A(D) \succeq B(D)$ stands for $A(D) \ge (1 - \varepsilon(D))B(D)$ where $\varepsilon(D) \to 0$ for $D \to \infty$.

(2) Establishing the opposite asymptotic inequality,

$$\#(D) \preceq D^n \operatorname{vol}(V)/\|\alpha_0\|^{lim}.$$

To prove (1), we have to generate sufficiently many maps $V \to X$ with dilatation $\le D$, starting with a certain collection of mappings $B^n \to X$. This is eventually done by taking, for each $D > 0$, a packing of V by metric balls B_j such that

(*a*) $\sum_j \operatorname{vol}(B_j) = (1 - \varepsilon) \operatorname{vol}(V)$, where $\varepsilon \to 0$ as $D \to \infty$.

(*b*) The radii r_j of the B_j satisfy $r_j \le \delta$, where $\delta \to 0$ as $D \to \infty$.

(*c*) For each value $r \in \{r_1, \dots, r_j, \dots\}$, there exist maps of the Euclidean ball $(B^n, \partial B^n) \to (S^n, s_0)$ of dilatation $\leq D/r$ and of degrees $\pm 1, \pm 2, \dots, \pm p$, such that

$$(D/r)^n \operatorname{vol}(B) \leq (1+\varepsilon)\, p\, \|\alpha\|^{lim},$$

where $\varepsilon \to 0$ as $D \to \infty$. (This implies in particular that the totality of these p as $D \to \infty$ belongs to a subsequence for which $\|\alpha_0^p\|/p \to \|\alpha_0\|^{lim}$.)

Such maps $B^n \to S^n$, composed with the rescaled inverse exponential maps $B_i(r_i) \to B^n$, give us, by a trivial computation, just enough mappings $V \to S^n$ constant outside $B_i(r_i)$ to satisfy (1) (compare with the proof of 2.11 and see the discussion below for a detailed argument).

To prove (2), we must reverse the preceding construction in order to transform maps $f\colon V \to S^n$ of dilatation D into maps $(B^n, \partial B^n) \to (S^n, s_0)$ whose dilatations are suitably controlled by D and which are of sufficiently high degree. For each D (where eventually $D \to \infty$) we again take a packing of V by $B_j(r_j)$ satisfying conditions (*a*) and (*b*) above and the following:

(c') The radii r_j satisfy $Dr_j \geq \triangle$, where $\triangle \to \infty$ as $D \to \infty$.

Then one can show (see below) that f can be homotoped to $f'\colon V \to S^n$ such that:

(*i*) f' maps the union $\bigcup_j \partial B_j$ to a single point $s_0 \in S^n$.

(*ii*) The dilatation of f' is only slightly greater than that of f, i.e.,

$$\operatorname{dil}(f') \leq (1+\nu)D,$$

where $\nu \to 0$ as $D \to \infty$. This f' gives rise to maps

$$(B^n, \partial B^n) \to (B_i, \partial B_i) \to (S^n, s_0),$$

among which one (obviously) finds those for which

$$\frac{\deg(g_i)}{(\operatorname{vol}(B^n))\operatorname{dil}(g_i)} \geq \frac{\deg(f)}{(\operatorname{vol}(V))D}.$$

Notice that while the proof of (1) applies to mappings into an arbitrary (X, x_0) and completely relies on the geometry of the source manifold V (specifically its property of being locally approximately Euclidean), the proof of (2) makes crucial use of the $(n-1)$-*connectivity* of the target sphere

S^n. This connectivity allows us not only to contract maps $S^{n-1} \to S^n$, but also to do it with negligible increase in dilatation over the course of the contraction.

Now we state the proposition encompassing (1) above.

2.22$_+$ Packing Proposition: *For every marked metric space (X, x_0), the sequence $\|\alpha^p\|/p$ always converges to $\|\alpha\|^{lim}$ for all $\alpha \in \pi_n(X, x_0)$. Moreover, if V is a triangulated n-dimensional Riemannian manifold of finite total volume, then for every $\alpha \in \pi_n(X, x_0)$, there exists a sequence of maps $f_p: V \to X$ with the following two properties:*

(i) *Every f_p is constant and equals $x_0 \in X$ on all $(n-1)$-simplices of the triangulation of V as well as on all but finitely many n-simplices. Furthermore, the map f represents α^p in $\pi_n(X, x_0)$. This means that the sum of the restrictions of f to the n-simplices Δ^n (homeomorphic to B^n) in V on which f is nonconstant, equals α^p in $\pi_n(X, x_0)$.*

(ii) $(\mathrm{dil}(f_p))^n/|p| \to \|\alpha\|^{lim} \operatorname{vol}(V)/\operatorname{vol}(B^n)$ *as $p \to \infty$.*

Throughout the proof of this proposition, we will use the notation $B(\omega)$ for the ball of *volume* ω in $\mathbb{R}^n$. Any class $\alpha \in \pi_n(X, x_0)$ defines a homotopy class of mappings $(B(\omega), \partial B(\omega)) \to (X, x_0)$ for each ω, and

$$\|\alpha\| = \inf \left\{ \omega : \begin{array}{l} \alpha \text{ admits a representative } (B(\omega), \partial B(\omega)) \to (X, x_0) \\ \text{having dilatation } < 1 \end{array} \right\}.$$

The idea of the proof is now that, given classes $\alpha_1, \dots, \alpha_k$ having "volumes" $\|\alpha_1\|, \dots, \|\alpha_k\|$, we will recover a ball of volume slightly greater than $\sum_{i=1}^k \|\alpha_i\|$ from the balls of volume $\|\alpha_i\|$. By juxtaposing the representatives of the α_i, we obtain a representative of $\prod_{i=1}^k \alpha_i$, and we will then prove that $\|\prod_{i=1}^k \alpha_i\|$ is close to $\sum_{i=1}^k \|\alpha_i\|$.

The first part of the proof describes how to pack a large ball with small balls of a given volume.

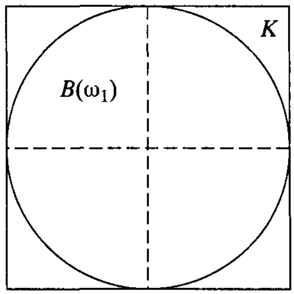
Lemma: *Let c_n be the volume of the unit ball in $\mathbb{R}^n$ and define $\lambda = 1 - 2^{-n-1}c_n$. Let $\omega_1 \geq \cdots \geq \omega_k$ be positive real numbers such that $\omega_j \geq 100\omega_{j+1}$. Then there exists a cube $K \subset \mathbb{R}^n$ and a union S of disjoint balls in K having volumes in $\{\omega_1, \dots, \omega_k\}$ such that* $\operatorname{vol}(K \setminus S) \leq \lambda^k \operatorname{vol}(K)$.

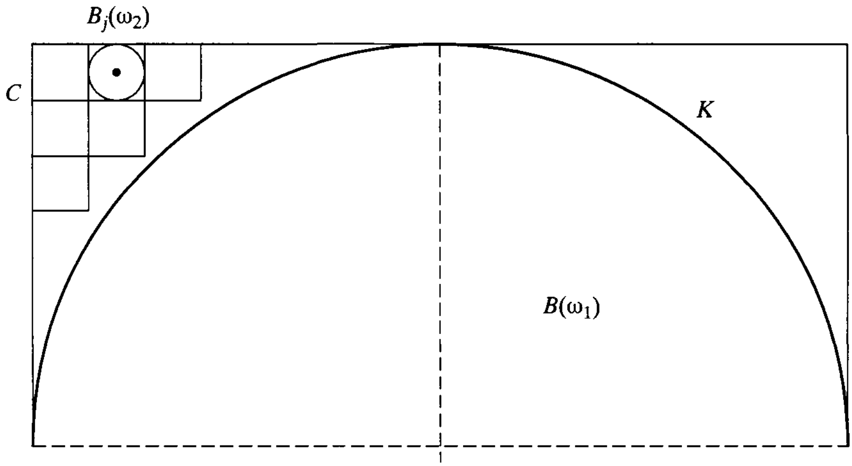
Proof. Let K be the n-cube circumscribed around $B(\omega_1)$ in $\mathbb{R}^n$, so that $\omega_1 = 2^{-n}c_n \operatorname{vol}(K)$ and

$$\frac{\operatorname{vol}(K \setminus B(\omega_1))}{\operatorname{vol}(K)} = (1 - 2^{-n}c_n) \leq \lambda.$$

Let K_2 be the n-cube circumscribing $B(\omega_2)$ in $\mathbb{R}^n$. Since $\operatorname{vol}(K_2)/\operatorname{vol}(K) = \omega_2/\omega_1 \leq 1/100$, the set C consisting of the union of the translates of K_2 represented in the figure covers almost all of $K \setminus B(\omega_1)$ or, more precisely,

$$\operatorname{vol}(C) \geq \frac{1}{2}\operatorname{vol}(K \setminus B(\omega_1)).$$





Next, we place a small ball $B_j(\omega_2)$ within each small cube so that

$$\operatorname{vol}\left(\bigcup_j B_j(\omega_2)\right) = 2^{-n} c_n \operatorname{vol}(C)$$

whence

$$\operatorname{vol}\left(\bigcup_j B_j(\omega_2)\right) = 2^{-n-1} c_n \operatorname{vol}(K \setminus B(\omega_1)).$$

In other words,

$$\operatorname{vol}\left(K \setminus \left(B(\omega_1) \cup \bigcup_j B_j(\omega_2)\right)\right) \leq \lambda \operatorname{vol}(K \setminus B(\omega_1)) \leq \lambda^2 \operatorname{vol}(K).$$

2.23. Lemma: *Let $(\omega_j)_{j\geq 1}$ be a sequence of positive real numbers satisfying $\omega_j \geq 100\omega_{j-1}$. For each $\varepsilon > 0$ and sufficiently large ω, there exists within $B(\omega)$ a union S of balls whose volumes are among the ω_j and*

$$\operatorname{vol}(B(\omega) \setminus S) \leq \varepsilon \operatorname{vol}(B(\omega)).$$

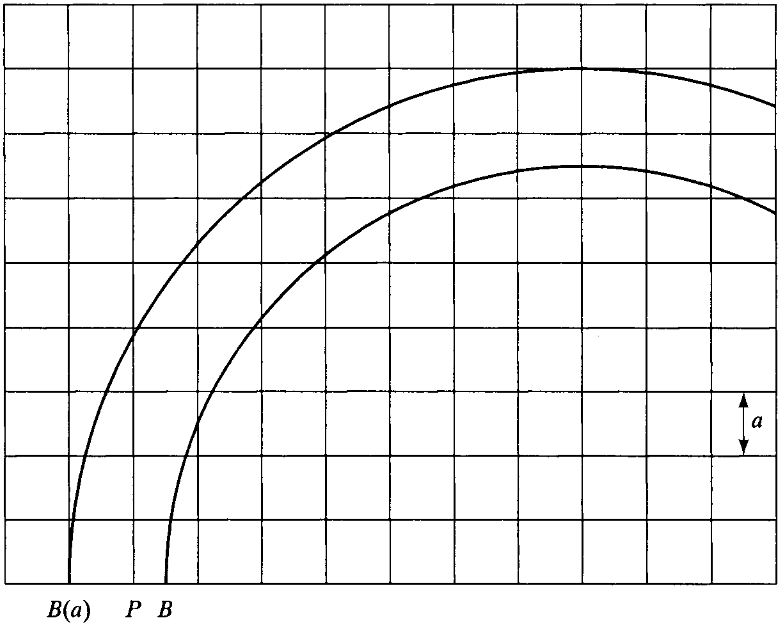
Proof. Let k be an integer such that $\lambda^k < \varepsilon/2$. By the preceding lemma, there is a union S_K of balls within a cube K of edge a, all of whose volumes lie among $\{\omega_k, \dots, \omega_1\}$, such that $\operatorname{vol}(K \setminus S) < \varepsilon \operatorname{vol}(K)/2$. Moreover, if R is the radius of $B(\omega)$, then there exists a union $P \subset B(\omega)$ of translates of K that completely covers the ball B' of radius $R - a\sqrt{n}$ (see the figure next page).

Thus,

$$\frac{\operatorname{vol}(B(\omega - P))}{\operatorname{vol}(B(\omega))} \leq \frac{\operatorname{vol}(B(\omega) - B')}{\omega} = \frac{R^n - (R - a\sqrt{n})^n}{R^n} \to 0$$

as $\omega \to 0$. Conversely, there exists ω_0 such that for $\omega \geq \omega_0$, we have $\operatorname{vol}(B(\omega) \setminus P) < \varepsilon\omega/2$. If S is the union of all collections of balls $S_K = v$ associated with the translates $K + v$ of K contained in P, then $\operatorname{vol}(P \setminus S) < \varepsilon \operatorname{vol}(P)/2$, and $\operatorname{vol}(B(\omega) \setminus S) < \varepsilon\omega$.

2.24. Proof of Proposition 2.22$_+$(1). The proof relies on the following, purely analytic property: If (α_p) is a sequence of positive numbers satisfying $\alpha_{i+j} \leq \alpha_i + \alpha_j$ for each i, j, then the sequence α_p/p is convergent (for an application of this fact to the spectral radius, see [Dieu], tome 2, Ch. 15, no. 2). When $n = 1$, our remarks in 2.21 imply that $\|\alpha^{p+q}\| \leq \|\alpha^p\| + \|\alpha^q\|$, and the proof is complete. If $\liminf_{p\to\infty} \|\alpha^p\|/p$, then inequality 2.21 again suffices. Let $\alpha \in \pi_n(X, x_0)$ and set $\nu = \liminf_{p\to\infty} \|\alpha^p\|/p$. Fix $\nu' > \nu$; we must show that, for sufficiently large p, we have $\|\alpha^p\| \leq p\nu'$. By hypothesis, there exists a sequence $p_j \to \infty$ such that $\liminf_{j\to\infty} \|\alpha^{p_j}\|/p_j = \nu$; we may assume that for each j, $\|\alpha^{p_j}\| < \nu' p_j$, and so there is a ω_j with $\|\alpha^{p_j}\| \leq \omega_j \leq \nu' p_j$, and a map $f_j\colon (B(\omega_j), \partial B(\omega_j)) \to (X, x_0)$ representing α^{p_j} and having dilatation ≤ 1. If the sequence (ω_j) is bounded, then we



cannot use Lemma 2.23, but the coarser line of argument in Proposition 2.11 suffices. Indeed, given a mapping $f_0\colon (B^n, \partial B^n) \to (X, x_0)$ representing an element β of $\pi_n(S^n, p)$, we can construct a representative f_r of $\beta^{N(r)}$ such that $N(r)(\mathrm{dil}(f_r))^{-n} \geq C(\mathrm{dil}(f_0))^{-n}$. Since $N(r)$ can assume any positive integer value, we can generalize Proposition 2.11 as follows: There exists a constant C depending only on n such that, for each $\beta \in \pi_n(X, x_0)$ and $q \in \mathbb{N}$, we have $\|\beta^q\|/q \leq C\|\beta\|$. On the other hand, given an integer p, we write $p = qp_j + r$, where $0 \leq r \leq p_j$, and set $\beta = \alpha^{p_j}$. Applying the inequality 2.21 to $\gamma = \alpha^{qp_j}$ and $\delta = \alpha^r$, we obtain

$$\begin{aligned}\left(\frac{\|\alpha^p\|}{p}\right)^{1/n} &\leq \left(\frac{\|\alpha^{qp_j}\|}{p}\right)^{1/n} + \left(\frac{\|\alpha^r\|}{p}\right)^{1/n}\\ &\leq \left(\frac{\|\alpha^{p_j}\|}{p_j + r/q}\right)^{1/n} + \left(\frac{\|\alpha^r\|}{p}\right)^{1/n}.\end{aligned}$$

For each $\varepsilon > 0$, there exists a j such that $C\|\alpha^{p_j}\|/p_j \leq (\varepsilon/2)^n$, and so there is a p_0 such that for $p \geq p_0$ and each r satisfying $0 \leq r \leq p_j$, we have $\|\alpha^r\|/p \leq (\varepsilon/2)^n$. Thus, for $p \geq p_0$, it follows that $\|\alpha^p\|/p \leq \varepsilon^n$, which shows that $\lim_{p\to\infty} \|\alpha^p\|/p = 0$.

If the sequence ω_j is not bounded, then by passing to a subsequence, we may assume that $\omega_j \geq 100\omega_{j-1}$ for each j. Fix $\varepsilon > 0$ so that by Lemma 2.23, there exists ω_0 such that for each $\omega \geq \omega_0$, there is a union S of balls in $B(\omega)$, each of whose volumes is among the ω_j, and $\mathrm{vol}(B(\omega)\setminus S) \le \varepsilon\omega$. Let $B_k(\omega_j)$ denote such a ball, $1 \le k \le m_j$. We define a map $f : (B(\omega), \partial B(\omega)) \to (X, x_0)$ by setting, for $1 \le k \le m_j$, $f|_{B_k(\omega_j)} = f_j$, and $f(B(\omega) \setminus S) = x_0$. Since $B(\omega)$ is a path metric space, $\mathrm{dil}(f) \le 1$ and f is a representative of $\sum_j m_j p_j$, this proves that $\|\sum_j m_j p_j\| \le \omega$. But $\mathrm{vol}(S) = \sum_j m_j p_j$ and $\mathrm{vol}(B(\omega) \setminus S) \le \varepsilon\,\mathrm{vol}(B(\omega))$, so that $\sum_j m_j\omega_j \ge (1-\varepsilon)\omega$. In other words, $\omega \le (1-\varepsilon\sum_j m_j\omega_j)^{-1} \le \nu'(1-\varepsilon\sum_j m_j p_j)^{-1}$, which is the desired inequality for the integer $p(\omega) = \sum_j m_j p_j$. When ω is a real number larger than ω_0, is the integer $p(\omega)$ greater than $(\nu'/(1-\varepsilon))\omega_0$? Not quite, since we know only that $(1-\varepsilon)\omega/\nu' \le p(\omega) \le \omega/\nu$ and $p(\omega)$ can exceed $\omega/\nu - (1-\varepsilon)\omega/\nu' \le (\nu'/\nu(1-\varepsilon) - 1)p(\omega)$. Note that, within the union of balls constructed in Lemma 2.23, the smaller balls comprise a nontrivial space, i.e., for a constant $\mu > 0$, we always have $m_1\omega_1 \ge \mu\sum_j m_j\omega_j$, and so $m_1 p_1 \le \mu p(\omega)$. For ν sufficiently close to ν' and sufficiently small ε, we have $(\nu'/\nu(1-\varepsilon) - 1) < \mu$, and the numbers p' of the form $km_1 + \sum_j m_j p_j$ for $0 \le k \le m_1$ may eventually exceed $p(\omega)$. By forgetting $m_1 - k$ balls of volume ω_1 within S, we construct a short representative of $\alpha^{p'}$ on the ball $B(\omega)$, which proves that, for those of these numbers that exceed $(v'/1-\varepsilon)\omega_0$, we have $\|\alpha^{p'}\| \le (\nu'/1-\varepsilon)p'$. We have therefore proved the first claim of Proposition 2.22 for almost all integers, and we conclude the proof with inequality 2.21.

2.25. Proof of Proposition 2.22$_+$ for general V. We will need the following version of Lemma 2.23 for each (Riemannian curved) simplex $\triangle^n$ in the triangulation of V. We rescale the metric in $\triangle^n$ so that the resulting simplex $\triangle^n(\omega)$ has (large) volume ω. We then pack almost all of $\triangle^n(\omega)$ with small, almost Euclidean balls whose volumes lie among the ω_i as in 2.24. As before, this gives us the desired short representative of α^p for $p = \sum_j m_j\omega_j$ mapping $(\triangle^n, \partial\triangle^n) \to (X, x_0)$. This proves 2.22 for $V = \triangle^n$, and the general case follows by choosing an enumeration of the simplices $\triangle_i^n$ in V and performing the above constructions for $\triangle_1^n, \triangle_2^n, \dots, \triangle_d^n$ with p_i proportional to $\mathrm{vol}(\triangle_i^n)$, $i = 1, \dots, d$, while the $\triangle_{i>d}^n$ go to x_0. This gives us all we need, since $p = \sum_{i=1}^d p_i \to \infty$ and $d \to \infty$.

Remark$_+$: The preceding argument is quite general in nature and relies exclusively on the following two facts:

(*a*) Riemannian manifolds are locally almost Euclidean.

(*b*) The Euclidean space $\mathbb{R}^n$ admits a nontrivial (i.e., having dilatation $\neq 1$) self-similarity.

The latter property is shared by many nilpotent Lie groups H equipped with left-invariant Carnot–Caratheodory metrics (such as the Heisenberg group $\mathbb{H}^{2m+1}$), and a general Carnot–Caratheodory manifold V can be locally approximated by such groups. This suggests the following H-norm on $\pi_n(X, x_0)$ for $n = \dim(H)$, which is defined using some precompact open topological ball $B \subset H$,

$$\|\alpha\|_B = \left(\inf_{f \in \alpha} \mathrm{dil}(f)\right)^h \mathrm{vol}_h,$$

where h denotes the Hausdorff dimension of H, vol_h is the corresponding Hausdorff measure, and the maps f representing $\alpha \in \pi_n(X, x_0)$ are those constant at infinity. As before, we set

$$\|\alpha\|_H = \|\alpha\|_B^{lim} = \lim_{p \to \infty} \frac{\|\alpha^p\|_B}{p},$$

and a straightforward generalization of the argument above shows that the limit exists and does not depend on $B \subset H$.

It is also easy to see that $\|\cdot\|_H$ (as well as our old norm $\|\cdot\|^{lim} = \|\cdot\|_H$ for $H = \mathbb{R}^n$) is a (possibly degenerate) *norm*, i.e.,

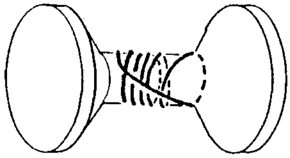
$$\|\alpha\beta\|_H \leq \|\alpha\| + \|\beta\|$$

for all $\alpha, \beta \in \pi_n(X)$ and $\|\alpha\|_H \neq 0$, provided that X is a compact Riemannian manifold and the rational Hurewicz homomorphism $\pi_n(X) \to H_n(X; \mathbb{Q})$ does not vanish on α (see 1.4.E′ in $[\mathrm{Gro}]_{\mathrm{CC}}$ for the proof and $[\mathrm{Gro}]_{\mathrm{DN}\ell\mathrm{S}}$ for the spectral interpretation of the phenomena described above).

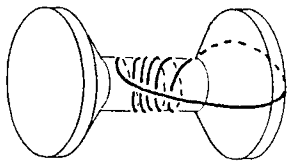
Turning to part (2) of the proof of Theorem 2.18 (sketched in $2.21\frac{1}{2}_+$), we must show that maps $V \to S^n$ which are concentrated on unions of small disjoint balls $B_j \subset V$ realize (asymptotically) the minimum dilatation in their homotopy class. For this purpose, we need the following fact.

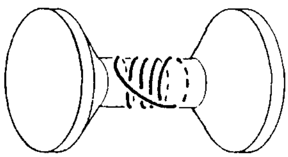
2.26. Proposition: *Let W be a compact, $(n-1)$-connected Riemannian n-manifold and let P be an $(n-1)$-dimensional polyhedron. For each integer p, there exist numbers $t_n(p)$ and $d_0(p)$ such that, for each $d \geq d_0(p)$ and for each mapping $f\colon P \to W$ of dilatation d, there is a homotopy between f and a constant map, defined on $P \times [0, d^{-1}t_n(p)]$ and having dilatation less than $(1 + 1/p)d$.*

It is useful to keep in mind the case when $n = 2$: suppose the map f describes a loop wound around a dumbbell W, making a sufficiently large number d of turns around the stem. Each of the d loops can be removed from the stem if we permit its length to be increased by a number K depends only on W.



We cannot release all of the turns simultaneously, since this operation multiplies the dilatation by K, which is too large. But, if we enlarge the loop by a length equal to K turns (which only increases the dilatation by a factor of $1 + K/d$), we can remove the loops one by one. The operation is lengthy, having on the order of d steps. If we permit ourselves a margin of $(1 + 1/p)$ for the dilatation and if d is large relative to p, i.e., $d = kpK$, then we can divide the loop in k equal parts, on each of which we perform the preceding deformation. This requires a time of pK and multiplies the dilatation by $1 + K/Kp = 1 + 1/p$, neither of which depend on d.





2.27. Scolie: *Let $\triangle$ be a metric space within Lipschitz distance (see Section 3.A) L from the cube $[0, \ell]^{n-1}$. Then there is a constant K depending only on L and the $(n-1)$-connected manifold W, such that each mapping $f : \triangle \to W$ of dilatation λ contracts to a point via a homotopy $\triangle \times [0, \lambda^{-1}] \to W$ of dilatation at most $K\lambda$.*

Proof. Since the question is unaffected by a change of scale, we may assume that $\lambda = 1$ and replace $\triangle$ by $[0, \ell]^{n-1}$, since the two are connected by a homeomorphism $\varphi : \triangle \to [0, \ell]^{n-1}$ for which $\mathrm{dil}(\varphi)$ and $\mathrm{dil}(\varphi^{-1})$ are bounded by e^L (as follows from the definition of L, see Section 3.A). Then we triangulate W, denote by δ the Lebesgue number of the corresponding covering W by the stars of the simplices, and use a standard triangulation of

$[0,\ell]^{n-1}$ by simplices of size $\delta' \simeq \delta$ (which is done by first dividing $[0,\ell]^{n-1}$ into smaller cubes of the form $[x, x+\varepsilon]^{n-1}$ for suitable points $x \in [0,\ell]$ and small ε, say $\varepsilon = \delta/n$).

Now $f\colon [0,\ell]^{n-1} \to W$ admits a simplicial approximation $f_1\colon [0,\ell]^{n-1} \to W$ which sends $[0,\ell]^{n-1}$ to the $(n-1)$-skeleton of W and has $\mathrm{dil}(f_1) \le \delta/\delta'$. Furthermore, if δ is small (less than a fraction of the injectivity radius of W), then there is a geodesic homotopy $[0,\ell]^{n-1} \times [0,1] \to W$ between f and f_1 sending each line $x \times [0,1]$ to the minimizing geodesic segment between $f(x)$ and $f_1(x)$. This exists since $\mathrm{dist}(f(x), f_1(x)) \le \delta$ and the dilatation of this geodesic homotopy is clearly bounded by a constant independent of f. Finally, we compose this homotopy with a contraction of the $(n-1)$-skeleton of W to a point (recall that W is $(n-1)$-connected) and thus obtain the required contraction of f.

2.28. From now on, we will replace any map $f : P \to W$ of dilatation d by the short map that it defines on the homothetic copy dP of P. For an integer p and real number K given by Scolie 2.27, we set $t_n = 3(1 + pK^2 + \cdots + p^n K^{2n})$ and let $i(P)$ be the smallest distance between two disjoint simplices in P having complementary dimension. We will show by induction that for each p there exists i_n such that if $i(P) \ge i_n$, then each short map $f\colon P \to W$ admits a homotopy H to a constant w_0, defined on $P \times [0, t_n]$ and having dilatation $< (1 + 1/p)$.

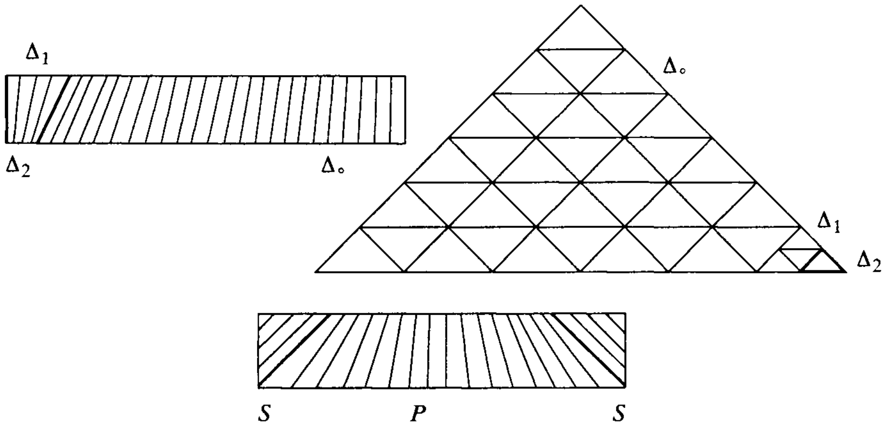
Assuming that this statement is true for $n-1$, and given an n polyhedron P such that $i(P) \ge i_{n-1}(pK)^{2n}$, there exists a subdivision of P into simplices $\triangle_0$ such that

$$\mathrm{dist}_L(\triangle, \text{standard simplex dilated by } (pK)^{2n}) \le L$$

(as in Scolie 2.27). We further subdivide each $\triangle_0$ into $2(pK)^n$ simplices $\triangle_1$, and each $\triangle_1$ into K^n simplices $\triangle_2$, in such a way that each $\triangle_2$ is almost standard, and such that there exists for each inclusion $\triangle_2 \subset \triangle_1 \subset \triangle_0$ a diffeomorphism D_1 (belonging to a deformation D_t, $0 \le t \le 1$) with $D_1(\triangle_1) = \triangle_2$, $\mathrm{dil}(D_1|_{\triangle_1}) \le 1/K$, and, for each t, $\mathrm{dil}(D_t) \le 1 + 1/2p$ (for $n = 1$, such a deformation is illustrated below).

By the inductive hypothesis, there exists a homotopy of the restriction of f to the $(n-1)$-skeleton (arising from the $\triangle_2$-subdivision) S defined on $S \times [0, t_{n-1}]$ and having dilatation less than $1 + 1/4p$. Since $i(\triangle_2)$ is very large in comparison with t_{n-1}, this homotopy can be used to produce an extension of f to $P \times [0, t_{n-1}]$ having dilatation again bounded by $1 + 1/2p$ (see figure).

Now we have $f(S, t_{n-1}) = w_0$, and we will describe an extension of f to



$\Delta_0 \times [0, t_{n-1}+3]$ such that $f(\Delta_2, [0, t_{n-1}+3]) = w_0$ and $f(x, t_{n-1}+3) = f(x, t_{n-1})$ if $x \in \Delta_0 \setminus \Delta_2$, with dilatation less than $1 + 1/p$. In fact, we will accomplish the extension on each simplex simultaneously Δ_0 (having already chosen the Δ_2 within each), in such a way as to obtain an extension to $P \times [0, t_{n-1}+3]$.

(1) We perform the deformation D_t, i.e., we set

$$f(x, t_{n-1}+t) = f(D_t(x), t_{n-1}).$$

By the construction of D_t, we have

$$\operatorname{dil}(f) \leq 1 + 1/2p + 1/2p \qquad \text{and} \qquad \operatorname{dil}(f|_{\Delta_2 \times [t_{n-1}+1]}) \leq 1/K.$$

(2) Since $f(\partial\Delta_2, t_{n-1}) = w_0$, $f(\partial\Delta_1, t_{n-1}+1) = w_0$, and since the extension of f to $\Delta_1 \times [t_{n-1}+1, t_{n-1}+2]$ provided by Scolie 2.27 (with $\operatorname{dil}(f) \leq K \cdot 1/K = 1$ and $f(\Delta_1, t_{n-1}+2) = w_0$) is compatible with the trivial extension of f to $(\Delta_0 \setminus \Delta_1) \times [t_{n-1}+1, t_{n-1}+2]$, we obtain an extension to all of Δ_0.

(3) We perform the inverse of the deformation D_t, i.e., we set

$$f(x, t_{n-1}+2+t) = f(D_{1-t}(x), t_{n-1}+2),$$

so that $\operatorname{dil}(f) \leq 1 + 1/K$, $f(\Delta_2, t_{n-1}+3) = w_0$, and $f(x, t_{n-1}+3) = f(x, t_{n-1})$ for $x \in \Delta_0 \setminus \Delta_2$.

To conclude, we perform this operation successively for each $\Delta_2 \subset \Delta_0$ which gives an extension of f to the product $P \times [0, t_{n-1} + 3(p^n K^{2n})]$ such that $f(P, t_n) = w_0$.

2.29. Contracting maps $V \to S^n$ on disjoint balls B_j. We will apply Proposition 2.26 (or, more precisely, the statement made at the beginning of 2.28) to a sphere $\partial B^n(R)$ of large radius R and a short map $f\colon B^n(R) \to S^n$. We can convert the homotopy defined on $\partial B^n \times [0, t_{n-1}]$ to an extension of f to the annulus $B^n(R+t_{n-1}) \setminus B^n(R)$ with $f(B^n(R+t_{n-1}))$ equal to the point $s_0 \in S^n$ and with dilatation less than $1 + 1/p$. In the manifold V, we assume that there is a union S of small, almost Euclidean balls such that $\operatorname{vol}(V \setminus S) < \varepsilon \operatorname{vol}(V)$. Fix a sufficiently large ω and a short map $f\colon V(\omega) \to S^n$ that almost minimizes the dilatation in its homotopy class, and set $k = 1 + 1/p + t_{n-1}(p)/\omega^{1/n}$. For each ball B_j of the union S, we consider $B_j(\omega)$ as a subset of $B_j(k\omega) \subset V(k\omega)$. Then f is well defined and almost short, and we construct two extensions g, h of f to $V(k\omega)$.

(1) There is an extension g of $f|_{B_j(\omega)}$ to $B_j(k\omega)$ of dilatation less than k, and such that $g(\partial B_j(k\omega)) = s_0$. We extend g to $V(k\omega)$ by the constant s_0, so that g is "concentrated on small balls."

(2) We set $h = f$ on $V(k\omega) \setminus S(k\omega)$ and also on $S(\omega)$, so that it remains to extend h radially on each annulus $B_j(k\omega) \setminus B_j(\omega)$. Since h is then homotopic to f, we have $\deg(h) = \deg(f)$; but $g = h$ on $S(\omega)$, whence

$$\begin{aligned} |\deg(f) - \deg(g)| &= (\operatorname{vol}(S^n))^{-1} \int_{V(k\omega)\setminus S(\omega)} (h^* - g^*)\nu_{S^n} \\ &\leq \frac{2k \operatorname{vol}(V(k\omega) \setminus S(\omega))}{\operatorname{vol}(S^n)} \\ &\leq \frac{2k(k-1+\varepsilon)\omega}{\operatorname{vol}(S^n)}, \end{aligned}$$

or, in other words, the ratio $\deg(f)/\deg(g)$ can be made arbitrarily close to 1 as ω tends to ∞. Since the mapping g is constructed with the aid of a union of arbitrary balls, we know that $(\operatorname{dil}(g))^n/\deg(g)$ is close to ν when $\deg(g)$ is large, and this completes part (2) of $2.21\frac{1}{2}{}_+$.

2.30. Proof of Theorem 2.18. Recall that by our preceding definitions,

$$\begin{aligned} \#(D) &= \operatorname{card}\{\text{homotopy classes of maps } V \to S^n \text{ having a representative} \\ &\qquad\qquad \text{with dilatation less than } D\} \\ &= \operatorname{card}\{p \in \mathbb{Z} : \exists\, f\colon V \to S^n \text{ with } \deg(f) = p \text{ and } \operatorname{dil}(f) \leq D\} \\ &= \inf_{D'>D} \#\{p \in \mathbb{Z} : \|\alpha_0^p\| \leq \operatorname{vol}(V)D^n\}. \end{aligned}$$

Since $\|\alpha_0^p\| \sim p\|\alpha_0\|^{lim}$ as $p \to \infty$, it is easy to see that as $D \to \infty$,

$$\operatorname{card}\{p \in \mathbb{N} : \|\alpha_0^p\| \leq \operatorname{vol}(V)D^n\} \sim \frac{D^n \operatorname{vol}(V)}{\|\alpha_0\|^{lim}}.$$

So that $\#(D) \sim \operatorname{vol}(V)c_g D^n$ as $D \to \infty$, where $c_g = 1/\|\alpha_0\|^{lim} + 1/\|\alpha_0^{-1}\|^{lim}$ does not depend on V.

$2.30\frac{1}{2}{}_+$ Lipschitz extension problem. Let Q and X be metric spaces and let $P \subset Q$ be a subspace. We then define the *Lipschitz extension function* $\operatorname{Ex}(\lambda) = \operatorname{Ex}_{P,Q,X}(\lambda)$ as follows: Consider all Lipschitz maps $f : P \to X$ which admit continuous extensions to maps $g\colon Q \to X$, and let $\operatorname{Ex}(\lambda)$ be the infimum of the numbers Λ such that *every* f with $\operatorname{dil}(f) \leq \lambda$ extends to some g with $\operatorname{dil}(g) \leq \Lambda$. For example, consider $Q = P \times [0, \ell]$ equipped with the product metric $\left(\text{say with } \operatorname{dist}_Q = \sqrt{\operatorname{dist}_P^2 + ds^2}\right)$, and let $\operatorname{Ex}_\ell(\lambda)$ be the corresponding extension function from $P \times 0 \cup P \times \ell$ to $Q = P \times [0, \ell]$. Note that $\operatorname{Ex}_\ell(\lambda)$ is monotone decreasing in ℓ and that the limit $\operatorname{Ex}_\infty(\lambda)$ measures the infimal Λ, such that every two homotopic maps $P \to X$ of dilatation $\leq \lambda$ can be joined by a homotopy of maps of dilatation $\leq \Lambda$. Finally, we can pose the same homotopy question for one of the maps $P \to X$ being constant and denote the corresponding *Lipschitz contractibility* function $\operatorname{Con}(\lambda) = \operatorname{Con}_{P,X}(\lambda)$.

Proposition 2.26 gives us quite satisfactory information about $\operatorname{Ex}_\ell(\lambda)$ and $\operatorname{Con}(\lambda)$ when X is a compact, n-connected Riemannian manifold (or a piecewise Riemannian space) and P is a compact n-dimensional polyhedron with a piecewise Riemannian metric. Specifically, 2.26 says that $\operatorname{Con}(\lambda)/\lambda \to 1$ as $\lambda \to \infty$ and, moreover, $\limsup_{\lambda\to\infty} \operatorname{Ex}_{\ell/\lambda}(\lambda) = 1 + \varepsilon(\ell)$, where $\varepsilon(\ell) \to 0$ for $\ell \to \infty$. (We suggest that the reader evaluate the actual rate of decay of this $\varepsilon(\ell)$.) Thus, $\operatorname{Ex}_{P,Q,X}(\lambda)/\lambda \to 1$ as $\lambda \to \infty$ for all compact polyhedra $Q \supset P$.

E_+ Degrees of short maps between compact and noncompact manifolds

2.31. Problem: What is the value of the constant c_g? The expression in 2.30 shows that $c_g > 0$, and Proposition 2.11 shows that $c_g \leq 1$. In particular, when do we have $c_g = 1$?

Flat example: If V, W are flat tori of the same dimension and volume, there exists a sequences of mappings $f_k\colon V \to W$ such that $\deg(f_k) \to \infty$

and $(\operatorname{dil}(f_k))^n/\deg(f_k) \to 1$ as $k \to \infty$.

Indeed, we write $V = \mathbb{R}^n/\Lambda$ and $W = \mathbb{R}^n/\Lambda'$, for two lattices Λ, Λ' in Euclidean n-space $\mathbb{R}^n$. By hypothesis, these lattices have the same volume, i.e., if (e_i) and (f_i) are the bases of Λ, Λ', and if L is an endomorphism of $\mathbb{R}^n$ mapping e_i onto f_i for each i, then $\det(L) = 1$. Let x_{ij} denote the j-th component of e_i with respect to the basis (f_i). From compactness of $\mathbb{R}^{n^2}/\mathbb{Z}^{n^2}$, if follows that for each $\varepsilon > 0$, there is an integer p such that for all i, j we have $d(px_{ij}, \mathbb{Z}) < \varepsilon$, or, in other words, if we denote by m_{ij} the closest integer to px_{ij} and $f'_j = \sum_k m_{jk} f_k$, then $\|f'_j - pe_j\| < \sqrt{n\varepsilon}$. From this, we can easily deduce that if L' is the endomorphism that maps each e_i to f'_i, then $|\operatorname{dil}(1/pL') - 1| < \sqrt{n\varepsilon}$ and $|(\det(1/pL'))^{1/n} - 1| < \sqrt{n\varepsilon}$. The endomorphism L', which sends Λ onto Λ', induces a map $\ell' \colon V \to W$ whose degree equals the index of the subgroup $L(\Lambda)$ in Λ', hence

$$\deg(\ell') = \frac{\operatorname{vol}(L'(\Lambda))}{\operatorname{vol}(\Lambda')} = \frac{\det(L')\operatorname{vol}(\Lambda)}{\operatorname{vol}\Lambda'} = \det(L') \leq p^n(1 + \sqrt{n\varepsilon})^n$$

so that $\operatorname{dil}(\ell) = \operatorname{dil}(L') \geq p(1 - \sqrt{n\varepsilon})$. We can therefore make the ratio $(\operatorname{dil}(\ell))^n/\deg(\ell')$ arbitrarily close to 1.

2.32. Converse to the flat example and proof of the inequality $c_g < 1$. We want to prove the following statement.

If closed, connected Riemannian n-manifolds V and W admit a sequence of maps $f_k \colon V \to W$ such that

$$\operatorname{dil}(f_k) \to \infty$$

for $k \to \infty$, and

$$\limsup_{k \to \infty} \frac{\deg(f_k)}{(\operatorname{dil}(f_k))^n} \geq \frac{\operatorname{vol}(V)}{\operatorname{vol}(W)}, \qquad (*)$$

then V and W are both flat.

This implies that the constant in Theorem 2.18 satisfies $c_g < 1$ for $n \geq 2$, since no metric g on S^n is flat. In fact, we will prove the more general

2.33. Volume rigidity theorem: *Let V and W be oriented n-dimensional manifolds, where we assume that W is compact and connected and that V satisfies* $\operatorname{vol}(V) < \infty$. *Suppose that there exist maps $f_k : V \to W$ with* $\operatorname{dil}(f_k) \to \infty$ *and satisfying*

$$\operatorname{vol}(f_k) \sim (\operatorname{dil} f_k)^n \operatorname{vol}(V) \qquad (\star)_V$$

(where $A(k) \sim B(k)$ signifies $A(k)/B(k) \to 1$). Then V and W are flat, and moreover W has no boundary.

Remark: This yields the above theorem on degrees, since

$$\deg(f) = \frac{\mathrm{vol}(f)}{\mathrm{vol}(W)},$$

where we recall that

$$\mathrm{vol}(f) \overset{\mathrm{def}}{=} \int_V J(f)\, d\nu_V = \int_V f^*(\nu_W)$$

for the Jacobian $J(f)$ of f.

Notice that the Jacobian of f (or, equivalently, the pullback volume form $f^*(\nu_W)$) is, a priori, defined only for smooth maps f. The necessary smoothness can always be achieved by a smooth approximation f_k of given map which represents a negligible perturbation of the dilatations and volumes of these maps. Alternatively, one can appeal to the *Rademacher-Stepanov theorem*, which states that every Lipschitz map is a.e. differentiable and then observe that $\int_V J(f)$ agrees with the topological degree for Lipschitz maps to the same extent as for smooth ones.

Proof. First we observe that the asymptotic relation $(\star)_V$ above is inherited by all open subsets $U \subset V$. In fact, if U is an arbitrary measurable subset in V, then the relation

$$\int_U J(f_k)\, d\nu_V \sim (\mathrm{dil}\, f_k)^n \,\mathrm{mes}(U) \qquad (\star)_U$$

follows trivially from the inequality $J(f_k) \leq (\mathrm{dil}(f_k))^n$ and the relation $(\star)_V$ written as

$$\int_U J(f_k)\, d\nu_V + \int_{V\setminus U} J(f_k)\, d\nu_V \sim (\mathrm{dil}\, f_k)^n \,\mathrm{mes}(U) + (\mathrm{dil}\, f_k)^n \,\mathrm{mes}(V \setminus U).$$

This applies in particular to each ball $B(v,\varepsilon)$ in V around a fixed point $v \in V$. Such balls, when rescaled to unit size become approximately Euclidean. More precisely, the unit Euclidean ball B admits (rescaled exponential) diffeomorphisms $e_\varepsilon \colon B \to B(v,\varepsilon)$ for all small $\varepsilon > 0$, such that $\mathrm{dil}(e_\varepsilon) \sim \varepsilon$ and $\mathrm{dil}(e_\varepsilon^{-1}) \sim \varepsilon^{-1}$ for $\varepsilon \to 0$. Such e_ε necessarily have $J(e_\varepsilon) \sim \varepsilon^n$, and therefore the composed maps $f_k^0 = f_k \circ e_{\varepsilon_k}$ for $f_k = f_k|_{B(v,\varepsilon_k)}$ and a suitable sequence ε_k (tending to 0 as $k \to \infty$) satisfy $(*)_B$, i.e.,

$$\mathrm{vol}(f_k^0) \sim (\mathrm{dil}\, f_k^0)\, \mathrm{vol}(B). \qquad (\star)_B$$

(In order to satisfy $(\star)_B$, the sequence ε_k must converge to zero quite slowly, say, much slower than $(\mathrm{dil}\, f_k)^{-1}$. In fact, the actual requirement on ε_k depends on the rate of convergence of

$$\frac{\mathrm{vol}(f_k)}{(\mathrm{dil}(f_k))^n\, \mathrm{vol}(V)} \to 1$$

implied by $(\star)_V$ and the convergences $\operatorname{dil} e_\varepsilon/\varepsilon \to 1$ and $\operatorname{dil} e_\varepsilon^{-1}/\varepsilon^{-1} \to 1$.

Thus, we have completed the first step of the proof by showing that *the existence of f_k with an arbitrary V implies the existence of a similar sequence for V equal to the flat Euclidean ball.*

2.34. Remarks: (*a*) A similar reduction $V \rightsquigarrow B$ is possible for maps f_k satisfying the (nonextremal) relation

$$\operatorname{vol}(f_k) \sim \lambda(\operatorname{dil} f_k)^n \operatorname{vol}(V) \qquad (\star)_{V,\lambda}$$

for a given $\lambda \in [0,1]$. Here again one can easily produce a sequence of maps $f_k^0 : B \to W$ satisfying $(\star)_{B,\lambda}$. But, now one should exercise some care in choosing the centers v_k of the balls $B(v_k, \varepsilon_k)$ in V in order to have the integral

$$\int_{B(v_k,\varepsilon_k)} J(f_k)\, \nu_V$$

sufficiently large, i.e., on the order of at least $\lambda \operatorname{vol}(B(v_k, \varepsilon_k))$.

(*b*) One could rescale the ball $B = B(1)$ to $B(R_k)$ for $R_k = \operatorname{dil}(f_k^0)$, thus obtaining *short* maps, say $f_k^0 : B(R_k) \to W$ satisfying $(\star)_{B(R_k),\lambda}$. Then by a simple compactness argument (left as an exercise for the reader), one can extract a suitable convergent subsequence, thus obtaining a *short map* $F : \mathbb{R}^n \to W$, where $\mathbb{R}^n$ admits an exhaustion by (nonconcentric!) balls $B(R_i)$, $i = 1, \ldots$ (where these R_i have nothing to do with the R_k's) such that

$$\operatorname{vol}(F|_{B(R_i)}) \sim \lambda R_i^n \operatorname{vol}(B). \qquad (\star)_\lambda$$

This can be expressed by saying that *the asymptotic degree* of F is $\geq \lambda$.

Next, in order to prepare the second step of the proof, we observe the following trivial

2.35. Lemma: *Let $J = J(x)$ be a measurable function on the Euclidean ball $B(R)$ such that $J \leq 1$ and*

$$\int_{B(R)} J(x)\, dx \geq (1-\delta) \operatorname{vol} B(R)$$

for a given $\delta \in [0,1]$. Then, for every given $\varepsilon > 0$, there exists a ball $B(x,r)$ contained in $B(R)$, where $r \geq \rho(R,\delta,\varepsilon)$ for a fixed function ρ satisfying $\rho(R,\delta,\varepsilon) \to \infty$ as $R \to \infty$ and fixed $\delta \in [0,1]$, and such that every ε-ball $B(y,\varepsilon)$ in $B(x,r)$ satisfies

$$\int_{B(y,\varepsilon)} J(x)\, dx \geq (1-\delta') \operatorname{vol} B(y,\varepsilon),$$

where $\delta' = \delta'(\delta, \varepsilon) \to 0$ as $\delta \to 0$ and each fixed $\varepsilon > 0$.

In other words, the "bad" ε-balls which violate the inequality above with δ' going to zero much slower than δ, cannot intersect every r-ball in $B(R)$ with moderately large r. Otherwise, the assumed lower $(1-\delta)$-bound on the integral $\int_{B(R)} J$ would be violated, as can be seen from an obvious integration argument.

2.36. Proposition: *Given $f_k : V \to W$ satisfying the assumptions of the volume rigidity theorem, there exists a short map $F\colon \mathbb{R}^n \to W$ such that*

$$\operatorname{vol}(F|_U) = \operatorname{vol}(U)$$

for every bounded domain $U \subset \mathbb{R}^n$.

Proof. We used the above map f_k° of the Euclidean ball B to W satisfying $(\star)_B$ which we transform to short maps $f_k^\square : B(R_k) \to W$ for $R_k = \operatorname{dil} f_k^\circ$. Then we apply the lemma above to the Jacobian $J = J_k = J(f_k^\circ)$ for a suitable sequence ε_k slowly going to zero as $k \to \infty$. Thus, we obtain balls $B(x_k, r_k)$ within $B(R_k)$ with $r_k \to \infty$ such that

$$\int_{B(y,\varepsilon)} J(f_k^\square)\, dx \xrightarrow{k\to\infty} \operatorname{vol} B(y, \varepsilon)$$

for every fixed $\varepsilon > 0$ and all $y \in B(x_k, r_k)$. Finally, we apply Ascoli's theorem to the maps $f_k^\square$ on the balls $B(x_k, r_k)$, which can be thought of as the balls in $\mathbb{R}^n$ around the origin, and obtain as a limit of a convergent subsequence the desired map $F\colon \mathbb{R}^n \to W$.

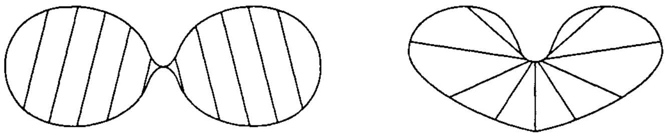
2.37. Remark: So far, we have not used the geometry of W. For example, we could allow up to this point a singular, e.g., piecewise linear (or piecewise smooth) metric. An example of such a metric on S^2 arises as the quotient metric, where S^2 is obtained from the flat torus T^2 via the involution $t \mapsto -t$. Here, "$-$" refers to the group structure on $T^2 = \mathbb{R}^2/\mathbb{Z}^2$. The quotient map $f\colon T^2 \to S^2$ is obviously short, of degree $2 = \operatorname{vol}(T^2)/\operatorname{vol}(S^2)$. This can be composed with self-mappings $t \mapsto kt$ of the torus, thus producing a sequence of maps $f_k\colon T^2 \to S^2$ with $\operatorname{dil} f_k = k \to \infty$ and

$$\operatorname{vol}(f_k) = (\operatorname{dil} f_k)^2 \operatorname{vol}(T^2).$$

This may seem to contradict our theorem, since S^2 is never flat. However, this $S^2 = T^2/\mathbb{Z}_2$ is nonsmooth, i.e., its metric does not come from a continuous Riemannian metric. In fact, the quotient metric on this S^2 is Riemannian flat apart from 4 singular points where it is isometric to the

Euclidean cone over the circle of circumference π (rather than 2π, as is the case at the regular points). This π-angle at the singular points contributes an infinite amount of positive curvature corresponding to the angle deficiency $\pi = 2\pi - \pi$. We shall indicate in 2.40 a possible generalization of the volume rigidity theorem to singular spaces.

2.38. Proving that W is flat. Our short map $F: \mathbb{R}^n \to W$ is almost everywhere differentiable, and its differential is a.e. isometric. The trouble is that this F is not *a priori* locally one-to-one (and it should not be, in general, as the remark above shows). The idea is to prove that F is one-to-one goes as follows. Take a small ball $B(\varepsilon)$ and suppose that F is not one-to-one on $B(\varepsilon)$. For example, F might identify two opposite points on the boundary of $B(\varepsilon)$, or it might map a boundary point close to the image of the center of $B(\varepsilon)$.

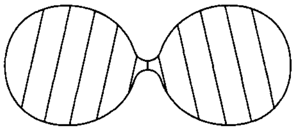


(a) (b)

Then, we note that $\mathrm{vol}(F|_B(\varepsilon))$ depends only on the map F on $S(\varepsilon) = B(\varepsilon)$, or rather on the resulting hypersurface $F(S(\varepsilon)) \subset W$ (where we should keep track of multiplicity if $\partial B(\varepsilon)$ goes through some hypersurface several times, such as the map $z \mapsto z^2$ on $S^1 \subset \mathbb{C}$ does). Thus, if we could fill $F(S(\varepsilon))$ by some ball (or more general n chain) in W with volume less than $\mathrm{vol}(B(\varepsilon))$, we would arrive at a contradiction, thus showing that F was indeed 1-1 after all. To this end, we start with the case $n = 2$ and look at the curve (a) with two opposite points identified. First, we assume that we are in $\mathbb{R}^2$ and apply the isoperimetric inequality to claim that each half of the curve bounds an area $\leq B(\varepsilon/2)$ and so the whole curve bounds and area $\leq \mathrm{Area}(B(\varepsilon))$. We then recall that a general *riemannian* W is locally almost Euclidean, and so every sufficiently short curve of length $\pi\varepsilon$ can be filled by a disk of

$$\mathrm{Area} \leq \pi\varepsilon^2/4 + o(\varepsilon^2),$$

which is smaller than $\mathrm{Area}\, B(\varepsilon)/2 = \pi\varepsilon^2/2$ for small ε. Since for any pair of collapsing points x and y in $\mathbb{R}^2$ one can find a disk B in $\mathbb{R}^2$ having them opposite on ∂B, we conclude that F *must be locally one-to-one* in order to have $\mathrm{vol}(F|_{B(\varepsilon)}) = \mathrm{vol}(B(\varepsilon))$ for small balls $B(\varepsilon)$. In fact, this argument also shows that the local inverse map F^{-1} is Lipschitz, since F cannot bring two opposite points on $B(\varepsilon)$ too close together,



(a')

since we have a good margin to use in our isoperimetric inequality. Now, once we know that the inverse of F is Lipschitz, it is necessarily isometric, since its differential is a.e. isometric. (Probably, this extra "Lipschitz" is unnecessary, but we get it for free anyway).

Now we turn to the case $n \geq 3$ and consider the figure (b) above, where the image $F(S(\varepsilon) = \partial B(\varepsilon))$ passes through the image of the center $F(0) \in W$. We fill $F(S(\varepsilon))$ radially from the center $F(0)$ assuming $W = \mathbb{R}^n$. The volume of the resulting conical filling is obviously bounded by

$$\frac{1}{n} \int_{S(\varepsilon)} \|F(0) - F(s)\| \, ds.$$

Since F is Lipschitz, the integrand is bounded by $\varepsilon \operatorname{vol}_{n-1} S(\varepsilon)/n = \operatorname{vol} B(\varepsilon)$. But, if some point $F(s_0)$ equals $F(0)$, then the integrand becomes smaller than $\varepsilon/2$ on a definite ball, say of radius $\varepsilon/2$ in $S(\varepsilon)$, which makes the resulting integral smaller than $\operatorname{vol}(B(\varepsilon))$ by a definite amount $\geq (2n)^{-n}\varepsilon^n$. This applies equally to non-Euclidean Riemannian W for small ε and where, furthermore, the condition $F(s_0) = F(0)$ can be relaxed to $\operatorname{dist}(F(s_0), F(0)) \leq \lambda\varepsilon$ for a given $\lambda < 1$. Thus, one arrives at the conclusion that F is locally invertible and is moreover a λ^{-1}-Lipschitz map. Then one sees that F^{-1} is actually an isometry by either letting $\lambda \to 1$ or by arguing as before using some $\lambda > 0$ and appealing to the a.e. isometric property of the differential DF.

Finally, we observe that a locally isometric map $\mathbb{R}^n \to V$ is necessarily a covering map of $\mathbb{R}^n$ onto (connected) V, and so V is flat without boundary.

2.39. Proving that V is flat. We return to the original maps $f_k : V \to W$ restricted to a small ball $B(v, \varepsilon) \subset V$ which we lift to the universal covering, $B(v, \varepsilon) \to \tilde{W} = \mathbb{R}^n$ and the scale by $x \mapsto (\operatorname{dil} f_k)^1 x$ in $\mathbb{R}^n$. Thus, we get short maps, say $\tilde{f}_k : B(v, \varepsilon) \to \mathbb{R}^n$ satisfying $\operatorname{vol} \tilde{f}_k \to \operatorname{vol} B(v, \varepsilon)$. These subconverge to a short map, say $\tilde{F} : B(v, \varepsilon) \to \mathbb{R}^n$ with $\operatorname{vol}(\tilde{F}) = \operatorname{vol} B(v, \varepsilon)$, which is necessarily locally isometric by the discussion of the previous section.

2.40. Remarks and questions. (a) Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a short map satisfying $\mathrm{vol}(f|_B) \geq \lambda \,\mathrm{vol}(B)$ for all balls $B \subset \mathbb{R}^n$. If $\lambda > 1 - \varepsilon_n$ for a small positive ε_n, then our reasoning in $2.3.8_+$ shows that f is one-to-one with $\varepsilon_2 = 1/2$, where the latter is sharp, as the map $z \mapsto z^2/2|z|$ shows. Probably $\varepsilon_n = 1/2$ works for all $n \geq 2$ with the extremal example $(x, z) \mapsto (x, z^2/2|z|)$.

(*b*) If V and W are singular (riemannian) manifolds, e.g., with bi-Lipschitz Euclidean metrics, then the best one can expect is that V is locally $\mathbb{R}^n/G$ for some finite group G, while W is locally a ramified covering of $\mathbb{R}^n$. For example, if $n = 2$ and W is piecewise linear, with angles around vertices $> \pi$ (nonsingular points have these angles $> 2\pi$), then W is necessarily flat, since it satisfies a suitable isoperimetric inequality (whose elementary proof is left to the reader). Similarly, if V is piecewise linear and flat with angles $< 4\pi$, then W is flat. Also, one expects nontrivial topological restrictions on V and W. Again, the strongest conclusion would be $W = \mathbb{R}^n/\Gamma$, for a discrete group Γ (where the action of Γ on $\mathbb{R}^n$ may have fixed points) while W must be some kind of ramified covering of a flat manifold.

(*c*) In fact, if one is concerned with the topology rather than the geometry of V and W, one can consider only Riemannian manifolds and define $\lambda(d)$ as the minimal possible dilation of maps $V \to W$ with $\mathrm{vol}(f) \geq d$. (If V and W are closed manifolds, then these are the same as maps of degree $d/\mathrm{vol}(W)$.) In the above discussion, we were concerned with the extreme case where $\lambda(d) \sim d^{1/n}$. This can be relaxed to $\lambda(d) = O(\alpha d^{1/n})$ for some $\alpha > 1$ (where $A(d) = O(B(d))$ means that $\liminf_{d \to \infty} A(d)/B(d) \leq 1$) and yet it yields strong topological restrictions on V and W (compare 2.43_+ below). Nevertheless, it is unclear whether W can be homeomorphic in this case to a connected sum such as $S^2 \times S^2 \# S^2 \times S^2$ or, more generally, to have hyperbolic rational homotopy type (compare [Grov–Halp]).

If V is a nonclosed manifold, then, obviously, $\lambda(d) = O(\alpha d^{1/(n1)})$ (as is seen with maps sending all of V to a small ball in W), but if V and W are both closed, then the mere existence of maps $f_k : V \to W$ with $\deg(f_k) \to \infty$ imposes non vacuous restrictions on V and W. For example, W does not admit a metric with negative sectional curvature if such f_k exist for some V (see Ch. $5.\mathrm{F}_+$). One may expect that in most (?) cases $\lambda(d) \sim cd^\alpha$ for some constants c and α, where $\alpha = \alpha(V, W)$ appears to be an interesting homotopy invariant of V and W.

Generalizations. One can allow $n = \dim(V) < \dim(W)$ and define the volume of a map $f\colon V \to W$ relative to a fixed n-form ω on W by

$$\mathrm{vol}_\omega f = \int_V f^*(\omega).$$

Then, the topological discussion above makes perfect sense. Furthermore, if $\text{comass}(\omega) \leq 1$ (compare Ch. 4.D), then the existence of maps $f_k: V \to W$ with $\text{vol}_\omega f_k \sim (\text{dil} f_k)^n$ implies the existence of a path-isometric map $F: \mathbb{R}^n \to W$ such that $F^*(\omega) = \nu_V$, and if W is nonsingular (and Riemannian), then this F is smooth.

Another avenue for generalization is opened by admitting non-Riemannian Carnot–Caratheodory metrics on V. Here, one can have, for example, $\lambda d \sim cd^\alpha$ for $\alpha > \dim V$. In fact, the maximal possible α here equals the *Hausdorff dimension* of V (see 2.47 and [Gro]$_{CC}$).

2.41. Asymptotic degree and elliptic manifolds. Consider equidimensional Riemannian manifolds V and W, where V is assumed to be connected and oriented, take a Lipschitz map $f: V \to W$, and let

$$\delta(U) = \text{vol}(f|_U)/\text{vol}(U)$$

for all bounded (i.e., precompact) domains $U \subset V$. Here, we are interested in the case of $\text{vol}(V) = \infty$, e.g., $V = \mathbb{R}^n$, and we study the asymptotics of $\delta(U)$ for domains U exhausting V. The simplest exhaustion is that given by concentric balls $B(R) = V$ for $R \to \infty$, but we can also use exhaustions by nonconcentric balls.

Example: Suppose $V = \mathbb{R}^n$ is exhausted by (possibly nonconcentric) balls $B(R)$ such that $\delta(B(R)) \to (\text{dil}(f))^n$. Then the volume rigidity theorem applies to the maps $f|_{B(R)}$ (where, strictly speaking, the balls $B(R)$ should all be scaled to unit size) and shows that W is flat, provided that it is a *complete riemannian* manifold.

Now, we shift into the topological gear and ask ourselves when a given Riemannian manifold W admits a short (or just Lipschitz) map $f: \mathbb{R}^n \to W$ with *nonzero asymptotic degree*, i.e., having $\limsup_{R\to\infty} \delta(B(R)) > 0$ for some exhaustion of $\mathbb{R}^n$ by balls. Such manifolds are called *elliptic*. Obviously, this property is a homotopy invariant of W when W is a *closed* manifold. One expects such closed elliptic W to have *elliptic* (i.e., nonhyperbolic, see [Grov–Halp]), *rational homotopy type*. But, we do not even know whether $S^2 \times S^2 \# S^2 \times S^2) \to$ is elliptic. On the other hand, if W is elliptic, then its fundamental group must be rather small.

2.42. Theorem: *If W is elliptic, then the balls $\tilde{B}(R)$ in the universal cover $\tilde{W}$ satisfy*

$$\limsup_{R\to\infty} \frac{\text{vol}\,\tilde{B}(R)}{R^n} < \infty \qquad (*)$$

Proof. Lift f to a map $\tilde{f}\colon \mathbb{R}^n \to \tilde{W}$ and observe that this does not change the normalized volume of f, namely

$$\delta_f(U) = \delta_{\tilde{f}}(U)$$

for all $U \subset V$. On the other hand, if $\pi_1(W)$ is infinite, then $\mathrm{vol}_{\tilde{f}}(U)$ depends only on $\tilde{f}|_{\partial U}$. In particular, if $\tilde{f}(\partial U) \subset W$ bounds an n-chain (i.e., a multi-domain in W) of volume A, then

$$\mathrm{vol}(f|_U) = \mathrm{vol}(\tilde{f}|_U) \le A.$$

We apply this to $U = B(R)$ for $R \to \infty$ and observe that $\mathrm{vol}_{n-1}\tilde{f}(\partial B(R)) = O(R^{n-1})$ since $\tilde{f}$ is short. Now, if it were true that $\limsup \mathrm{vol}(\tilde{B}(R))/R^n = \infty$, then we could fill the hypersurface $\tilde{f}(\partial B(R))$ in $\tilde{W}$ by a chain of volume $\varepsilon(R)R^n$, with $\varepsilon(R) \to 0$ for $R \to \infty$ according to a theorem by Varopoulos (see Ch. 6.E).

2.43. Corollary: *If W is elliptic and its universal covering is contractible, then the fundamental group $\pi_1(W)$ is virtually abelian, i.e., it contains an abelian subgroup of finite index.*

Proof. The relation $(*)$ says that $\tilde{W}$, and hence the fundamental group $\pi_1(W)$ has polynomial growth of degree n, and so $\pi_1(W)$ is virtually nilpotent (see 5.7_+–5.10) But, it was observed a long time ago by Bass (see [Bass]) that if a torsion free nilpotent group Γ has its degree of growth less than or equal to it homological dimension, then Γ is abelian. Since our $\Gamma \subset \pi_1(W)$ appears as the fundamental group of a closed, aspherical n-manifold (a finite covering of W), it has homological dimension n, and Bass' observation applies.

2.44. Question: *Is the assumption of asphericity of W (i.e., of the contractibility of $\tilde{W}$) truly needed to conclude that $\pi_1(W)$ is virtually abelian?*

What is immediately clear at this stage is that $\pi_1(W)$ is virtually nilpotent of growth degree $\le n$. This does not exclude, for example, $W = S^2 \times N/\Gamma$, where N is the 3-dimensional Heisenberg group and Γ is a lattice in N. One knows that every large domain $\Omega \subset \tilde{W}$ for this W satisfies the *Pansu isoperimetric inequality*

$$\mathrm{vol}_5(\Omega) \le \mathrm{const}(\mathrm{vol}_4(\partial\Omega)^{4/3}$$

(compare Ch. $6.E_+$). In our case, the boundary $\partial\Omega$ comes by a short map from the Euclidean 4-sphere $S(R)$, and thus we have $\mathrm{vol}_4(\partial\Omega) = O(R^4)$.

What we need in order to prove that this W is nonelliptic is the asymptotic relation

$$\mathrm{vol}_5(\Omega) = o(R^5)$$

for $R \to \infty$, rather than $O(R^{16/3})$ by somehow using the special geometry of $\partial\Omega$. (Notice that Ω in this case is a multiple domain, but this may only improve the isoperimetric inequality.) Some nontrivial information concerning the shape of $\partial\Omega = f(S(R))$ for $R \to \infty$ can be obtained by rescaling $\tilde{W}$ by $\varepsilon = R^{-1}$ and passing to the limit as $\varepsilon \to 0$ (see $3.18\frac{1}{2}{}_+$), where the (Carnot–Caratheodory) geometry of the limit space $\lim_{\varepsilon\to 0} \varepsilon\tilde{W}$ severely restricts the size of $f(S(R)) \subset \tilde{W}$ (as well as that of $f(B(R))$) for large R (compare 3_+ in 2.17). This allows us, for example, to rule out nonabelian nilpotent groups of growth exactly n. Namely, if the fundamental group $\pi_1(W)$ is *not* virtually abelian, then the relation $(*)$ in 2.42 can be improved as follows:

$$\limsup_{R\to\infty} \frac{\mathrm{vol}(\tilde{B}(R))}{R^{n-1}} < \infty. \tag{$**$}$$

Sketch of the proof. By rescaling our maps by $\varepsilon = R^{-1}$, we get maps $f_\varepsilon\colon B(1) \to \varepsilon\tilde{W}$ which subconverge as $\varepsilon \to 0$ to a Lipschitz map $f_0\colon B(1) \to N$, where N is the Hausdorff limit of $\varepsilon\tilde{W}$ as $\varepsilon \to 0$ (compare 3_+ in 2.17), which is in our case a nonabelian nilpotent Lie group with a (non-Riemannian!) Carnot–Caratheodory metric. Hence, by the Pansu-Rademacher theorem, f_0 is a.e. differentiable, and $\operatorname{rank} Df_0 \le n-1$ a.e. Now assume that $(**)$ fails to be true, which yields that the volume growth is exactly of degree n for $\pi_1(W)$ and $\tilde{W}$. Then, the inequality $\operatorname{rank} Df_0 \le n-1$ a.e. for the limit map f_0 implies that the image of $B(R)$ under f satisfies

$$\frac{\mathrm{vol}(f(B(R)))}{R^n} \to 0$$

as $R \to \infty$. A priori, this does not rule out the possibility of $\mathrm{vol}(f|_{B(R)}) \sim R^n$, since the latter volume is counted with multiplicity. Nevertheless, the presence of high multiplicity enhances the power of the (Varopoulos) isoperimetric inequality in $\tilde{W}$, which reads in our case (compare Ch. $6.\mathrm{E}_+$)

$$\mathrm{vol}(\Omega) \le \mathrm{const}(\mathrm{vol}_{n-1}(\partial\Omega))^{n/(n-1)}. \tag{$\star$}$$

But, if Ω is a multiple domain given by a map $I : \Omega \to \tilde{W}$ (which can be assumed to be an immersion), and $\mu(w) = \mathrm{card}(I^{-1}(w))$ is the multiplicity of Ω at the points $w \in \Omega_1 = \mathrm{im}_I(\Omega)$, then the oriented volume

$$\mathrm{vol}(\Omega) \overset{\mathrm{def}}{=} \int_\Omega I^*(d\nu_{\tilde{W}})$$

can be bounded in terms of the average multiplicity $\overline{\mu} = \left(\int_\Omega \mu \, d\nu_{\tilde{W}}\right) / \operatorname{vol}(\Omega_1)$ by

$$\operatorname{vol}(\Omega) \leq \operatorname{const} \mu^{-1/(n-1)} (\operatorname{vol}(\partial\Omega))^{n/(n-1)}, \qquad (\star\star)$$

as follows from $(\star)$ applied to the subdomains $\Omega_i \subset \Omega_1$, where $\mu \geq i$, for $i = 1, 2, \ldots$. This applies to $\partial\Omega = f(\partial B(R))$ (although f does not have to be an immersion on $B(R)$), and shows that $\operatorname{vol}(f|_{B(R)})/R^n \to 0$, thus proving our assertion by contradiction.

Example: The manifold $S^2 \times N/\Gamma \times N/\Gamma$, where N is the Heisenberg group, is nonelliptic.

2.45. Elliptic manifolds W with $\pi_1(W) = \mathbb{Z}^n$ for $n = \dim(W)$.

Theorem: *If $\pi_1(W) = \mathbb{Z}^n$ and the universal covering $\tilde{W}$ admits a sequence of short maps f_i from the Euclidean balls $f_i \colon B(R_i) \to \tilde{W}$ such that* $\operatorname{vol} f_i \geq \operatorname{const} R_i^n$ *for some* const > 0 *and $R_i \to \infty$, then $\tilde{W}$ is contractible.*

Proof. Take a smooth map $\alpha \colon W \to T^n$ which induces an isomorphism on π_1, and look at the corresponding maps $\tilde{\alpha} \colon \tilde{W} \to \mathbb{R}^n = \tilde{T}^n$.

2.46. Lemma: *Let $f_i : B(R_i) \to \tilde{W}$ be short maps such that* $\operatorname{vol}(f_i) \geq \operatorname{const} R_i^n$ *for $R_i \to \infty$. Then there exist (Euclidean) balls $B(r_i) = B(x_i, r_i) = B(R_i)$ with $r_i \to \infty$ such that the image of the sphere $S(r_i)$ under the composition $\alpha \circ f_i$ does not intersect the ball*

$$B(y_i, cr_i) \subset \mathbb{R}^n \quad \text{for} \quad y_i = \alpha \circ f_i(x_i)$$

and a fixed $c > 0$ (independent of i) and such that the normal projection P to the boundary $\partial B(y_i, cr_i)$ gives us a map of degree 1 or -1 from the sphere $S(r_i)$ to $S(\rho) = \partial B(y_i, cr_i)$, i.e., the composition $P \circ \alpha \circ f_i : S(r) \to S(\rho)$ has $|\deg| = 1$.

Proof. We rescale the maps f_i by $\varepsilon_i = R_i^{-1}$ as before, obtaining a sublimit map $f_\infty : B(1) \to \mathbb{R}^n = \lim_{\varepsilon_i \to 0} \varepsilon_i \tilde{W}$. This composes with the rescaled *linear* map $\tilde{\alpha}$, say $\tilde{\alpha}_0 : \mathbb{R}^n = \lim_{\varepsilon_i \to 0} \varepsilon_i \tilde{W} \to \mathbb{R}^n = \lim_{\varepsilon_i \to 0} \varepsilon_i \mathbb{R}^n$, where the resulting composition $\tilde{\alpha}_0 \circ f_\infty$ could have been obtained without any Hausdorff limits by just composing the maps

$$B(1) \xrightarrow{R} B(R) \xrightarrow{\tilde{\alpha} \circ f} \mathbb{R}^n \xrightarrow{\varepsilon} \mathbb{R}^n,$$

(where R stands for the map $x \mapsto Rx$ and ε stands for $y \mapsto \varepsilon y$ with $\varepsilon = R^{-1}$) and passing to the sublimit as $R = R_i \to \infty$.

The limit map $\tilde{\alpha}_0 \circ f_\infty$ is a.e. differentiable by the Rademacher theorem, and $\operatorname{rank}(\tilde{\alpha}_0 \circ f_\infty(x_0)) = n$ at some point of differentiability in the interior of $B(1)$, since otherwise rank $\leq n-1$ would enhance the (Euclidean) isoperimetric inequality for multiple domains bounded by $f_i(S(R_i)) \subset \tilde{W}$ and would imply $\operatorname{vol}(f_i|_{B(R_i)}) = o(R_i^n)$ as before. Now, since f_0 has a nonsingular differential at x_0, it is approximately nonsingular linear on a small ball $B(x_0, \delta)$ in $B(1)$, and so the approximating maps $\tilde{\alpha} \circ f_i$ are approximately non singular linear on the corresponding balls $B(x_i, r_i = \delta R_i) \subset B(R_i)$.

Now, in order to show that W is a homotopy torus, we will prove that $H_i(\tilde{W}) = 0$ for $i > 0$. In fact, suppose that there is a nontrivial cycle $\mathcal{C}$ in $\tilde{W}$, and take an infinite cycle $\mathcal{C}'$ of complimentary dimension having nonzero intersection index with $\mathcal{C}$. One can assume that $\mathcal{C}$ is located near the point $f_i(x_i) \in \tilde{W}$ for large i, so that the boundary of $B(x_i, r_i)$ is sent by f_i far from $\mathcal{C}$, and the maps $f_i|_{B(r_i)}$ has $|\deg| = 1$ over $\mathcal{C}$. Thus,

$$f_i^{-1}(\mathcal{C}) \cap f_i^{-1}(\mathcal{C}') = \pm\mathcal{C} \cap \mathcal{C}' \neq 0,$$

where f_i^{-1} refers to $f_i = f_i|_{B(r_i)}$, which is impossible (since $H_*(\mathbb{R}^n) = 0$), and so $\mathcal{C} \cap \mathcal{C}' = 0$ after all. Consequently, every cycle $\mathcal{C}$ in $\tilde{W}$ of positive dimension is homologous to zero.

Example: The connected sum $T^4 \# S^2 \times S^2$ is non elliptic.

Question: Does every elliptic W admit a nonconstant quasiregular map $f_0 \colon \mathbb{R}^n \to W$? It would be interesting to prove this by finding some extremal map $f \colon \mathbb{R}^n \to W$ with maximal possible volumes $\operatorname{vol}(f|_{B(R)})$ for $R \to \infty$ and then showing that such an extremal map is quasiregular. (Such a proof could go beyond mere topology and apply to some open Riemannian manifolds W). Conversely, the Bloch–Brody principle suggests that the existence of f_0 makes W elliptic (compare $6.43\frac{2}{3}{}_+$).

Exercise: Show that the torus T^n minus a point equipped with the metric induced from T^n is nonelliptic. *Hint:* Study the isoperimetry of Euclidean spheres $S^{n-1}(R)$ mapped by Lipschitz maps to $\mathbb{R}^n$ minus a δ-dense subset, where $R \to \infty$ while δ is held constant.

Question: Is the sphere S^n minus an infinite subset ever elliptic? (The negative answer is suggested by a theorem of Rickman, which claims that every quasiregular map $\mathbb{R}^n \to S^n$ can omit only finitely many points.)

2.47. Non-Euclidean generalizations. Let us generalize the discussion above by looking at volume efficient ways of Lipschitz-wrapping up W in non-Euclidean paper. Namely, we take some (complete, open) manifold V and study λ-Lipschitz maps $f: V \to W$ satisfying $\mathrm{vol}(f|_{\Omega_i}) \sim \mu\,\mathrm{vol}(\Omega_i)$ for some exhaustion of V by suitable bounded domains Ω_i, for $i = 1, 2, \ldots$, where we wish $\mu = \mu(\lambda)$ to be relatively large, in particular > 0. If $V = \mathbb{R}^n$, then $\mu(\lambda) = \lambda^n \mu(1)$, but for general V, one may necessarily have $\mu(\lambda)/\lambda^n \to 0$ for $\lambda \to 0$, even for such simple manifolds W as the sphere S^n. On the other hand, for a fixed λ, say $\lambda = 1$, the Euclidean theory extends with minor adjustments to several classes of V. For example, if V is a nilpotent Lie group with a left-invariant Riemannian metric, then V-ellipticity of W implies that $\Gamma = \pi_1(W)$ is virtually nilpotent of growth $\Gamma \leq$ growth V. Furthermore, if V and $\pi_1(W)$ have equal growth, then the limit spaces $\lim_{\varepsilon \to 0} \varepsilon V$ and $\lim_{\varepsilon \to 0} \varepsilon \Gamma$ are bi-Lipschitz equivalent (which implies that the graded Lie algebras associated to V and Γ are isomorphic) and the universal covering $\tilde{W}$ of W is contractible.

The proofs of the results above can be easily derived along the Euclidean guidelines with the use of the Varopoulos isoperimetric inequality in nilpotent groups (see Ch. 6.E$_+$). We suggest that the reader look into this.

The "rigid" example $2.41\frac{1}{2}_+$ generalizes to those (not only nilpotent) V which admit an isometry group Γ with V/Γ compact and where the exhaustion Ω_i in question satisfies $\mathrm{vol}(\Omega_i)/\mathrm{vol}_{n-1}(\partial\Omega_i) \to \infty$ for $i \to \infty$. Here, an easy argument shows that the existence of a short map $f: V \to W$ with $\mathrm{vol}(f|_{\Omega_i}) \sim \mathrm{vol}(\Omega_i)$ implies that there exists a locally isometric (and hence, covering) map $V \to W$.

The volume rigidity theorem also generalizes, at least to the nilpotent framework, as long as one passes to Carnot–Caratheodory metrics. For example, if V and W are compact, contact C-C manifolds of dimension $n = 2m + 1$ admitting a sequence of maps $f_k: V \to W$ with $\mathrm{dil}(f_k) \to \infty$ and

$$\limsup \frac{\deg(f_k)}{(\mathrm{dil}\, f_k)^{n+1}} \geq \frac{\mathrm{vol}(V)}{\mathrm{vol}(W)},$$

where "vol" refers to the $(n+1)$-dimensional Hausdorff measure, then the universal covering of W is isometric to the n-dimensional Heisenberg group H with the standard (self-similar!) C-C metric, while W is locally isometric to H as well.

Next, one can turn to more general V, such as solvable Lie groups for instance, where one should not use exhaustions by balls $B(R_i)$ by rather by domains Ω_i having $\mathrm{vol}_{n-1}(\partial\Omega_i) = o(\mathrm{vol}(\Omega_i))$. Here, it seems more difficult to identify topological consequences of V-ellipticity since one does not know if the solvable Lie groups are distinguished by their isoperimetry (see [Va–Sa–Co] for what is known in this regard). Nevertheless, it seems not to be too hard to show that $W_0 \# V/\Gamma$ can not be V-elliptic for nonspherical W_0. (In fact, this may be expected with more general, e.g. hyperbolic manifolds V with a suitably defined V-ellipticity, compare sections $9\frac{3}{4}$ in $[\mathrm{Gro}]_{\mathrm{PCMD}}$.)

Another interesting line of thought is suggested by Lipschitz maps $f: V \to W$ for $n = \dim(V) < \dim(W)$ with

$$\int_{\Omega_i} f^*(\omega) \geq \text{const} \operatorname{vol}(\Omega_i)$$

for $i \to \infty$ and for some closed n-form ω in W. This probably can be carried over pretty far for $V = \mathbb{R}^n$, but in more general cases, we are impeded by the absence of (proofs of) suitable n-dimensional isoperimetric inequalities in non-Euclidean spaces W of dimension $> n$ (compare $[\mathrm{Gro}]_{\mathrm{CC}}$).

Chapter 3

Metric Structures on Families of Metric Spaces

A. Lipschitz and Hausdorff distance

3.1. Definition: The *Lipschitz distance* between two metric spaces X, Y, denoted $d_{\mathrm{L}}(X, Y)$, is the infimum of the numbers

$$|\log \operatorname{dil}(f)| + |\log \operatorname{dil}(f^{-1})|$$

as f varies over the set of bi-Lipschitz homeomorphisms between X and Y.

By convention, we set $d_{\mathrm{L}}(X, Y) = \infty$ whenever there is no bi-Lipschitz homeomorphism $X \to Y$. Evidently d_{L} is symmetric, satisfies the triangle inequality, and $d_{\mathrm{L}}(X, Y) = 0$ when X, Y are isometric. Conversely, we have

3.2. Proposition: *Any two compact metric spaces X, Y satisfying $d_{\mathrm{L}}(X, Y) = 0$ are isometric.*

Proof. For each integer $n > 0$, there exists a bi-Lipschitz homeomorphism $f_n : X \to Y$ such that $1 - 1/n \leq \operatorname{dil}(f_n) \leq 1 + 1/n$. Since the sequence (f_n) is equicontinuous, it contains a uniformly convergent subsequence, and the limit mapping is necessarily an isometry.

Example: If X is a compact surface of genus $g > 1$, the Lipschitz distance defines a metric on its moduli space, viewed as the space of Riemannian metrics of curvature -1, modulo the group of diffeomorphisms of X.

3.3. We now define a distance between metric spaces that are not necessarily homeomorphic, using the classical notion of Hausdorff distance (cf. [Berger]$_{\text{Cours}}$ or [Rinow], §7). Recall that for two subsets A, B of a metric space Z, this is defined as

$$d_{\mathrm{H}}^{Z}(A,B) = \inf\{\varepsilon > 0 : B \subset U_\varepsilon(A) \text{ and } A \subset U_\varepsilon(B)\},$$

where $U_\varepsilon(A) = \{z : d(z,A) \le \varepsilon\}$. It is a classical fact (and easy to check) that d_{H}^{Z} is a metric on the space of compact subsets of Z.

3.4. Definition: The *Hausdorff distance* between two metric spaces X, Y, denoted $d_{\mathrm{H}}(X,Y)$, is defined as the infimum of the numbers

$$d_{\mathrm{H}}^{Z}(f(X), g(Y))$$

for all metric spaces Z and isometric embeddings f, g of X, Y, respectively, into Z.

Remarks: (a) If X, Y are compact, then $d_{\mathrm{H}}(X,Y) < \infty$. To see this, it suffices to embed X and Y into their union, equipped with a metric d that restricts to the given metrics on X, Y and satisfies $d(x,y) = \sup(\operatorname{diam}(X), \operatorname{diam}(Y))$ for $x \in X$ and $y \in Y$.

(b) Two metric spaces with finite diameter can be separated by zero Hausdorff distance without being isometric, as, for example, in the case of $[0,1]$ and $\mathbb{Q} \cap [0,1]$.

(c) Even for very simple domains in $\mathbb{R}^n$, the Hausdorff distance is not realized by embeddings into Euclidean space. Indeed, if $A = \{a_1, a_2, a_3\}$ is an equilateral triangle of edge 1 and B consists of a point, then $d_{\mathrm{H}}(A,B) = 1/2$, although any isometric embeddings f, g of A, B into $\mathbb{R}^n$ satisfy

$$d_{\mathrm{H}}^{\mathbb{R}^n}(f(A), g(B)) > \frac{1}{\sqrt{3}}.$$

(d_+) Here is a sleeker definition of a distance. Consider surjective maps φ and ψ from the disjoint union $Z = X \sqcup Y$ to X and Y, respectively, and let d_X^* and d_Y^* be the functions on $Z \times Z$ induced from dist_X and dist_Y by these maps. Then take the sup-norm of the difference $d_X^* - d_Y^*$ on $Z \times Z$ and the infimum of this norm $\|d_X^* - d_Y^*\|_{L_\infty}$ over all possible surjective maps φ and ψ. This infimum defines a metric, say $d_{\mathrm{H}}^*(X,Y)$, which obviously satisfies $\frac{1}{2} d_{\mathrm{H}} \le d_{\mathrm{H}}^* \le 2 d_{\mathrm{H}}$. (Notice that one could minimize over all abstract sets Z and surjective maps $\varphi \colon Z \to X$ and $\psi \colon Z \to Y$.)

Roughly speaking, the convergence of a sequence (X_i) of metric spaces with respect to the Hausdorff metric corresponds to Lipschitz convergence of ε-nets in the X_i (cf. 2.14) to an ε-net in the limit space. More precisely,

3.5. Proposition:

(a) If a sequence (X_i) of metric spaces converges to X with respect to the Hausdorff distance, then for each $\varepsilon > 0$ and $\varepsilon' < \varepsilon$, every ε'-net in X with strictly positive separation is the limit of a sequence (N_i) of ε-nets in the X_i with respect to the Lipschitz distance.

(b) Conversely, if $\sup(\mathrm{diam}(X_i), \mathrm{diam}(X)) < \infty$, and if, for each $\varepsilon > 0$, there exists an ε-net of X that is the Lipschitz limit of a sequence of ε-nets $N_i \subset X_i$, then the X_i converge to X in the Hausdorff metric.

Proof. (a) By hypothesis, there exists a sequence η_i tending to zero and isometric embeddings f_i, g_i of X, X_i, respectively, into a metric space Z_i such that

$$d_{\mathrm{H}}^{Z_i}(f_i(X), g_i(X_i)) \leq \eta_i.$$

If $(x_p)_{p\in P}$ is an ε-net of X, then since $f_i(X) \subset U_{\eta_i}(g_i(X_i))$, there exist points x_p^i such that

$$d_{\mathrm{H}}^{Z_i}(f_i(x_p), g_i(x_p^i)) < \eta_i,$$

and since moreover $g_i(X_i) \subset U_{\eta_i}(f_i(X))$, the $(x_p^i)_{p\in P}$ form a $(\varepsilon + 2\eta_i)$-net in X_i. Additionally, we have

$$|d^X(x_p, x_{p'}) - d^{X_i}(x_p^i, x_{p'}^i)| \leq 2\eta_i,$$

which shows that $(x_p^i)_{p\in P}$ Lipschitz converges to $(x_p)_{p\in P}$ since $(x_p)_{p\in P}$ has a strictly positive separation.

(b) Let $(x_p)_{p\in P}$ be an ε-net of X and $(y_p^i)_{p\in P} \subset X_i$ a sequence of ε-nets that converges to X with respect to the Lipschitz distance. On the union $X \cup X_i$, we define — using the same idea as in Remark 3.4(a) — a metric d that restricts to the given metrics on X and X_i by setting

$$d(x, y) = \inf_{p\in P} d^X(x, x_p) + d^{X_i}(y_p^i, y) + \varepsilon$$

for each $x \in X$ and $y \in X_i$. The only nontrivial step in showing that d is a metric is to check the inequalitites of the form

$$d(x, x') \leq d(x, y) + d(y, x')$$

for $x, x' \in X$ and $y \in X_i$. To this end, we write $d(x,x') \leq d(x,x_p) + d(x_p,x_q) + d(x_q,x')$ and note that

$$d(x_p,x_q) \leq (1+\eta_i)\, d(y_p^i, y_q^i),$$

where $\eta_i \to 0$. For sufficiently large i, we then have

$$\begin{aligned} d(x,x') &\leq d(x,x_p) + d(y_p^i,y_q^i) + d(x_q,x') + \varepsilon \\ &\leq (d(x,x_p) + d(y_p^i,y)) + (d(y,y_q^i) + d(x_q,x')) + \varepsilon, \end{aligned}$$

which gives the desired inequality. With respect to this metric on $X \cup X_i$, we clearly have $d_{\mathrm{H}}(X,X_i) < 2\varepsilon$.

An important consequence of Proposition 3.5(a) is the following.

3.6. Proposition: *Any two compact metric spaces X, Y satisfying $d_{\mathrm{H}}(X,Y) = 0$ are isometric.*

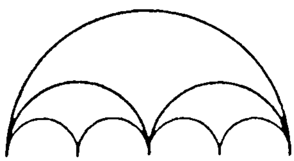
Proof. For each integer n and each $(1/n)$-net $N^{1/n}$ of Y, there exists a sequence $N_i^{2/n}$ of $(2/n)$-nets in X such that $N_i^{2/n}$ Lipschitz converges to $N^{1/n}$. For each $\varepsilon > 0$, there is an integer i_n and a mapping f_n of $N^{1/n}$ into $N_{i_n}^{2/n}$ such that both f_n and f_n^{-1} have dilatation less than $1+\varepsilon$.

By choosing the nets $N^{1/n}$ to be nested, we can use a diagonal procedure to obtain a homeomorphism $g\colon Y \to X$ having the same properties, and so the assertion is reduced to the case of Proposition 3.2.

3.7. Proposition: *If X_n are compact metric spaces and $X_n \xrightarrow{\mathrm{Lip}} X$, then $X_n \xrightarrow{\mathrm{Hau}} X$.*

Proof. The statement is an immediate consequence of Proposition 3.5(b).

The converse of this statement is certainly false, even if X and each X_n are homeomorphic, as seen in the picture below.



While we cannot expect much stability of topological properties under Hausdorff convergence, certain *metric* properties are stable. More precisely,

3.8. Proposition: *If X is a complete metric space that is the Hausdorff limit of a sequence of path metric spaces, then X is a path metric space.*

Proof. It suffices to apply Theorem 1.8, noting that the existence of approximate midpoints (Property 1.8(1)) is stable under Hausdorff convergence.

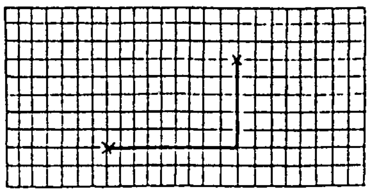
3.9. Examples: Let (X, d) be a metric space, and for $\lambda > 0$, let λX denote the metric space $(X, \lambda d)$. If $\mathrm{diam}(X) < \infty$, then λX tends to a point as $\lambda \to 0$.

If X_n is a sequence of metric spaces and if $\mathrm{diam}(X_n)$ tends to 0, then for any space X, we have $\lim_{n\to\infty}(X \times X_n) = X$. More generally, if W is the total space of a Riemannian fibration (cf. [Ber–Gau–Maz]) whose base V and fibers are compact, then by change of scale in the fibers we can construct a family g_t of metrics on W for which $\lim_{t\to 0}(W, g_t) = V$. This remark applies, for example, to the Hopf fibrations of S^{2n+1} over $\mathbb{C}\mathrm{P}^n$ and of S^{4n+3} over $\mathbb{H}\mathrm{P}^n$.

3.10. Hausdorff convergence and dimension. The preceding, rather crude examples seem to suggest that dimension is at least lower semicontinuous with respect to Hausdorff convergence. This is not the case, however, as illustrated by a sequence of increasingly fine grids in the square, which, by Proposition 3.5, converge to the square equipped with the norm

$$\|(x, y)\| = |x| + |y|.$$

One can ensure that the dimension of $X = \lim_{\mathrm{H}} X_n$ does not exceed that of the X_n by imposing the requirement that the X_n be convex (if each X_n is a subset of Euclidean space, then the proof of this assertion is elementary), or by assuming that the X_n are Riemannian manifolds with uniformly bounded curvature (the subject of Chapter 8).



3.11. The examples of sequences of Riemannian manifolds tending to manifold of smaller dimension given in 3.9 are mildly disturbing since the curvatures (seem to) tend to infinity, and one should feel relieved upon learning that this situation also arises in the case of manifolds with bounded sectional curvature.

(a_+) Berger examples

Let G be a compact Lie group with a left-invariant metric, and let H_i be a sequence of closed, connected subgroups of fixed dimension k. Then, the quotient spaces $X_i = G/H_i$ have curvatures bounded by a constant independent of i. In fact, the curvature of G/H_i, being a local quantity, depends *continuously* on the Lie algebra $\mathfrak{h} = \mathfrak{h}_i \subset \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of G. The family $\{\mathfrak{h}_i\}$ is precompact, since it is contained in the compact Grassmannian manifold $\mathrm{Gr}_k(\mathfrak{g})$. And, if $\mathfrak{h}_i \to \mathfrak{h}_\infty \subset \mathfrak{g}$, then the infinitesimal (and properly understood) local geometry of X_i converges to that of the local quotient of G by the connected subgroup H_∞ corresponding to the (limit) Lie algebra $\mathfrak{h}_\infty$. Furthermore, if the subgroup $H_\infty \subset G$ is closed, then the global quotient $X_\infty = G/H_\infty$ exists, and obviously $X_i \overset{Lip}{\to} X_\infty$. (If H_∞ is not closed, and this is the case of interest, then one may only speak of the "local quotient" obtained by dividing a small neighborhood U of the identity in G into the *connected components* of the intersections of the cosets of H_∞ with U.)

Next, for a *nonclosed* H_∞, we take its closure $\overline{H}_\infty \subset G$, which is again a connected Lie group of dimension $\dim(H_\infty) > k = \dim(H_i)$, and we observe that the manifolds $X_i = G/H_i$ Hausdorff-converge to $X_\infty = G/\overline{H}_\infty$, where $\dim(X_\infty) < \dim(X_i)$. The simplest of these are the *Berger spheres* $X_i = (SU(2) \times SU(1))/H_i$, where H_i are 1-dimensional subgroups (circles) contained in the torus $SU(1) \times SU(1)$, such that the limit H_∞ is dense in this torus. Thus, the limit manifold is the 2-sphere $S^2 = SU(2)/SU(1)$, while all X_i are 3-dimensional. In fact, we may choose H_i so that their projections to $SU(1)$ via $SU(2) \times SU(1) \to SU(1)$ are injective, and then, clearly, all X_i are simply connected and in fact are diffeomorphic to S^3. Here, the local geometries of X_i converge to that of the product $S^2 \times \mathbb{R}$, while globally the X_i converge to S^2. Each $X_i \simeq S^3$ is Hopf-fibered over $S^2 = SU(2)/SU(1)$, with the circle fibers of length $\ell_i \to 0$. If we restrict the fibrations X_i to a fixed disk $D^2 \subset S^2$, then the resulting submanifolds $Y_i \subset X_i$ are diffeomorphic to $D \times S^1$, and their universal coverings, diffeomorphic to $D \times \mathbb{R}$, Lipschitz-converge to the actual *Riemannian product* $D \times \mathbb{R}$, as the group-theoretic picture above shows.

Exercise: Exhibit a sequence of spaces $X_i = (SU(2) \times SU(2))/H_i$ which Hausdorff-converges to $S^2 \times S^2$, while their infinitesimal geometries converge to that of $S^2 \times S^3$. (One can even make all X_i diffeomorphic to $S^2 \times S^3$.)

Remark: The above collapse of X_i is due to the nonvanishing of $\pi_2(S^2 \times S^3)$. In fact, if manifolds X_i with $\pi_1(X_i) = 0$ have the property that their infinitesimal geometries converge to that of some X with $\pi_1(X) = \pi_2(X) = 0$, then $X_i \xrightarrow{\text{Hau}} X$ (see $[\text{Che}]_{\text{CFT}}$, $[\text{Gro}]_{\text{Stab}}$).

Exercise: Let ∂ be a nonvanishing Killing field on a Riemannian manifold (X, g), where "Killing" signifies that ∂ locally integrates to an *isometric* action of $\mathbb{R}$ on X, or, equivalently, that the Lie derivative ∂g vanishes. Denote by g_ε the Riemannian metrics on X defined by the conditions that $g_\varepsilon(\partial, \partial) = \varepsilon^2 g(\partial, \partial)$ and $g_\varepsilon = g$ on the vectors τ orthogonal to ∂, i.e., $g_\varepsilon(\tau, \tau') = g(\tau, \tau')$ for all τ, τ' satisfying $g(\partial, \tau) = 0$, $g(\partial, \tau') = 0$. Show that the curvature of g_ε remains bounded as $\varepsilon \to 0$ on each compact subset in X. In fact, that local geometry of $X_\varepsilon = (X, g_\varepsilon)$ converges to that of $Y \times \mathbb{R}$, where Y denotes the local quotient $X/\mathbb{R}$ for the local action of $\mathbb{R}$ on X given by ∂. Next, assume that X is a closed manifold and show that X_ε Hausdorff-converges as $\varepsilon \to 0$ to X/H, where H denotes the closure of the $\mathbb{R}$-groups defined by ∂ in the full isometry group $\text{Iso}(X)$ (compare Ch. 8_+A). Apply this to the sphere S^{2k+1} with the field ∂ tangent to the Hopf fibers and show again that the ε-shrinking of (the spherical metric along) these fibers leads to the Hausdorff convergence $S_\varepsilon^{2k+1} \xrightarrow{\text{Hau}} \mathbb{C}\text{P}^k$ with local geometries Lipschitz convergent to $\mathbb{C}\text{P}^k \times \mathbb{R}$. (Compare [Che–Ebin], p. 70, [Che–Gro], and Ch. 8_+A).

(*b*) Flat manifolds

In terms of the Hausdorff distance, the Mahler compactness theorem (cf. [Cass], p. 137) states that the closure of the space of flat n-tori $\mathbb{T}^n = \mathbb{R}^n/\Lambda$ consists of all flat tori of dimension less than or equal to n. The limit of a sequence $\mathbb{T}_i^n = \mathbb{R}^n/\Lambda_i$ has dimension n if and only if the numbers $c(\Lambda_i) = \inf\{|\lambda| : \lambda \in \Lambda_i \setminus \{0\}\}$ (in Riemannian terms, $c(\Lambda_i)$ is twice the radius of injectivity of $\mathbb{T}_i^n$) has a strictly positive lower bound.

If we now consider the closure of the space of flat n-manifolds, we obtain certain quotients $\mathbb{R}^k/\Gamma$, where $k \leq n$ and Γ is a discrete group of isometries of $\mathbb{R}^k$ that may have fixed points. This phenomenon already occurs for $n = 2$.

Moreover, every singular path metric space of the form $\mathbb{R}^k/\Gamma$ is the limit of a sequence of flat manifolds of dimension $n(\Gamma)$. To see this, it suffices to construct (by elementary means) an extension of Γ by a lattice that acts without fixed points on some $\mathbb{R}^n$ and then to tend the lengths of its generators to zero.

These examples illustrate that in order to ensure that a sequence of Riemannian n-manifolds converges to an n-manifold, a reasonable hypothesis is provided by a uniform lower bound on the injectivity radii. In Chapter 8, we see several results of this sort, as well as more precise information as to the singularities that can appear in the limit space when the injectivity radii tend to zero.

$3.11\frac{1}{2}{}_+$ The Hausdorff moduli space. The Hausdorff metric allows us to bring all metric spaces together. For example, the set $\mathcal{X}_c$ of isometry classes of compact metric spaces has an interesting geometry with respect to the Hausdorff metric $d_{\mathrm{H}}(X,Y)$ for $X,Y \in \mathcal{X}_c$. It is easy to see that $\mathcal{X}_c$ *is a complete, contractible metric space, in which the (isometry classes of) finite spaces X form a dense subset.* (It is not locally compact; in fact, the space of continuous functions on $[0,1]$ admits a topological embedding into $\mathcal{X}_c$.) One can also make a moduli space of isometry classes of noncompact spaces X lying within a finite Hausdorff distance from a given X_0, e.g., $X_0 = \mathbb{R}^n$. Such moduli spaces are also complete and contractible.

On the functorality of $X_i \rightsquigarrow \lim_{\mathrm{H}} X_i$. The limit space, whenever it exists, is defined only up to isometry, and morphisms (e.g., isometric or, say, Lipschitz maps) $X_i \to X'_i$ do not pass, *a priori*, to any limit maps $\lim_{\mathrm{H}} X_i \to \lim_{\mathrm{H}} X'_i$. But this can be remedied by a more functorial definition of the limit (see 3.29). To clarify the idea of functorality, we suggest the following:

Exercises: (a) Suppose that $X_i \xrightarrow{\mathrm{Hau}} X$, $X'_i \xrightarrow{\mathrm{Hau}} X'$ and there exist isometric embeddings $X_i \to X'_i$ for all $i = 1, 2, \ldots$. Then there exists an isometric embedding $X \to X'$, provided that X' is *compact.*

(b) If $X_i \xrightarrow{\mathrm{Hau}} X$ and all X_i are *homogeneous*, (i.e., the isometry group of X_i is transitive on X_i for every i), then X is also homogeneous provided that it is a locally compact space, decomposable into a union of countably many compact subsets.

(c) Let $X_i \xrightarrow{\mathrm{Hau}} X$, $X'_i \xrightarrow{\mathrm{Hau}} X'$, and suppose that there are short surjective maps $X_i \to X'_i$ for all i. Then there is such a map $X \to X'$, provided that X and X' are compact.

$3.11\frac{2}{3}{}_+$ Uryson spaces. A metric space Y isometrically containing another space X is called a *U_d-extension of X* if Y isometrically contains every (iso-)metric extension X_* of X by a single point, $X_* = X \sqcup \{x_*\}$, provided that $\mathrm{dist}_*(x_*, x) \le d$ for all $x \in X$. This means that for an arbitrary function $d_*(x) \le d$ on X which may serve as a distance function, $d_*(x) = \mathrm{dist}_*(x_*, x)$ for some metric dist_* on $X_* = X \sqcup \{x_*\}$, there exists a point $y = y_{d_*} \in Y$ such that $\mathrm{dist}_Y(y, x) = d_*(x)$ for all $x \in X$.

Example: Let Y be the space of 1-Lipschitz maps $y : X \to [0, d]$ with $\mathrm{dist}(y, y') = \sup_{x \in X} \mathrm{dist}(y(x), y'(x))$. This is called the *$K_d$-extension* of X, since X is isometrically embedded into Y by the *Kuratowski map* (see [Kur]), which sends each point $x_0 \in X$ to the distance function $y(x) = \mathrm{dist}_X(x, x_0)$. Clearly, this Y has the U_d-property, since one can take $y_{d_*} = d_*(x) \in Y$.

A metric space Z is called *U_d-universal* if it U_d-extends every *finite* subset $F \subset Z$.

(1) Example: Start with an arbitrary metric space X_0 and define $X_0 \subset X_1 \subset X_2 \subset \cdots$, where each X_i is the K_d-extension of X_{i-1}. Then the union $X_\infty = \bigcup_{i=0}^{\infty} X_i$ is U_d-universal, since every finite subset is contained in some X_{i_0} which is U_d-extended by X_{i_0+1}.

Here is an obvious, useful

(2) Lemma: *Let Z be a U_d-universal space and let A be a countable metric space. Then every isometric embedding of a finite subset of A into Z, say $A \supset F_0 \xrightarrow{\varphi} Z$, extends to an isometric embedding $A \to Z$, provided that* $\mathrm{diam}(A) \leq d$.

Proof. Write $A = F_0 \cup \{a_1, a_2, \dots\}$ and keep on extending isometric embeddings from F_i to $F_{i+1} = F_i \cup \{a_{i+1}\}$.

(3) Corollary: *If Z is complete as well as U_d-universal, then the isometric extension above is possible from finite subsets to arbitrary second countable metric spaces $X \supset F_0$, provided that* $\mathrm{diam}(X) \leq d$. *In particular, every second countable metric space of diameter $\leq d$ isometrically embeds into Z.*

Proof. Use a countable dense subset $A \subset X$ containing F_0.

(4) Theorem (Uryson, see [Ury]): *Every two Polish (i.e., complete, second countable) U_d-universal spaces Z and Z' are isometric.*

Proof. If suffices to construct increasing sequences of finite subsets $\cdots \subset F_i \subset F_{i+1} \subset \cdots$ in Z and $\cdots \subset F'_i \subset F'_{i+1} \subset \cdots$ for $i = 1, 2, \dots$ and isometries $F_i \overset{I_i}{\longleftrightarrow} F'_i$, where the I_{i+1} extend the I_i, where we want the (necessarily isometric) unions $F_\infty = \bigcup_{i=1}^{\infty} F_i$ and $F'_\infty = \bigcup_{i=1}^{\infty} F'_i$ to be dense in Z and Z', respectively. These F_i are constructed in steps by induction on i. Given F_i and F'_i for an *even* i, we add an arbitrary point $z_{i+1} \in Z$

to F_i and then extend the isometry $F_i \xrightarrow{I_i} F_i'$ to $F_{i+1} = F_i \cup \{z_{i+1}\}$ using the U_d-universality of Z'. Similarly, on each *odd* step, we add z_{j+1}' to F_j' and extend the isometry $F_j' \xrightarrow{I_j} F_j$. Thus, we can arrange our sets to be eventually dense in Z and Z'.

(5) Corollary: *For each $d \in [0, \infty]$, there exists a unique, up to isometry, Polish U_d-universal space, called the Uryson space U_d.*

In fact, the uniqueness is shown above, while the existence is obtained with the K_d-extensions. Namely, the metric completion $\overline{X}_\infty$ of the union $X_\infty = \bigcup_{i=0}^{\infty} X_i$ is U_d-universal for an arbitrary starting compact metric space X_0. Here one should note that

(a) if X_0 is compact, then so are its K_d-extensions for $d < \infty$ (by the Ascoli theorem) and consequently the union X_∞ and its completion $\overline{X}_\infty$ are second countable,

(b) the U_d-property of a space implies that of its metric completions. In fact, *if a complete metric space Y admits a dense U_d-universal subspace, then Y is U_d-universal.*

Proof. It is obvious that Y is *approximately* U_d-universal. This means that for each finite subset $F \subset Y$, every $\varepsilon > 0$ and an arbitrary potential distance function d_* on F, there exists $y \in Y$ such that

$$|\operatorname{dist}(y, f) - d_*(f)| \leq \varepsilon$$

for all $f \in F$. (This is seen by ε-approximating F by a finite subset in the dense subspace and appealing to the U_d-universality of the latter.) But since Y is complete, *the approximate U_d-universality implies the actual one* as follows. We assume without loss of generality that $d_*(f) \geq \delta > 0$ for all $f \in F$ and we find points $y_1, \ldots, y_i, \ldots$ in Y by induction on i such that

(i) $|\operatorname{dist}(f, y_i) - d_*(f)| \leq \delta 2^{-i}$, and

(ii) $\operatorname{dist}(y_j, y_{j+1}) \leq \delta 2^{-j+2}$ for $j = 2, \ldots, i$.

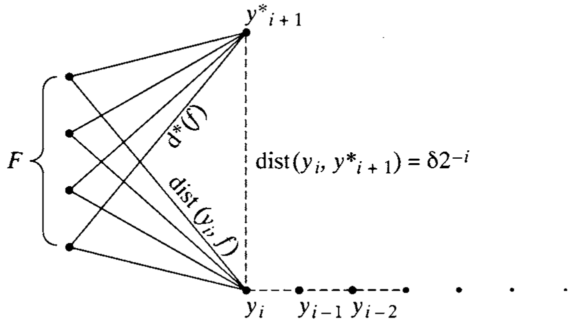
In fact, assume we have already found such $y_1, \ldots, y_i$, let $F_i = F \cup \{y_1, \ldots, y_i\}$, adjoin an extra (abstract) point, say y_{i+1}^* to F_i, and define a metric dist_* on $F_{i+1}^* = F_i \sqcup \{y_{i+1}^*\}$, such that

$$\operatorname{dist}_* |_{F_i} = \operatorname{dist}_Y |_{F_i}, \tag{$*$}$$

$$\operatorname{dist}_*(f, y_{i+1}^*) = d^*(f), \tag{$**$}$$

$$\operatorname{dist}_*(y^*_{i+1}, y_i) = \delta 2^{-i}. \qquad (***)$$

To do this, we observe that the conditions $(*) - (***)$ define *metrics* on $F_i \subset F^*_{i+1}$ and on $F \cup \{y_i, y^*_{i+1}\} \subset F^*_{i+1}$, where the triangle inequality for the latter trivially follows from (i).



Since these metrics agree on the intersection

$$F_i \cap (F \cup \{y_i, y^*_{i+1}\}) = F \cup \{y_i\},$$

there exists a metric dist_* on $F^*_{i+1} = F_i \cup (F \cup \{y_i, y^*_{i+1}\})$ extending these metrics on F_i and $F \cup \{y_i, y^*_{i+1}\}$.

Now, we use the approximate U_d-property and extend the embedding $F_i \subset Y$ to an ε_i-isometric embedding $F^*_{i+1} \to Y$ with a sufficiently small ε_i, say $\varepsilon_i \leq \delta 2^{-i-1}$. This means that the distance function from y_{i+1} (coming from y^*_{i+1}) to each point in $F_i \subset Y$ is ε_i-close to the distance dist_* from y^*_{i+1} to this point; therefore, $(**)$ yields

$$|\operatorname{dist}(f, y_{i+1}) - d^*(f)| \leq \varepsilon_{i+1} \leq \delta 2^{-i-1},$$

for $f \in F$, while $(***)$ implies that

$$\operatorname{dist}(y_{i+1}, y_i) \leq \delta 2^{-i} + \varepsilon_{i+1} \leq \delta 2^{-i+1}.$$

This tells us that the relations (i) and (ii) are extended from i to $i+1$ and thus we get our infinite sequence $y_1, \dots, y_i, y_{i+1}, \dots$ satisfying (i) and (ii) for $i = 1, 2, \dots$. This converges (due to (ii) and the completeness of Y) to the required (because of (i)) y.

Remark: My attention was recently drawn to Uryson's spaces by A. Vershik, who warmed up my enthusiasm by suggesting the following.

I. A possible relation between the Uryson spaces U_d and the Hausdorff moduli space.

II. Identifying the isometry group of U_d by some intrinsic (universal) property (compare Exercise (a) below).

III. Raising the question of the existence of nice (?) measures on U_d and/or finding a meaningful counterpart of U_d in the category of metric spaces with measures.

Exercises: (a) By adjusting the proof of (4), show that the isometry group is transitive on U_d. Moreover, it is transitive on mutually isometric r-tuples of points in U_d (as are the isometry groups of $\mathbb{R}^n$ and S^n). Notice that U_d contains proper subsets isometric to itself and so not every isometry between subsets in U_d isometrically extends to U_d, but this is true for compact subsets in U_d (see below).

(b) Show that the U_d-universality for finite subsets in a complete space X implies the U_d-universality for all compact subsets in X. (*Hint:* Argue as in the derivation of U_d from approximate U_d.)

(b') Show that for every pair of isometric compact subsets A and B in U_d, there exists an isometry of U_d moving A to B which agrees with a given isometry $A \to B$.

(c) Construct the Uryson space $U_{d=\infty}$ by using $\cdots \subset X_i \subset X_{i+1} \subset \cdots$, where X_{i+1} is the K_i-extension of X_i.

(d) Use homotheties in the spaces of functions $y(x) \mapsto ty(x) \in X_{i+1}, x \in X_i$ and thus obtain homotheties of $U_d = X_\infty$, say $\varphi_t : U_d \to U_d$, shrinking U_d to a (given) point for $t = 0$ and having the images $\varphi_\varepsilon(U_d) \subset U_d$ isometric to $U_{td} \subset U_d$. In particular, U_d is contractible. It is also a path metric space with infinitely many shortest paths between every pair of points (Uryson). And the homothety above selects particular geodesics $t \mapsto \varphi_t(x)$, $t \in [0,1]$, from all $x \in U_d$ to $x_0 = \varphi_0(U_d)$.

(e) (Uryson) Show that the spheres of radius d' in U_d are isometric to $U_{d'}$.

Question to the reader: Do you find this huge U_d ugly, or, on the contrary, quite beautiful? (A couple of good theorems would make it more respectable in any case.)

There is a discrete version of U_d which may feel more palatable to some people. Namely, we stick to *countable* spaces X with *integer* valued metrics dist : $X \times X \to \mathbb{Z}_+$. The corresponding universal space, say $U_d^{\mathbb{Z}}$ comes as $X_\infty = \bigcup_{i=0}^{\infty} X_i$, where X_0 is finite, and X_{i+1} consists of integer-valued (1-Lipschitz if you wish) functions $X_i \to [0,d]$. (We need a variable d, say $d = i$, in order to arrive at $U_\infty^{\mathbb{Z}}$.)

The isometry groups of $U_d^{\mathbb{Z}}$ is uncountable for all $d \geq 1$. In fact, it contains the full permutation group of a countable set. To see this, observe that the construction of X_∞ can start with an arbitrary countable X_0 if X_{i+1} consists, by definition, of $\mathbb{Z}_+$-valued functions y on X_i, such that $y(x) = \mathrm{dist}_{X_i}(x_0, x)$ for some $x_0 = x_0(y) \in X_i$, and *all but finitely many* $x \in X_i$ (which keeps X_{i+1} countable). Thus, the isometry group of X_0 acts on X_∞ and one may take X_0 with $\mathrm{dist}(x, x') = 1$ for all $x \neq x'$. (Or one can play with $X_0 = X_\infty$.)

It would be interesting to find a nice embedding $\mathrm{Isom}\, U_d^{\mathbb{Z}} \subset \mathrm{Isom}\, U_d$ and to look at the quotient space. Also, it seems logical to represent U_∞ as some kind of limit of $\varepsilon U_\infty^{\mathbb{Z}}$, but I do not know if this makes sense (compare $3.29\frac{1}{2}{}_+$).

Finally, we can ask for which geometric and combinatorial structure (e.g., colored graphs, hypergraphs, etc.) Uryson type universal objects exist. One may try here a construction of such a space by gluing together representatives of our spaces along isomorphic subsets in the spirit of the definition of the Hausdorff moduli space. For example, let us construct $U_\infty^{\mathbb{Z}}$ in this way. First, we enumerate all *finite* $\mathbb{Z}_+$-metric spaces, say they are $X_1, X_2, \ldots$. Then we construct finite spaces Y_{i+1} by induction as follows. Take Y_i and consider all isometries between subsets $Y' \subset Y_i$ and subsets in the spaces $X_1, \ldots, X_{i+1}$. Enumerate these isometries by $\nu = 1, 2, \ldots, N = N(i)$ and glue the spaces $X_{j(\nu)}$ to Y_i successively according to these isometries I_ν. Thus, we get $Y_i \subset Y_i^1 \subset \cdots \subset Y_i^\nu \subset \cdots Y_i^N = Y_{i+1}$, where $Y_i^{\nu+1}$ is obtained from Y_i^ν by gluing $X_{j(\nu)}$ to it along some $X'_\nu \overset{I_\nu}{\leftrightarrow} Y'_\nu \subset Y_i \subset Y_i^\nu$. Clearly, the union $Y_\infty = \bigcup_{i=0}^\infty Y_i$ is U_∞-universal and thus isometric to $U_\infty^{\mathbb{Z}}$.

Exercises: (*a*) Make a similar construction for U_∞ by applying some gluings to a (universal) *continuous family* of compact metric spaces (similar to the Hausdorff moduli space).

(*b*) Consider the space of compact subsets in U_∞ with the Hausdorff metric, call it U_∞^+, and show that $U_\infty^+ / \mathrm{Isom}(U_\infty)$ is isometric to the Hausdorff moduli space $\mathcal{H}_c$ (thus vindicating Vershik's idea). The above makes U_∞^+ a good candidate for the universal ramified covering $\tilde{\mathcal{X}}_c$ of $\mathcal{X}_c$ indicated below.

(*c*) Show that every second countable metrizable group G admits a free isometric action on U_∞.

(*d*) Vershik also suggested that Uryson spaces come from suitable measures on the space M_∞ of infinite symmetric matrices viewed as random metrics on $\mathbb{N}$ (compare 3.39(4)). Another way to put it is that a (sufficiently) random metric space is U_∞. To make this precise, consider the space M_∞ of metrics on $\mathbb{N} = \{1, 2, \dots\}$ and let μ be a probability measure on M_∞ with the following properties.

(1) μ is invariant and ergodic under the group of permutations of $\mathbb{N}$ with finite support.

(2) the natural projection (push-forward) of μ to the space M_r of metrics on $\{1, 2, \dots, r\}$ is a Borel measure whose support equals all of M_r for every $r = 1, 2, \dots$. Then almost all (for this μ) metrics $m \in M_\infty$ are approximately U_∞-universal, and hence the completion of $(\mathbb{N}, m)$ is isometric to U_∞ for almost all $m \in M_\infty$. Checking everything is just another exercise for the reader.

(e) Define U_d-universality in the category of spaces with a free proper isometric action of a locally compact, second countable group G. Show the existence of these spaces and prove that they are all isometric to U_∞.

(f) Return to the construction of U_d as the tower of the K_d-extensions and find other (stronger) structures on these (besides the metric) with the transitive automorphism group. (I have not fully solved this problem.)

3.11$\frac{3}{4}$$_+$ Orbit structures in the moduli spaces. Let us look at the Hausdorff moduli space $\mathcal{X}_c$ in a small neighborhood U of some space $X_0 \in \mathcal{X}_c$ which has a nontrivial isometry group G_0. If G_0 is a finite group, then a (sufficiently small!) U can be rather naturally represented as $\tilde{U}/G_0$, where $\tilde{U}$ consists of isometry classes of suitably marked spaces X close to X_0 as explained in the simplest case in 3.32. Ideally, we want to have some natural "universal branched covering" $\tilde{\mathcal{X}}_c \to \mathcal{X}_c$ which has "Iso(X)-ramification" at each $X \in \mathcal{X}_c$ such that (the local isometries of) $\tilde{\mathcal{X}}_c$ keep track of isometric (even better, Lipschitz) maps $X \to X'$ for all $X, X' \in \mathcal{X}_c$. But I do not know if there is a suitable realization of this idea, for example, in the spirit of the theory of stacks, giving some $\tilde{\mathcal{X}}_c$ smaller than U_∞.

Motivating Example: Consider the space of all Riemannian manifolds diffeomorphic to a fixed V. Then this space $\mathcal{X} = \mathcal{X}_V$ equals $\mathcal{G}/\operatorname{Diff}(V)$, where $\mathcal{G}$ denotes the space of all Riemannian metrics on V. Consider the subset $\mathcal{X}_0 = \mathcal{G}_0/\operatorname{Diff}(V)$ corresponding to Riemannian manifolds (V, g) with *trivial* isometry groups and equip $\mathcal{X}_0$ with the Lipschitz (instead of Hausdorff) metric. Then $\mathcal{G}_0 \to \mathcal{X}_0$ constitutes a principle $\operatorname{Diff}(V)$-fibration, and $\mathcal{G}_0$ is a contractible space for $\dim(V) \geq 2$. Thus, the fundamental group Γ of $\mathcal{G}_0$ equals $\pi_0(\operatorname{Diff}(V))$. So, we can define $\tilde{\mathcal{X}}$ as the metric completion of the universal covering $\tilde{\mathcal{X}}_0$ and observe that Γ isometrically acts on $\tilde{\mathcal{X}}$ with $\tilde{\mathcal{X}}/\Gamma = \mathcal{X}$.

B. The noncompact case

3.12. When dealing with noncompact spaces, it is convenient to work in the category of pointed spaces. The Lipschitz and Hausdorff distances between pointed metric spaces are defined as in 3.1 or 3.3, except that only pointed bi-Lipschitz homeomorphisms and isometric embeddings are allowed.

The connection between convergence and pointed convergence can be described as follows:

3.13. Proposition: *Let (X_n, a_n) and (X, a) be complete, locally compact path metric spaces. In order for* $\lim_{\mathrm{H}}(X_n, a_n) = (X, a)$, *it is necessary and sufficient that the sequence of closed balls $B(a_n, r)$ Hausdorff converges to $B(a, r)$ uniformly in r.*

Proof. Apply "abstract nonsense" (à la Zorn); in any case, we will not have use for this proposition in the remainder of the book.

The uniformity in the r condition is actually unreasonably strong, since it does not even allow the sphere $S^n(r)$ to converge to $\mathbb{R}^n$ as $r \to \infty$. For complete, locally compact path metric spaces, we will use the following:

3.14. Definition: We will say that (X_n, a_n) tends to (X, a) in the Hausdorff (resp. Lipschitz) sense if, for each $r > 0$, the ball $B(a_n, r + \varepsilon_n)$ Hausdorff (resp. Lipschitz) converges to $B(a, r)$ for some sequence (ε_n) tending to 0.

For a pointed space (X, a), the study of $\lambda X = (X, \lambda\,\mathrm{dist}_X)$ (cf. 3.9) for large λ amounts to looking at a neighborhood of a under a magnifying glass. For example,

3.15. Proposition: *Let (V, g) be a Riemannian manifold with g continuous. For each $v \in V$, the spaces $\lambda(V, v)$ Lipschitz converge as $\lambda \to \infty$ to the tangent space $(T_v V, 0)$ with its Euclidean metric g_v.*

Proof$_+$. Start with a C^1 map $(\mathbb{R}^n, 0) \to (V, v)$ whose differential is isometric at 0. The λ-scalings of this provide almost isometries between large balls in $\mathbb{R}^n$ and those in λV for $\lambda \to \infty$.

Remark: In fact, we can *define* Riemannian manifolds as locally compact path metric spaces that satisfy the conclusion of Proposition 3.15.

Evidently, it is more interesting to tend λ to zero, in which case the limit space (if it exists) will depend on the global properties of the original space (cf. [Pr]). A trivial, yet fundamental remark is that for each metric d on $\mathbb{R}^n$ arising from a norm, we have $\lim \lambda\mathbb{R}^n = \mathbb{R}^n$ as $\lambda \to 0$. This is a special case of the following observation:

3.16. Proposition: *Let V be a path metric space whose fundamental group is isomorphic to $\mathbb{Z}^n$, and let $\tilde{V}$ be its universal cover, equipped with the natural length structure. For $\tilde{v} \in \tilde{V}$, the sequence of homothetic spaces $2^{-k}(\tilde{V}, \tilde{v})$ Hausdorff converges to $(\mathbb{R}^n, 0)$ equipped with a metric arising from a (possibly noneuclidean) norm.*

The norm $\|\cdot\|$ described in the proposition is obtained as follows. By Proposition 2.11, there exists a limit-norm, denoted $\|\cdot\|^{lim}$, on the group $\pi_1(V, v_0)$ having the property that if $q \in \mathbb{Z}$, then $\|q\alpha\|^{lim} = |q|\|\alpha\|^{lim}$. By means of an isomorphism, we equip the subgroup $\mathbb{Z}^n$ of $\mathbb{R}^n$ with this norm and extend by homogeneity to $\mathbb{Q}^n$ and then by continuity to $\mathbb{R}^n$.

Proof. Choose an isomorphism $\varphi\colon \mathbb{Z}^n \to \Gamma = \pi_1(V, v_0)$ and let $\alpha_1, \dots, \alpha_n$ be the image of the canonical basis of $\mathbb{Z}^n$. We will use the same notation for an element of Γ and the deck transformation of $\tilde{V}$ that it induces. Fix $\varepsilon > 0$ and an integer k_0 such that

$$2^{-k_0}\left(\sum_{i=1}^{n} \operatorname{length}(\alpha_i) + \operatorname{diam}(V)\right) < \varepsilon$$

and such that the subgroup $2^{-k_0}\mathbb{Z}^n$ is an ε-net of $(\mathbb{R}^n, \|\cdot\|)$. Set $N = 2^{-k_0}\mathbb{Z}^n$ and for $k \geq k_0$, define $\varphi_k\colon N \to \tilde{V}$ by

$$\varphi_k(2^{-k_0}(x_1, \dots, x_n)) = \prod_{i=1}^{n} (\alpha_i^{x_i})^{2^{k-k_0}} \cdot \tilde{v}.$$

Then $N_k = \varphi_k(N)$ is an ε-net in $2^{-k}\tilde{V}$; indeed, if $\alpha \in \Gamma$, then

$$\begin{aligned} d(\alpha\tilde{v}, N_k)_{\tilde{V}} &= d(\tilde{v}, \alpha^{-1}N_k)_{\tilde{V}} \\ &\leq d\left(\tilde{v}, \prod_{i=1}^{n} \alpha_i^{(2^{k-k_0}[x_i 2^{k_0-k}] - x_i)} \cdot \tilde{v}\right)_{\tilde{V}} \\ &\leq 2^{k-k_0}\left(\sum_{i=1}^{n} \operatorname{length}(\alpha_i)\right), \end{aligned}$$

so that if $\tilde{x} \in \tilde{V}$ we have

$$\begin{aligned} d(\tilde{x}, N_k)_{\tilde{V}} &\leq d(x, \tilde{\Gamma v})_{\tilde{V}} + d(\Gamma\tilde{v}, N_k)_{\tilde{V}} \\ &\leq \operatorname{diam}(V) + 2^{k-k_0}\left(\sum_{i=1}^{n} \operatorname{length}(\alpha_i)\right) \\ &< 2^k\varepsilon, \end{aligned}$$

whence $d(\tilde{x}, N_k)_{2^{-k}\tilde{V}} < \varepsilon$. Finally, we verify that if $x, y \in N$, then the sequence $d(\varphi_k(x), \varphi_k(y))_{2^{-k}\tilde{V}}$ tends to $\|y-x\|$ as $k \to \infty$. Let $x = 2^{-k_0}x', y = 2^{-k_0}y'$ so that

$$\begin{aligned} d(\varphi_k(x), \varphi_k(y))_{2^{-k}\tilde{V}} &= 2^{-k} d\left(v, \left(\prod_{i=1}^{n} \alpha_i^{x'_i - y'_i}\right)^{2^{k-k_0}} \cdot \tilde{v}\right)_{\tilde{V}} \\ &\to 2^{-k_0}\left\|\prod_{i=1}^{n} \alpha_i^{x'_i - y'_i}\right\|^{lim} \\ &= \|x-y\| \end{aligned}$$

as $k \to \infty$. In other words, $\operatorname{dil}_{(x,y)}(\varphi_k) \to 1$ as $k \to \infty$, and so, since $N \cap B(0,r)$ is finite, we have

$$\operatorname{dil}(\varphi_k|_{N\cap B(0,r)}) \to 1 \quad \text{and} \quad \operatorname{dil}(\varphi_k^{-1}|_{N_k \cap B(v,r)}) \to 1,$$

which proves that the restrictions of the N_k to radius-r balls Lipschitz converge to $N \cap B(0,r)$. Thus, $2^{-k}V$ converges to $(\mathbb{R}^n, \|\cdot\|)$.

Remarks$_+$: (a) What one truly needs is not $\pi_1(V) = \mathbb{Z}^n$ but rather a surjective homomorphism $\pi_1(V) \to \mathbb{Z}^n$ giving us a Galois $\mathbb{Z}^n$-cover $\tilde{V}$ of V to which the proposition above applies.

(b) Typically, there is *no Lipschitz* convergence $2^{-k}\tilde{V}$ to $\mathbb{R}^n$, since $\tilde{V}$ may even be nonhomeomorphic to $\mathbb{R}^n$, as in the case of the $\mathbb{Z}^n$-covering of the connected sums $V = T^n \# V_0$, where V_0 is not a homotopy sphere. In fact, if the convergence is Lipschitz, then obviously $\tilde{V}$ is isometric to $(\mathbb{R}^n, \|\cdot\|^{lim})$, and since $\tilde{V}$ is Riemannian, it is isometric to the Euclidean space $\mathbb{R}^n$.

(c) **D. Burago's theorem.** *$\tilde{V}$ lies within a finite Hausdorff distance from* $(\mathbb{R}^n, \|\cdot\|^{lim})$ (see [Bur]$_{\text{PM}}$ and $2.C'_1$ in [Gro]$_{\text{AI}}$).

The key step in the proof is the following property of the lengths of the shortest loops in (V, v_0) representing the elements α of $\mathbb{Z}^n$ (where $\pi_1(V) = \mathbb{Z}^n$, or, more generally, we are given an epimorphism $\pi_1(V) \to \mathbb{Z}^n$)

$$\operatorname{length}(\alpha) \leq \frac{1}{2}\operatorname{length}(2\alpha) + \text{const}\ .$$

Problem$_+$: What are the possible limit norms for Riemannian manifolds $\tilde{V}$ with $\mathbb{Z}^n$-periodic metrics? This question has been addressed by V. Bangert and more recently by Burago and Ivanov, who conjecture that these norms cannot be smooth unless $\tilde{V} = \mathbb{R}^n$. However, we don't even know if there are non-Euclidean $\tilde{V}$ with Euclidean limit norms.

3.17. If $X = \lim_{\lambda \to 0} \lambda V$, then clearly X is conical, i.e., isometric to λX for all $\lambda \neq 0$. The simplest path metric space after $\mathbb{R}^n$ having this property is the 3-dimensional Heisenberg group $\mathbb{H}^3$, equipped with its Carnot metric (cf. 1.18).

If a, b, c are generators of its Lie algebra $\mathfrak{h}_3$ such that $[a, c] = [b, c] = 0$ and $[a, b] = c$, and f_λ is defined by $f_\lambda(a, b, c) = (\lambda a, \lambda b, \lambda^2 c)$, then one can see that for $x, y \in \mathbb{H}^3$ and $\lambda > 0$, we have $d(f_\lambda(x), f_\lambda(y)) = \lambda d(x, y)$. The limit space is given by

3.18. Proposition: *Let g be the left-invariant Riemannian metric on $\mathbb{H}^3$ with respect to which a, b, c are orthonormal. As $\lambda \to 0$, the metric λg tends to the Carnot metric defined by g and the left-invariant plane field spanned by left-translates of $\{a, b\}$. Consequently, the Carnot space $(H^3, \lim_{\lambda \to 0} \lambda g)$ is indeed conical, and $(H^3, \lambda g)$ Hausdorff-converges to $(H^3, \lim_{\lambda \to 0} \lambda g)$.*

Proof. Apply the same principle as in Proposition 3.15, using the fact that if the elements of $\mathbb{H}^3$ are represented by matrices of the form

$$\begin{pmatrix} 1 & u & w \\ 0 & 1 & v \\ 0 & 0 & 1 \end{pmatrix},$$

then the metric g is given by $du^2 + dv^2 + (dw - u\,dv)^2$.

3.18$\frac{1}{2}_+$ The Pansu theorem. Let V be a Riemannian manifold admitting a discrete isometric action of a nilpotent group Γ with compact quotient. It is easy to see that the family λV is precompact (compare Ch. 5A–B), and thus there exists a (pointed) Hausdorff limit $X = \lambda_i V$ for a sequence $\lambda_i \to 0$. Such an X is, in fact, a nilpotent Lie group equipped with a Carnot metric which admits a nontrivial self-similarity, i.e., an isometry $X \to \alpha X$ for $\alpha \neq 1$ which is moreover a group isomorphism. All this is easy; what is nontrivial is the following

Convergence theorem (see [Pan]$_{\mathrm{CBC}}$). *The spaces λV converge to X as $\lambda \to 0$ in the pointed Hausdorff topology.*

C. The Hausdorff–Lipschitz metric, quasi-isometries, and word metrics

The Lipschitz metric introduced earlier is too restrictive because of the simple fact that two spaces must be homeomorphic in order for their Lipschitz distance apart to be finite. Thus, on the one hand, the Hausdorff distance is not in general finite for unbounded spaces and is therefore too restrictive "at infinity," as we have observed in Proposition 3.13, whereas the Lipschitz distance can be rendered infinite by a small singularity at finite distances. This is why we need a distance that combines the two.

3.19. Definition: The *Hausdorff–Lipschitz distance* between two metric spaces X, Y, denoted $d_{\mathrm{HL}}(X,Y)$, is the infimum of the numbers

$$d_{\mathrm{H}}(X, X_1) + d_{\mathrm{L}}(X_1, Y_1) + d_{\mathrm{H}}(Y_1, Y),$$

where X_1, Y_1 are arbitrary metric spaces. One says that X and Y are *quasi-isometric* if $d_{\mathrm{HL}}(X,Y) < \infty$.

Before presenting some properties of this metric, we first describe two means for defining a distance on a finitely-generated discrete group.

3.20. Definition: Let Γ be a discrete group of finite type and $\{\gamma_i\}$ a finite set of generators for Γ. For each $\alpha \in \Gamma$, we define $\|\alpha\|_{\mathrm{alg}}$ as the smallest length of a word in the γ_i and their inverses that represents α. The function $\|\cdot\|_{\mathrm{alg}} : \Gamma \to \mathbb{R}_+$ is a norm on Γ, i.e., a function satisfying

1. $\|\alpha\|_{\mathrm{alg}} = 0$ if and only if $\alpha = 1$.

2. $\|\alpha + \beta\|_{\mathrm{alg}} \leq \|\alpha\|_{\mathrm{alg}} + \|\beta\|_{\mathrm{alg}}$.

3. $\|\alpha^{-1}\|_{\mathrm{alg}} = \|\alpha\|_{\mathrm{alg}}$.

We call $\|\cdot\|_{\mathrm{alg}}$ the *algebraic (word) norm associated with the set* $\{\gamma_i\}$ *of generators.* It is the largest norm that assumes the value 1 on each γ_i.

$3.20\frac{1}{2}{}_+$ Changing generators. If we replace γ_i by another finite system of generators γ'_j, then the resulting norm remains *equivalent* to the old one, i.e.,

$$c\|\cdot\|^{\mathrm{old}} \leq \|\cdot\|^{\mathrm{new}} \leq c'\|\cdot\|^{\mathrm{old}}.$$

In fact, this is (obviously) true when $c = (\sup_i \|\gamma_i\|^{\mathrm{new}})^{-1}$ and when $c' = \sup_j \|\gamma'_j\|^{\mathrm{old}}$, and a more general geometric equivalence is stated in Proposition 3.22 below.

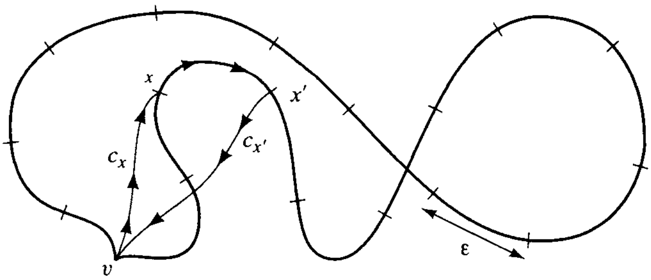
3.21. Definition: Let V be a compact Riemannian manifold and let $\tilde{V}$ be its universal cover, so that the fundamental group Γ of V acts by isometries on $\tilde{V}$. The *geometric norm* associated with a point $v \in V$ is defined as

$$\|\alpha\|_{\text{geo}} = d(\tilde{v}, \alpha\tilde{v})_{\tilde{V}},$$

where $\tilde{v}$ is any point lying over v. This norm equals the length of a shortest loop representing α.

3.22. Proposition: *For each $v \in V$, the fundamental group $\pi(V, v)$ is generated by a finite number of classes represented by loops based at v having length at most $2\,\text{diam}(V)$. All of the norms (a) associated with finite sets of generators and (b) associated with points of V are equivalent.*

Proof. For each element α of the fundamental group Γ of V, we choose a representative loop a based at $v \in V$. For $\varepsilon > 0$ we divide each loop a into pieces of length less than ε and connect each of the endpoints x thus obtained to v by a minimizing arc c_x. Then a is homotopic to the product of the loops $c_x a|_{[x,x']} c_{x'}^{-1}$ based at v, where (x, x') denotes a consecutive pair of endpoints along a.



Thus, we have found a system of generators for Γ represented by loops of length less than $2\,\text{diam}(V) + \varepsilon$. The set of lengths of minimizing geodesics is discrete (cf. 1.13), however, and so for sufficiently small ε, the interval $(2\,\text{diam}(V), 2\,\text{diam}(V) + \varepsilon)$ contains no length of a closed, minimizing geodesic based at v. Consequently, each of the curves we have constructed is homotopic to a curve of length at most $2\,\text{diam}(V)$.

We begin by showing that the algebraic norms associated with two finite sets G^1, G^2 of generators are equivalent. Let $\|\cdot\|^1, \|\cdot\|^2$ denote these norms, and set $S^1 = \sup\{\|\alpha\|^2 : \alpha \in G^1\}$. Then $(1/S^1)\|\cdot\|^2$ is a norm on Γ that assumes values less than 1 on G^1. Thus,

$$(1/S^1)\|\cdot\|^2 \leq \|\cdot\|^1.$$

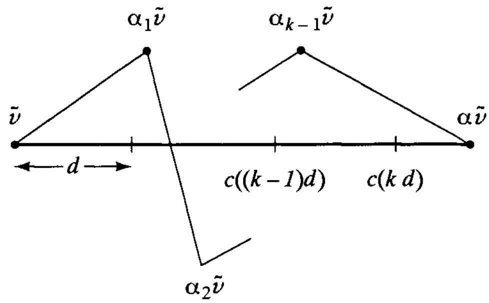
Given a point $\tilde{v} \in \tilde{V}$, let G be a finite set of generators of Γ containing all elements of Γ for which $\|\alpha\|_{\text{geo}} \leq 3\,\text{diam}(V) = 3d$ for the norm associated with the point v. Again, the norm $(1/3d)\|\cdot\|_{\text{geo}}$ assumes values less than 1 on the generators belonging to G and is therefore bounded above by the algebraic norm $\|\cdot\|_{\text{alg}}$ associated with G. Conversely, suppose we are given an element α of Γ and a minimizing geodesic c from $\tilde{v}$ to $\alpha\tilde{v}$ in $\tilde{V}$ parametrized by arc length. Then the points $\tilde{v} = c(0), c(d), c(2d), \ldots, c(kd)$ (where $k = [d^{-1}\|\alpha\|_{\text{geo}}]$) are spaced by d. For each $i \leq k$, there exists an element α_i of Γ such that $d(c(id), \alpha_i\tilde{v}) \leq d$. Thus, $d(\alpha_i\tilde{v}, \alpha_{i+1}\tilde{v}) \leq 3d$, and so

$$\alpha_i^{-1}\alpha_{i+1} \in G$$

and

$$\|\alpha\|_{\text{alg}} \leq \|\alpha_1\|_{\text{alg}} + \sum_{i=1}^{k-2} \|\alpha_i^{-1}\alpha_{i+1}\|_{\text{alg}} + \|\alpha_{k-1}^{-1}\alpha\|_{\text{alg}} \leq k \leq \frac{1}{d}\|\alpha\|_{\text{geo}}.$$

We conclude that $d\|\cdot\|_{\text{alg}} \leq \|\cdot\|_{\text{geo}} \leq 3d\|\cdot\|_{\text{alg}}$.



Every norm on a group induces a metric by the formula

$$d(\alpha, \beta) = \|\alpha^{-1}\beta\| = \|\beta^{-1}\alpha\|.$$

The above-mentioned property can be reformulated thus: Each of the metric spaces $(\Gamma, \|\cdot\|_{\text{alg}})$ or $(\Gamma, \|\cdot\|_{\text{geo}})$ lie within a finite Lipschitz distance from one another.

3.23. Proposition: *If two manifolds V_1, V_2 have isomorphic fundamental groups, then $d_{\text{HL}}(\tilde{V}_1, \tilde{V}_2) < \infty$. If two groups Γ_1, Γ_2 act on the same manifold V and are cocompact in V (i.e., the quotients V/Γ_i are compact), then $d_{\text{HL}}(\Gamma_1, \Gamma_2) < \infty$.*

Proof. This follows from the fact that, in both cases, $d_{\text{H}}(\tilde{V}, \Gamma\tilde{v}) < \infty$.

Example: The fundamental groups of compact hyperbolic n-manifolds are quasi-isometric.

Example: If $\tilde{V}$ is the universal cover of a compact manifold V whose fundamental group is isomorphic to the discrete Heisenberg group,

$$\left\{\begin{pmatrix} 1 & m & n \\ 0 & 1 & p \\ 0 & 0 & 1 \end{pmatrix} : m, n, p \in \mathbb{Z}\right\},$$

then $d_{\mathrm{HL}}(\tilde{V}, \mathbb{H}^3) < \infty$, where $\mathbb{H}^3$ is again the continuous 3-dimensional Heisenberg group.

3.24$_+$ The word metric. The metric on Γ associated to an algebraic norm is called a *word metric.* Word metrics associated to different finite generating subsets of Γ are bi-Lipschitz equivalent, i.e., the identity map $(\Gamma, \mathrm{dist}_1) \leftrightarrow (\Gamma, \mathrm{dist}_2)$ is bi-Lipschitz, as follows from the equivalence of the norms. This suggests the geometric study of finitely generated groups, where we are interested in the those properties which are invariant under bi-Lipschitz maps. In fact, it is more convenient to relax "bi-Lipschitz convergence" to "quasi-isometry," since $\Gamma = \pi_1(V)$ is *quasi-isometric to* $\tilde{V}$, as we have just seen, rather than bi-Lipschitz equivalent to it. On the other hand, since the word metrics are discrete, one can think that the two notions coincide for groups, i.e.,

$$d_{\mathrm{HL}}(\Gamma_1, \Gamma_2) < \infty \;\Rightarrow\; d_{\mathrm{L}}(\Gamma_1, \Gamma_2) < \infty,$$

but this question remains unsettled at the present date. Observe that the existence of a quasi-isometry between metric spaces X_1 and X_2 is equivalent to the existence of bi-Lipschitz equivalent *nets* $\Delta_i \subset X_i$, where $\Gamma \subset X$ is called a net if $\mathrm{dist}(x, \Delta) \leq \mathrm{const}$ for all $x \in X$. Furthermore, one can insist that these nets are *separated* (where $\Delta \subset X$ is said to be separated or *uniformly discrete* if $\mathrm{dist}(\delta_1, \delta_2) \geq \varepsilon > 0$ for all δ_1 and δ_2 in Δ), since every net lies within a finite Hausdorff distance from a separated net. Now the algebraic problem above can be stated in the following geometric terms:

For which metric spaces X are every two separated nets in X bi-Lipschitz equivalent, i.e., have $d_{\mathrm{L}} < \infty$?

It is obvious that all separated nets in $\mathbb{R}$ are bi-Lipschitz, but this is unknown in $\mathbb{R}^n$ for $n \geq 2$. On the other hand, the question is resolved for free groups by Papasogly. In particular, *the free groups on p and q generators are mutually bi-Lipschitz equivalent for all $q \geq p \geq 2$* (see [Pap]$_{\mathrm{HTBE}}$).

The key property of the free groups Γ_p, $p \geq 2$, is *nonamenability* (see Ch. 6). In fact, it is not very hard to show that *for every net Δ in a* *nonamenable group* Γ, *there exists a bijective map* $f : \Gamma \to \Delta$ *such that* $\operatorname{dist}(\gamma, f(\gamma)) \leq \text{const}$[1].

Exercise: Show that the above fails to be true in $\Gamma = \mathbb{Z}^n$, as well as in any finitely generated amenable group.

The notion of quasi-isometry, originated in the work by Mostow and then Margulis (see [Most]), generalizes the purely algebraic notion of *commensurability* of discrete groups, which means the existence of isomorphic subgroups $\Delta_1 \subset \Gamma_1$ and $\Delta_2 \subset \Gamma_2$ of finite index. It is believed that for many classes of groups, quasi-isometry implies commensurability, but one must be aware of the following:

3.25$_+$ Counterexamples: (*a*) Let Γ_1 and Γ_2 be cocompact discrete subgroups in a given connected Lie group G (or, more generally, in a compactly generated group). Then, clearly, Γ_1 and Γ_2 are quasi-isometric, but one knows that they do not have to be commensurable, say for $G = SL_2(\mathbb{C})$.

(*b*) Here is a more dramatic example. Let Γ_p denote the fundamental group of an oriented surface of genus p and let Γ'_p be the fundamental group of the unit tangent bundle of this surface. Then, *for* $p \geq 2$, Γ'_p *is quasi-isometric (and hence bi-Lipschitz) to* $\Gamma_p \times \mathbb{Z}$, as was (independently) observed by S. Gersten and D. Epstein. (Note that this is false for $p = 1$, since Γ'_p is a nilpotent nonvirtually abelian group.)

Proof. For $p \geq 2$, the group Γ_p is quasi-isometric to the hyperbolic plane H, while Γ'_p is quasi-isometric to the universal cover of the unit tangent bundle UH. The cover $\widetilde{UH}$ constitutes a (topologically trivial) $\mathbb{R}$-fibration over H with connection given by Levi-Civita on H. Now fix a point $v_0 \in H$ and bring each fiber $\widetilde{UH}_v \subset \widetilde{UH}$, $v \in H$, to v_0 via parallel transport along the geodesic segment between v and v_0. This gives us a projection $\varphi \colon \widetilde{UH} \to \mathbb{R} = \widetilde{UH}_{v_0}$ which is *Lipschitz*, since the holonomy around every geodesic triangle $\Delta(v_0, v, v')$ in H is bounded by its area and thus by *const.*$\cdot \operatorname{dist}(v, v')$, since geodesics issuing from v_0 exponentially diverge. It follows that the map $f : \widetilde{UH} \to H \times \mathbb{R}$ satisfying $f(\tilde{u}) = (p(\tilde{u}, \varphi(\tilde{u}))$, where p is the projection $\widetilde{UH} \to H$, is Lipschitz, and since it is isometric on the fibers $\widetilde{UH}_v = \mathbb{R} \to v \times \mathbb{R}$, it is bi-Lipschitz as well.

3.26$_+$ Exercises: (*a*) Generalize the above to the universal covering of a circle bundle with connection over a complete, simply connected manifold

[1] "Exotic" nets in $\mathbb{R}^n$ are constructed in the recent papers [McMl] and [Bur–Klei], and nonamenable spaces are treated in [Why], [Bog], and [Necr].

with negative curvature $\leq -\kappa < 0$, provided that the connection in question has bounded curvature form (e.g., coming from a compact manifold V via a covering $X \to V$).

(*b*) Let X be a *Hermitian* symmetric space with nonpositive curvature, and let $U \to X$ be the circle bundle associated to the *canonical bundle* of X, i.e., $\bigwedge^n T(X)$, where $n = \dim_{\mathbb{C}} X$ and $\bigwedge^n$ denotes the (top) exterior power. Show that $\tilde{U}$ is bi-Lipschitz equivalent to $X \times \mathbb{R}$ and apply this to central extensions of cocompact discrete groups acting on X. *Hint:* Observe that the curvature form ω of U vanishes on flat geodesic subspaces in X.

(*c*) **A dynamical criterion for quasi-isometry:** *Γ_1 and Γ_2 are quasi-isometric if and only if there exists a metric space X admitting mutually commuting cocompact discrete actions of Γ_1 and Γ_2.*

Hint: The "if" part is straightforward. To prove "only if," construct X as the space of λ-Lipschitz maps $f: \Gamma_1 \to \Gamma_2$, where λ is a fixed large number, and where f admits *quasi-inverse* λ-Lipschitz maps $g: \Gamma_2 \to \Gamma_1$, i.e., such that $f \circ g$ and $g \circ f$ lie within a finite distance from the identity maps $\Gamma_2 \to \Gamma_2$ and $\Gamma_1 \to \Gamma_1$, respectively (where $\operatorname{dist}(\varphi, \psi) \stackrel{\text{def}}{=} \sup_\gamma \operatorname{dist}(\varphi(\gamma), \psi(\gamma))$).

Example: If $\Gamma_1 = \Gamma_2 = \Gamma$, one can use Γ for X with the left and right action of Γ on itself.

References: Geometric group theory has recently exploded into a vast field, see [Groups].

D_+ First-order metric invariants and ultralimits

3.27. Hausdorff convergence via $K_r(X)$. We recall that the subsets $K_r(X)$ in the space M_r of symmetric $(r \times r)$-matrices are comprised of the distances between all r-tuples of points in X (compare 1.19_+). We begin with the following elementary observation.

$3.27\frac{1}{2}$. *Two compact metric spaces X, Y are isometric if and only if $K_r(X) = K_r(Y)$ for all $r = 1, 2, \ldots$.*

Proof. The inclusion $K_r(X) \subset K_r(Y)$ says that every r-point subset in X embeds isometrically into Y. Since X and Y are compact, this property for all $r = 1, 2, \ldots$ implies that X itself embeds isometrically into Y, and similarly, Y embeds isometrically into X. Thus, X is isometric to Y, since both are compact.

Note that compactness is an essential condition, as the example of $X = \mathbb{R}$ and $Y = \mathbb{R}_+$ shows.

Now we can introduce *local* Hausdorff convergence $X_i \to Y$ as the convergence of the subset $K_r(X_i) \to K_r(Y)$ for every given $r = 1, 2, \ldots$, where the space of subsets in $M_r = \mathbb{R}^{r(r-1)/2}$ is given the ordinary Hausdorff metric. Clearly, this new local convergence $X_i \xrightarrow{H_{\mathrm{loc}}} Y$ implies our old one if the family $\{X_i\}$ is precompact in the topology of d_{H}. On the other hand, any sequence X_i with $\operatorname{diam}(X_i) \leq \text{const}$ admits a subsequence, say X_j, such that $K_r(X_j)$ Hausdorff-converges to some subset $N_r \subset M_r$ for each $r = 1, 2, \ldots$ (and the same applies without the bound on the diameter if one speaks of Hausdorff convergence in M_r on compact subsets). It is not hard to show (an exercise for the reader) that there exists a (nonunique) metric space Y with $K_r(Y) = N_r$ for all $r = 1, 2, \ldots$. In fact, there is a canonical choice of such Y (explained in 3.29) which absorbs full information on the limit behavior of the geometries of finite subsets in X_i.

3.28. First order theory of metric spaces. Finite configurations of points in X carry more information than is encoded in $K_r(X) \subset M_r$. For example,

$$\operatorname{Rad}(X) \overset{\text{def}}{=} \inf_x \sup_y \operatorname{dist}(x, y)$$

cannot be expressed in terms of $K_r(X)$. Then one can get even trickier and define something more sophisticated (and ugly) like this:

$$X' = \{x \in X : \sup_y \operatorname{dist}(x, y) \leq 1.1 \operatorname{Rad}(X)\},$$

and $X'' = (X')'$, $X''' = (X'')'$, $\ldots$, etc. Then set $\operatorname{Rad}^{\cdots}(X) = \operatorname{Rad}(X^{\cdots})$. The invariants of this type can be expressed with (long) chains of quantifiers $\exists x \forall y \exists z \cdots$, or set-theoretically as follows. We think of K_r as an M_r-valued map in r variables $x_1, \ldots, x_r$. If we temporarily fix $x_1, \ldots, x_{r-1}$ and vary $x = x_r \in X$, we get a map $X \to M_r$ whose image in M_r is denoted by $\operatorname{im}_{x_r} K(x_1, \ldots, x_r) \in M_r'$, where M_r' denotes the power set of M_r, i.e., the set of subsets in M_r. Thus, we get an M_r'-valued function in the variables $x_1, \ldots, x_{r-1}$, say $K_r'(x_1, \ldots, x_{r-1})$, and then by repeating the procedure above, we get the $M_r'' = (M_r')'$-valued function $K_r''(x_1, \ldots, x_{r-1})$, then $K_r'''(x_1, \ldots, x_{r-1})$, etc., terminating with a single point in $M_r^{(r)}$ denoted by $K_r^{(r)} \in M_r^{(r)}$, which conveys the full information on the geometry of r-tuples of points in X. Then each space $M_r^{(i)}$ can be given, by induction on i, the metric corresponding to the Hausdorff distance between subsets in $M_r^{(i-1)}$. This gives a (semi-)metric on metric spaces leading to a refinement of the above notion of local Hausdorff convergence.

Examples: (*a*) The spaces $X = \mathbb{R}$ and $Y = \mathbb{R}_+$ are distinguishable by 3-point configurations, since every point in $\mathbb{R}$ can appear as the center of a triple of points isometric to $\{-1, 0, 1\}$, while this is not true in $\mathbb{R}_+$.

(*b*) The coverability of X by, say, 100 balls of given radii can be expressed in terms of the above M'_{101}.

(*c*) The metric connectedness of X, i.e., the existence of chains $x_1 = x, x_2, \ldots, x_r = y$ between each $x, y \in X$ with $\mathrm{dist}(x_i, x_{i+1}) < \varepsilon$ for every $\varepsilon > 0$ can be expressed in terms of (the totality of) M'_r, $r = 1, 2, \ldots$. But, the ordinary topological connectedness seems to be out of the reach of $M_r^{(i)}$.

(*d*) It is not hard to express properties of coverings of X by finite unions of balls in terms of K's, but the full-fledged notion of the topological dimension is beyond $K_r^{(i)}$; the same can be said about the Hausdorff dimension, where the usual definition is even more transcendental.

(*e*) Start with the real line and attach to it unit segments at the point of a given subset $V \subset \mathbb{R}$. For example, take $V \subset \mathbb{Z} \subset \mathbb{R}$ by choosing the points in $\mathbb{Z}$ independently at random with probability 1/2. Given two such random subsets V, V' in $\mathbb{Z} \subset \mathbb{R}$, we have two different spaces, say X_V and $X_{V'}$ which are obviously nonisometric (use $K_r^{(2)}$) but which cannot be distinguished by sequences of finite measurements. (For example, for each $x \in X_V$ and every $R > 0$, there exists a point $x' \in X_{V'}$ such that x and x' have isometric R-balls around them. But, if we replace $\mathbb{R}$ by $\mathbb{R}_+$, it is easy to tell $X_V(\mathbb{R})$ from $X_{V'}(\mathbb{R}_+)$ by looking at balls). A hierarchy of different systems of invariants is possible in the framework of model theory, but I do not know if anybody has carried out such investigations. On the other hand, one is keen on particular invariants with a geometric flavor to them (see $[\mathrm{Grov}]_{\mathrm{MTM}}$). For example, one can take a function I on M_5 and define

$$\mathrm{Inv}_I(X) = \inf_{x_1} \sup_{x_2} \inf_{x_3} \inf_{x_4} \sup_{x_5} I(K_5(x_1, x_2, x_3, x_4, x_5)).$$

The question is for which I (if any) one gets something interesting (compare $[\mathrm{Grov}]_{\mathrm{CPT}}$, $[\mathrm{Grov}]_{MTM}$).

3.29. Ultraproducts. Suppose we have distinguished a class of $\mathbb{R}$-valued metric invariants of metric spaces. For example, we could use the above In_I, where we allow all bounded, continuous functions I on M_r, $r = 1, 2, \ldots$, and we take arbitrary inf sup sequences. Then we say that a sequence X_i converges to X if $\mathrm{Inv}(X_i) \to \mathrm{Inv}(X)$ for each invariant Inv from our class. If we want to get away with this, we must face the following problem. Suppose $\lim_{i\to\infty} \mathrm{Inv}(X_i)$ exists for each of our Inv. Does there then exist a space X with these limit invariants and which can therefore be called $\lim_{i\to\infty} X_i$?

Here is a universal construction. Start with the cartesian product $Y = \{x_1, x_2, \ldots, x_i, \ldots\}$ with $x_i \in X_i$, where the distance takes values in $\mathbb{R}_+^\infty$ rather than in $\mathbb{R}_+$, namely $\mathrm{DIST}(\{x_i\}, \{x_i'\}) = \{\mathrm{dist}_{X_i}(x_i, x_i')\}$, $i = 1, 2, \ldots$. Now we want to pass from sequences $d_1, d_2, \ldots,$ to numbers, i.e., to assign some kind of limit to $\{d_i\}$ for $i \to \infty$. In other words, we need a suitable projection $\mathbb{R}_+^\infty \to \mathbb{R}_+$, and such is known to exist under the name of *nonprincipal ultrafilters* on the natural numbers. Given such an ultrafilter, say ω, we can speak of $\lim_\omega d_i$ for an arbitrary bounded sequence d_i, and we define dist_ω on Y as $\lim_\omega \mathrm{dist}_{X_i}(x_i, x_i')$. Notice that this distance can vanish on some distinct pairs $(\{x_i\}, \{x_i'\})$ (for example, it is zero if $x_i = x_i'$ for all but finitely many i's) and so we must factor Y by the $\mathrm{dist}_\omega = 0$ equivalence relation. The resulting space is our ultraproduct, X_ω. (Recall that ω can be thought of as a point in the Stone-Čech compactification of $\mathbb{N}$, and X_ω appears as the limit of the X_i for $i \to \omega$.)

This definition must be adapted slightly for pointed spaces X_i, and it becomes particularly useful for defining the *asymptotic cone* $\mathrm{Con}_\infty(X) = \lim_{\varepsilon\to 0} \varepsilon X$ as follows (compare [Dri–Wil], $[\mathrm{Gro}]_{\mathrm{AI}}$). Fix a point $x_0 \in X$, consider the sequences x_i satisfying $\mathrm{dist}(x_i, x_0) \leq \mathrm{const}\, i$, where the constant can depend on $\{x_i\}$ (but not on i, of course!) and set

$$\mathrm{dist}_\omega(\{x_i\}, \{x_i'\}) = \lim_\omega \frac{1}{i} \mathrm{dist}(x_i, x_i').$$

3.29½ Asymptotic cones of hyperbolic spaces. Let X be a complete, simply connected space with negative sectional curvature $K(X) \leq -\kappa < 0$. For instance, one can take the hyperbolic plane H^2, or the n-dimensional H^n (including $n = \infty$). Then this X is *tree-like at infinity* (see $[\mathrm{Gro}]_{\mathrm{HG}}$), which can be adequately expressed by saying that $\mathrm{Con}_\omega(X)$ is a *tree*, i.e., a 1-dimensional contractible path metric space, where, moreover, there is a *unique* topological segment containing two given distinct points as its endpoints. Notice that this tree has uncountable branching at every point. (In fact, and this is especially clear for $X = H^n$, the isometry group is transitive on $\mathrm{Con}_\omega(X)$.) Yet, this is not as bad as it may seem, and this limit tree turns out to be quite useful for the study of *hyperbolic groups* Γ, which can actually be defined by requiring $\mathrm{Con}_\omega(\Gamma)$ to be a tree for the word metric in Γ (see [Paul]). Also, there are nonhyperbolic spaces (especially groups) with reasonable (e.g., finite-dimensional) Con_ω. This is probably the case for all connected Lie groups, p-adic Lie groups and lattices (i.e., finite co-volume subgroups) in them (compare $[\mathrm{Gro}]_{\mathrm{AI}}$).

E_+ Convergence with control

3.30. Examples of regular and irregular Hausdorff convergence. The ideal situation is where the Hausdorff convergence $X_i \xrightarrow{\mathrm{Hau}} X$ reduces to Lipschitz convergence $X_i \xrightarrow{\mathrm{Lip}} X$ such that the implied bi-Lipschitz homeomorphisms $f_i\colon X \to X_i$ satisfy $f_i(x) \xrightarrow{\mathrm{Hau}} x$ for all $x \in X$. (Notice that the Hausdorff convergence $X_i \xrightarrow{\mathrm{Hau}} X$ allows us to define convergence $x_i \xrightarrow{\mathrm{Hau}} x \in X$ for sequences $x_i \in X_i$.) This, of course, is an exceptionally rare case, and we would be satisfied by just having continuous maps (preferably homeomorphisms) $f_i : X_i \to X$ satisfying $f(x_i) \xrightarrow{\mathrm{Hau}} x$ for all $x \in X$.

Another class of examples is suggested by the degeneration of algebraic and semialgebraic subsets in $\mathbb{R}^n$. Namely, if a sequence of such subsets X_i Hausdorff-converges to X, and the degrees of X_i are bounded by a constant independent of i, then X is also semialgebraic of $\dim(X) \leq \sup_i \dim(X_i)$ and $\deg(X) \leq \sup_i \deg(X_i)$. Furthermore, the degeneration occurring in the course of the convergence is no worse than in the following cases:

(a) the circle $x^2 + y^2 = \varepsilon^2$ converges to a single point as $\varepsilon \to 0$;

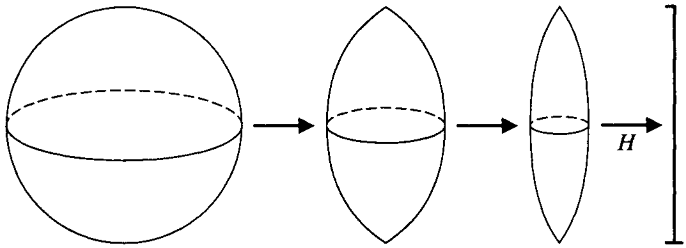
(b) the hyperbola $xy = \varepsilon$ converges to the pair of lines $xy = 0$;

(c) let X be an arbitrary semialgebraic, e.g., piecewise linear subset in $\mathbb{R}^n$ of positive codimension and let X be the boundary of the ε-neighborhood of X_ε. Then $X_\varepsilon \to X$ as $\varepsilon \to 0$, where X can consist of several pieces of different dimensions, and where one may use the induced path metric in X_ε for $\mathrm{codim}(X) \geq 2$.

Next, one can look at convergent Riemannian manifolds, $X_i \xrightarrow{\mathrm{Hau}} X$, where $\dim X_i = n$. A basic example is where we are given an isometric action of a compact, connected Lie group G on X_0, and we collapse X_0 to $X = X_0/G$ by shrinking the original Riemannian metric g_0 on X along the orbits. Specifically, if the orbits have equal dimension and X_0 fibers over X, then we split g_0 as $g_0 = g_0^{\mathrm{vert}} + g_0^{\mathrm{hor}}$ according to the splitting of the tangent bundle $T(X_0)$ into the parts which are normal and tangent to the orbits. Then, we take

$$X_i = (X_0, g_i = i^{-1} g_0^{\mathrm{vert}} + g_0^{\mathrm{hor}})$$

and observe that g_0^{hor} is induced by the projection $X_0 \to X = X_0/G$; thus, $X_0 \xrightarrow{\mathrm{Hau}} X$ (compare 3.9). In general, if the orbits are not equidimensional, we apply the shrinking to the maximal open subset $U_0 \subset X_0$ where the orbits have the maximal dimension. The resulting Riemannian metric on U_0 gives us a (singular) path metric on all of X_0 as $\mathrm{codim}(X_0 \setminus U_0) \geq 2$ (since G is connected) and U_0 is connected. Thus, we get the convergence $X_i \xrightarrow{\mathrm{Hau}} X$ anyway. For example, S^2 with the obvious S^1-action collapses to the segment $[0, \pi]$.

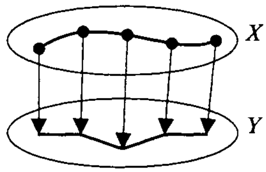


Now, we do not introduce any group action but just assume that $X_i \xrightarrow{\mathrm{Hau}} X$, while the sectional curvatures of X_i are uniformly bounded, $|K(X_i)| \leq$ const. Then again, the picture is very similar to the above group actions. Namely, either there is *no collapse*, i.e., $\dim(X) = n = \dim(X_i)$, and then $X_i \xrightarrow{\mathrm{Lip}} X$ or, if $\dim(X) < n$, then the convergent manifolds X_i can be visualized as fibrations over X with possibly singular fibers. (See Ch. 8_+). In fact, much of this stratified fibration picture persists under the one-sided curvature bound $K(X_i) \geq -\mathrm{const}$ and something even remains for $\mathrm{Ricci}(X_i) \geq -\mathrm{const}$ especially if X_i are *Einstein* manifolds. (See [Che–Col], [Col]$_{\mathrm{ARC}}$).

Finally we turn to "bad" examples, among which the foremost is given by discrete ε-nets X_ε^o in X which Hausdorff-converge to X. One can turn such a net into a path metric space, namely a graph X_ε^1 with an ε-net $X_\varepsilon^o \subset X$ serving as the vertex set of X_ε^1, and where every two points x and y in X_ε within distance $\delta \leq \delta_\varepsilon$ are joined by an edge of length δ. We chose $\delta_\varepsilon \to 0$ for $\varepsilon \to 0$ with $\varepsilon/\delta_\varepsilon \to \infty$ and then, clearly, $X_\varepsilon^1 \xrightarrow{\mathrm{Hau}} X$ provided that X is a path metric space. Here, we achieved connectedness of the spaces X_ε approximating X but paid by high nonsimple connectedness. Furthermore, if one prefers manifolds, one can replace X_ε^1 by surfaces: just embed each X_ε^o into $\mathbb{R}^3$ and take the boundary of a small ε'-neighborhood for an approximating surface (compare [Cass]).

One may perform further surgery, say fill in all small triangles of edges in the graphs X_ε^1 by actual 2-dimensional triangles, thus obtaining a sequence of 2-polyhedra $X^2_\varepsilon \to X$. If X is locally simply-connected, then the fundamental groups of X^2_ε for small ε are naturally isomorphic to $\pi_1(X)$ (compare below).

3.31. Remark on surjectivity $\pi_1(X_\varepsilon) \to \pi_1(X)$. If two path metric spaces, say X and Y, are ε-Hausdorff close, then every path in X can be (nonuniquely) moved to Y by subdividing it into segments of length $\leq \varepsilon$, moving the division points to nearest points in Y and joining them in Y by minimal geodesic segments (of length $\leq 3\varepsilon$).



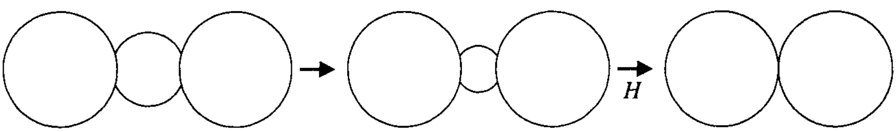
(Recall that X and Y may be isometrically embedded into some metric space Z, where their ordinary Hausdorff distance becomes $\leq \varepsilon$ and "moving points by ε" refers to the metric on $X \cup Y$ induced from Z). Now let $X_\varepsilon \xrightarrow{\text{Hau}} Y$ for $\varepsilon \to 0$, take a closed loop α in Y, move it to X_ε and then go back from X_ε to Y. The resulting loop α' in Y is about ε-close to α and so it is homotopic to α if Y is locally simply connected. Thus, the convergence $X_\varepsilon \xrightarrow{\text{Hau}} X$ ensures an *epimorphism* $\pi_1(X_\varepsilon) \to \pi_1(X)$ for $\varepsilon \to 0$.

This argument applies both ways to the above $X^2_\varepsilon \to X$, since the spaces X^2_ε are uniformly locally simply connected in an obvious sense, and so $\pi_1(X^2_\varepsilon) \simeq \pi_1(X)$ for small ε.

3.32. Hausdorff moduli spaces of graphs and surfaces. We have just seen that every compact path metric space appears as a limit of graphs (i.e., 1-complexes), but the topological type of these X^1_ε goes to infinity for $\varepsilon \to 0$. Now let us look at the finite 1-complexes X of bounded topological type, where *topological complexity* $\text{Top}(X)$ is defined as the minimal number of edges needed to present (or subdivide) X.

Here is an obvious but pleasant fact.

Observation: *Let $X_\varepsilon \xrightarrow{\text{Hau}} X$ for $\varepsilon \to 0$, where X_ε are graphs satisfying* $\text{Top}(X_\varepsilon) \leq t < \infty$. *Then X is also a graph with* $\text{Top}(X) \leq t$.

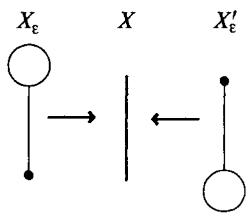


(Notice that the above remains valid when X_ε and/or X are allowed to be noncompact.)

Corollary: *The space $\mathcal{X} = \mathcal{X}(1,t)$ of finite connected graphs X with* $\mathrm{Top}(X) \leq t$ *is complete with respect to the Hausdorff topology. Furthermore, the subset $\mathcal{X}_\ell \subset \mathcal{X}$ consisting of X with length ℓ is compact.*

Notice that $\mathcal{X}$ is naturally a cone, where $\mathbb{R}_+$ acts by $X \mapsto \lambda X = (X, \lambda \cdot \mathrm{dist}_X)$ for $\lambda \in \mathbb{R}_+$, and each $\mathbb{R}_+$ orbit intersects $\mathcal{X}_\ell$ for $\ell > 0$ at a single point. Also observe that $\mathcal{X}$ admits a natural finite cell decomposition: each cell consists of the set of mutually homeomorphic graphs X.

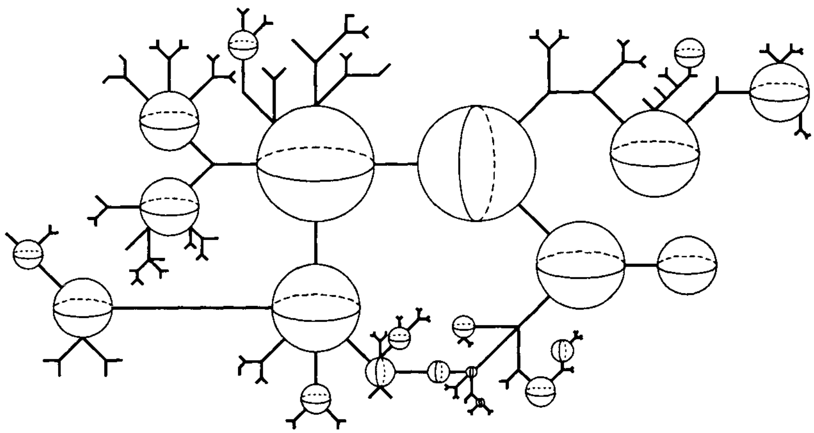
To better understand $\mathcal{X}$, one should observe that it carries a natural *orbispace* structure, i.e., for each point $X \in \mathcal{X}$, there exists a neighborhood $\mathcal{U} \subset \mathcal{X}$ of X which is represented as $\tilde{U}/G$ for some metric space $\tilde{U}$ with an isometric action of a finite group G on $\tilde{U}$. The group G here is the isometry group of X (where one should be aware of $X = S^1$ with infinite G which needs the extra case) and $\tilde{U}$ is obtained by distinguishing the isometric deformations X_ε and X'_ε of X if the implied isometry between X_ε and X'_ε correspond to a non-identity in G.



The (suitably defined for orbispaces) *universal covering* of the (slightly modified) $\mathcal{X}$ plays the key role in the study of the automorphism groups of free groups. (See [Bri–Vog])

Polyhedra of dimension $n \geq 2$. The topological complexity of spaces X of dimension $n > 1$ may increase under Haudorff limits (see below), but one can bound the *geometric complexity*, $\mathrm{geo}(X)$, defined as the minimal number of Euclidean simplices needed to partition X. (For $\dim(X) = 1$, obviously, $\mathrm{Top}(X) = \mathrm{geo}(X)$). Now again, whenever $X_\varepsilon \xrightarrow{\varepsilon} X$ with $\mathrm{geo}(X) \leq g$, we have $\dim(X) \leq n \leq \dim(X_\varepsilon)$ and $\mathrm{geo}(X) \leq g$, but this has little topological significance. (One could measure the complexity in terms of the combinatorial structure of the stratification of our spaces into locally Euclidean strata, keeping an eye on the Bieberbach theorem, which implies that every compact, flat n-dimensional manifold X can be triangulated into at most $g \leq g(n)$ Euclidean simplices). In any case, the corresponding moduli space $\mathcal{X}(n,g)$ is a conical space of dimension $\leq g_1$, where g_1 is the minimal number of 1-simplices that a Euclidean triangulation of X with $\mathrm{geo}(X) \leq g$ can have. In fact, X lies in $\mathbb{R}^{g_1}$ as a semialgebraic set. Some fragments of $\mathcal{X}(2)$ appear in Teichmüller theory, since Riemann surfaces admit piecewise Euclidean metrics with few singular points. Notice that the Teichmüller space itself can be defined as the universal orbispace covering of the Hausdorff space of surfaces of constant curvature -1 and of a fixed genus. Also, topologists have studied the parts of $\mathcal{X}(n)$ corresponding to polyhedra of a given simple homotopy type, but it is unclear whether there is a meaningful geometric theory of $\mathcal{X}(n,g)$ for all n.

Surfaces of genus g. Consider all closed surfaces X of genus g equipped with path (e.g., Riemannian) metrics, and look at the Hausdorff completion of the space of these X's. If such surfaces Hausdorff-converge to some metric space $X_\varepsilon \xrightarrow{\mathrm{Hau}} Y$, then the assumption $\mathrm{genus}(X_\varepsilon) = g$ severely restricts the topology of Y. For example, if $g = 1$, then the limit may look no worse than the following *bubble (or branching) space*,



In fact, we have the following

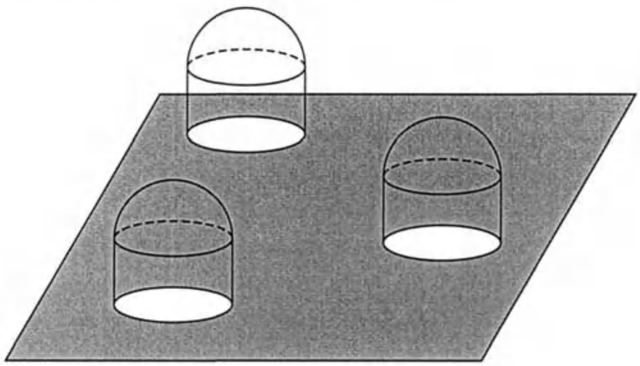
Proposition (Compare [Shio]): *Every point $y \in Y$ either admits a neighborhood homeomorphic to $\mathbb{R}^2$ or y is a local cut point, i.e., $U \setminus \{u\}$ is disconnected for every small neighborhood $U \subset Y$ of y.*

Sketch of the proof: Suppose that there is a sequence of closed curves $C_\varepsilon \subset X_i$ such that $\mathrm{diam}(C_\varepsilon) = \delta_\varepsilon \to 0$ as $\varepsilon \to 0$, but C_ε is nonhomologous to zero in its ρ-neighborhood in X_ε for a fixed $\rho > 0$. Then, clearly, the C_ε (sub)converge to a local cut point in Y. So, if Y has no local cut points, then the surfaces X_ε are *controllably locally acyclic*, i.e., there exists a (control) function $\rho(\delta)$, $\rho(\delta) \to 0$ for $\delta \to 0$, such that every closed curve C in each X_ε of diameter $\leq \delta$ bounds in its ρ-neighborhood. One knows (see [White]) that Y is a topological surface in this case, and similarly, every $U \subset Y$ without local cut points is locally $\mathbb{R}^2$. (The topology of the cut points in Y is rather transparent here, since Y is locally simply connected by an easy argument similar to the proof of surjectivity $\pi_1(X_\varepsilon) \to \pi_1(X)$.)

The above suggests nice properties of the Hausdorff (orbi-)space of surfaces of genus g reminiscent of the moduli spaces of holomorphic maps of surfaces (see [McD–Sal]). And, one is also tempted to look at general 2-polyhedra rather than surfaces.

Corollary*(Pointed out to me by S. Ivanov): Let Y' be the Hausdorff limit of surfaces X'_ε of genus g with boundaries. Then,* $\dim(Y) \leq 2$.

In fact, such Y' topologically embeds into the Y above, since every surface X' can be isometrically embedded into a closed surface X by filling each boundary component with a spherical cap.



Notice (also following S. Ivanov) that every *planar graph* appears in the limit of surfaces of genus zero, and so the Hausdorff limits of planar graphs embed into the bubble spaces above (of genus zero).

3.33. Approximating compact metric spaces by Riemannian manifolds. If $n \geq 3$, then there is little correlation between the topology of approximating spaces X_ε^n and the limit space X. In particular, if X is a compact, connected ANR, e.g., a finite polyhedron, then one has the:

Ferry–Okun Approximation Theorem. *Let X_0 be a compact. smooth manifold of dimension* > 3. *Then X_0 admits a family of Riemannian metrics g_ε, such that $(X_0, g_\varepsilon) \xrightarrow{\text{Hau}} (X, d)$ for a given path metric d on X, if and only if there is a continuous map $X_0 \to X$ surjective on the fundamental groups.*

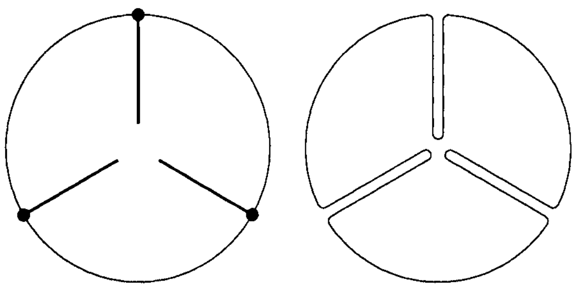
Remarks: (*a*) The "only if" part is trivial, see 3.31.

(*b*) According to Bing's theorem cited in 1.26_+, every compact, connected ANR admits a path metric since every compact ANR is locally contractible and vice versa.

(*c*) One does not truly need X to be an ANR, but just the existence of a continuous map $f : X_0 \to X$, where $f^{-1}(x) \subset X_0$ is connected for all $x \in X$. Such a map is known to exist under our assumptions (see $[\mathrm{Fer}]_{\mathrm{UV}^k}$), and starting from such an f, the construction of g_ε is not hard (see $[\text{Fer–Ok}]_{\mathrm{ATM}}$).

Examples: (*a*) Let X be a closed n-dimensional manifold admitting a closed subset Y of codimension ≥ 2, such that the complement is homeomorphic to $\mathbb{R}^n$. (This is equivalent to $\pi_1(X) = 0$ by Morse-Smale theory for $n \geq 5$ and by Freedman's theorem for $n = 4$, while the case $n = 3$ amounts to the Poincaré conjecture.) Then X can be approximated by n-balls with path metrics, which is especially clear if X is a Riemannian manifold and Y is a piecewise smooth subpolyhedron. In this case, we consider small ε-neighborhoods $U_\varepsilon(Y)$ and observe that the complements $X_\varepsilon = X \setminus U_\varepsilon(Y)$ with induced *path* metrics Hausdorff-converge to X. (This fails to be true for, say, hypersurfaces Y in X since the complement is disconnected near the points $y \in Y$.)

(*b*) Take the n-ball B^n with some triangulation, take the cone from the center over the k-skeleton and cut away a small ball around the center. Then move the boundary sphere inside B along the resulting $(k+1)$-dimensional protrusion.



If $n - 2 > k \geq 1$, then the induced Riemannian (i.e., path) metric on the protruded (or rather intruded) sphere is close to the induced metric, and for a sufficiently fine triangulation we get a reasonable approximation of

B^n by (S^{n-1}, g_ε). This is easy to adjust to an actual convergence $(S^{n-1}, g'_\varepsilon) \to B^n$, which allows the Hausdorff approximation of n-balls by $(n-1)$-spheres for $n \geq 4$. Since n-spheres can be approximated by n-balls for $n \geq 2$ (just look at $S^n \setminus \{0\}$), we see with (a) that every compact, simply connected manifold X can be Hausdorff-approximated by n-spheres for each $n \geq 3$. (Notice that we excluded the Poincaré 3-spheres Σ in (a), but these can be approximated by $\Sigma \times \varepsilon S^2$ for $\varepsilon \to 0$ and (a) does apply to $\Sigma \times S^2$ by Smale's theorem.)

3.34. On the regularity of approximation. Notice that the approximation in (a) above is by far more regular than in (b). In fact, one can use the construction of (a) in order to achieve the following

Regular Approximation. *Let X be a closed, simply connected manifold of dimension ≥ 5 with a piecewise Euclidean path metric. Then X can be Hausdorff-approximated by piecewise Euclidean n-balls X_ε with* $\mathrm{geo}(X_\varepsilon) \leq \mathrm{const} = \mathrm{const}(X)$.

(The proof is straightforward and is left to the reader).

Remarks: The approximation above works with $(n+1)$-spheres instead of n-balls, since $B^n = S^{n+1}/S^1$. But, it is impossible with n-spheres X_ε if X is nonhomeomorphic to S^n, since the maps $\varphi\colon X_\varepsilon \to X$ will have rather disconnected preimages as $\deg(\varphi_\varepsilon) \neq 1$ in the above sense.

Conjecture: The regular approximation $X_\varepsilon \xrightarrow{\text{Hau}} X$ with all X_ε homeomorphic to a fixed X_0 is possible whenever $\dim X_0 > \dim X$ and there is a map $X_0 \to X$ surjective on π_1. Furthermore, if $\dim X_0 = \dim X$, one probably only needs a map $X_0 \to X$ of degree one for such an approximation.

Question: Is it possible to define some regularity in purely topological terms, e.g., by requiring the spaces X_ε, or even better λX_ε for $\lambda \geq 1$, to have a uniform bound on their local topology? Some steps in this direction are indicated in 3.35 below.

$3.34\frac{1}{2}_+$ Regular convergence in $\mathbb{R}^n$. Our guiding example is given by semialgebraic subsets $X_\varepsilon \subset \mathbb{R}^n$ of degrees bounded by a constant as they Hausdorff-converge to some $X \subset \mathbb{R}^n$. This limit is again a semialgebraic subset with $\deg(X) \leq \sup \deg(X_i)$. We want to extend this picture to more general closed subsets X in $\mathbb{R}^n$, i.e., to define some measure of complexity, or a sequence of such measures, such that the complexity of every compact, C^∞-smooth submanifold is *finite*, and then we would study the Hausdorff limits $X_\varepsilon \to X$, where the complexities of X_ε are bounded by a fixed constant. We would like such limits X to be as similar as possible to diffeomorphic images of semialgebraic sets. Here is a possible candidate for such complexity.

Definition of $\mathrm{vol}^{(r)}(X)$. Consider a smooth, n-dimensional submanifold X in a smooth manifold Y and let $Y' = \mathrm{Gr}_n(Y)$ denote the space of the tangent n-planes in Y. The submanifold X naturally lifts to Y', then to $Y'' = \mathrm{Gr}_n(Y')$, to $Y''' = \mathrm{Gr}_n(Y'')$, etc., and the i-th lift is denoted by $X^{(r)} \subset Y^r$. We apply this construction to $Y = \mathbb{R}^N$, where all Y^r carry natural Riemannian metrics and set

$$\mathrm{vol}^{(r)}(X) = \mathrm{vol}_n X^{(r)}.$$

(One can think of $\mathrm{vol}^{(r)}$ as a geometric counterpart of Sobolev's norm $||\partial^r f||_{L^1}$ on functions f.)

Observation: *If* $X \subset \mathbb{R}^N$ *is semialgebraic, then*

$$\mathrm{vol}^{(r)}(X) \leq \mathrm{const} = \mathrm{const}(r, N, \deg(X), \mathrm{diam}(X)).$$

In fact, intersect X with a generic $(N-n)$-plane $P \subset \mathbb{R}^N$ and notice that $\mathrm{card}(P \cap X) \leq \deg(X)$ by Bezout's theorem. Then by the Buffon–Crofton formula, we evaluate $\mathrm{vol}^{(1)} = \mathrm{vol}_n(X)$ by

$$\mathrm{vol}_n(X) = \int \mathrm{card}(P \cap X)\, dP \leq c_N(\mathrm{diam}(X))^n \deg(X).$$

This equally applies to all $X^{(t)}$ and yields our inequality (where the reader is welcome to furnish the details).

Problem: Study the geometry of compact subsets $X \subset \mathbb{R}^N$ which are Hausdorff limits of smooth n-dimensional submanifolds X_ε with $\mathrm{vol}^{(r)}(X_\varepsilon) \leq \mathrm{const}_r$ for $r = 1, 2, \ldots$.

Exercise: Show that if such a limit for $n = 1$ is a piecewise smooth curve in $\mathbb{R}^N$, then the number of breaks of the k-th derivative of X is controlled by vol^{k+2}.

Remark: The above probably generalizes to piecewise smooth $X \subset \mathbb{R}^N$ with $\dim X \geq 2$. But more complicated singularities are not necessarily reflected in $\mathrm{vol}^{(r)}(X)$, not even for $\dim X = 1$, as was recently pointed out to me by Joseph Fu. For example, the graph $X \subset \mathbb{R}^2$ of the function $y = x^{(k+1)/k}$, $k = 1, 2, \ldots$ becomes nonsingular as we pass to $X^{(r)}$ with ≥ 1. Then, following Fu, one easily makes a C^1-curve X in the unit disk with infinitely many C^2-singularities, where all $X^{(r)}$ are C^∞-smooth for $r \geq 1$:

take the graph of $f(x) = \int \varphi(x)\,dx$, where the graph of φ is C^∞-smooth with $\varphi'(x) = \infty$ at infinitely many values of x.

Questions: How much does the condition $\operatorname{vol}^{(r)} X_\varepsilon \le \mathrm{const}_r < \infty$ control the singularities of $X = \lim_{\varepsilon\to\infty} X_\varepsilon$? What should one add to $\operatorname{vol}^{(r)} < \infty$ to make X stratifiable? Which global (topological and geometric) invariants of smooth (and more generally, stratified) X can be bounded in terms of $\operatorname{vol}^{(r)} X$, $r = 1, 2, \dots$? (See [Fu]$_{\mathrm{CMCC}}$ and [Fu]$_{\mathrm{MAF}}$ for the case $r = 1$.) What is the effect of $\operatorname{vol}^{(r)}(X)$ (and other geometric complexities for that matter) on the rate of Hausdorff-approximation of X by semialgebraic subsets Y? Here we seek $Y = Y_\varepsilon$ of minimal possible degree $\deg_\varepsilon = \deg Y_\varepsilon$ such tht $\mathrm{dist}_{\mathrm{H}}(X, Y_\varepsilon) \le \varepsilon$ and study the asymptotic behavior of $\deg_\varepsilon$ for $\varepsilon \to 0$. Also, we can look at $\mathrm{dist}_{\mathrm{H}}(X^{(r)}, Y_\varepsilon^{(r)})$ for $r \ge 2$.

Intrinsic $\operatorname{vol}^{(r)}$. Now let X be a Riemannian manifold and try to define some $\operatorname{vol}^r(X)$ in Riemannian terms. One possibility is to look at all path isometric embeddings $X \to \mathbb{R}^N$ for large N and minimize $\operatorname{vol}^{(r)}$ for the images of these embeddings. (Or, one may use some canonical embedding of X into an L_2-space of functions or tensors on X.) Similarly, one may define $\mathrm{dist}_{\mathrm{H}}^{(r)}(X, Y)$ via path-isometric embeddings $X, Y \to \mathbb{R}^N$.

Alternatively, one may look at the integrals of suitable expressions in the curvature tensor and its derivatives. Here one must keep an eye on the picture of $X = X_0 \times S^1_\varepsilon$, where X_0 is a badly curved manifold and S^1_ε is a very short circle. We want the curvature of X_0 to be felt by our integrals over X no matter how small ε can be. This can be achieved by normalizing such integrals with $(\operatorname{vol}(X))^{-\alpha}$ or $(\mathrm{InjRad})^{-\beta}$, or with something else of this kind.

Another possibility is to look at the manifold X' of orthonormal frames in X with the metric naturally defined with the Levi-Civita connection in X. Then one defines $X^{[2]} = (X')', \dots, X^{[r]} = (X^{[r-1]})'$ and measures the r-th complexity of X by the Hausdorff distance from $X^{[r]}$ to the cartesian product of X by the fiber of the fibration $X^{[r]} \to X$ (or something else in this spirit).

Ultimately, we can measure the complexity of X by the minimal number $d = d_\varepsilon(X)$ of "ε-standard pieces" into which one can decompose X. Here, one has to specify what "ε-standard" means. For example, one might declare the convex simplices in $\mathbb{R}^n$ to be "standard," so that "ε-standard" would then mean "being within Lipschitz distance $\le \varepsilon$ from something standard." (Being more generous, one may also regard simplices like $\{x, y : 0 \le x \le 1, 0 \le y \le x^\alpha\}$ as "standard" (compare [Gro]$_{\mathrm{SGSS}}$).)

Concluding remark. To bring the discussion back to solid ground, one has to:

(a) analyse the relation between different notions of complexity.

(b) evaluate the complexity for concrete examples, e.g., decide when $\mathrm{vol}^{(r)}$ of the smooth locus of a minimal variety in $\mathbb{R}^N$ is finite; pose and solve similar problems for Einstein manifolds and their singular limits.

We hope that the reader finds the perspective attractive. A step in a somewhat different direction is indicated below.

3.35. Hausdorff convergence and collapse. There are several definitions expressing the idea that a family X_ε of n-dimensional spaces collapses, for $\varepsilon \to 0$, to something of dimension $< n$.

(1) *Uryson width and collapse.* Define $\mathrm{wid}_k(X)$, also denoted $\mathrm{diam}_k(X)$, as the infimum of the numbers δ such that X admits a continuous map into a k-dimensional polyhedron, say $f\colon X \to P$, such that

$$\mathrm{diam}\, f^{-1}(p) \leq \delta$$

for all $p \in P$. Then the convergence $\mathrm{wid}_{n-1} X_\varepsilon \to 0$ signifies some kind of collapse, called the *Uryson collapse* of X_ε.

(2) *Kuratowski embedding and the filling radius.* Let X be a closed, oriented manifold embedded into a metric space Y. Then $\mathrm{FilRad}(X \subset Y)$ is defined as the infimal ρ such that X bounds in its ρ-neighborhood, i.e., the fundamental class $[X] \in H_n(X)$ goes to zero under the inclusion $X \hookrightarrow U_\varepsilon(X)$. (If X is nonorientable, one makes this definition with $\mathbb{Z}_2$-coefficients.) Next, we observe that every compact metric space X isometrically embeds into the space Y of functions on X with the uniform norm, $x \mapsto \mathrm{dist}(\cdot, x) \in Y$. Then

$$\mathrm{FilRad}(X) \stackrel{\mathrm{def}}{=} \mathrm{FilRad}(X \subset Y)$$

for the Kuratowski embedding $X \subset Y$.

Exercise: Show that $\mathrm{FilRad}(X)$ equals the infimal ρ such that there exists a pseudomanifold X with a boundary endowed with a metric such that X is isometric to the boundary $\partial X' | \mathrm{dist}_{X'}$ and $\mathrm{dist}(x', \partial X') \leq \rho$ for all $x' \in X$.

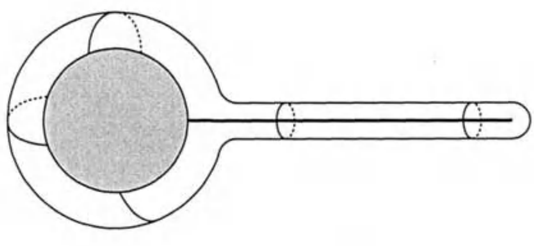
The definition of FilRad extends to manifolds X with boundaries by tracing the vanishing of $[X] \in H_n(X, \partial X)$ as one enlarges X to $U_\rho(X)$ and ∂X to $U_\rho(\partial X)$ in the space of functions on X (see $[\mathrm{Gro}]_{\mathrm{FRM}}$).

The *filling collapse* of X_ε is defined by requiring that $\mathrm{FilRad}\, X_\varepsilon \to 0$. Since $\mathrm{FilRad} \leq \mathrm{wid}_{n-1}$ (by a trivial argument, see $[\mathrm{Gro}]_{\mathrm{FRM}}$), the Uryson collapse implies the filling collapse.

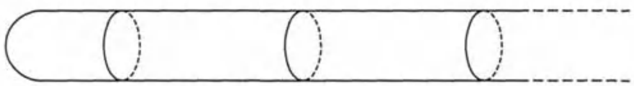
Example: We shall see in Ch. 8_+ that if X_ε are Riemannian manifolds with bounded curvatures $|K(X_\varepsilon)| \leq \text{const}$, then the filling collapse is equivalent to injRad $\to 0$ (at all points in X_ε) which is, in turn, equivalent to Uryson's collapse. In fact, the Uryson and filling collapses are also equivalent under the one-sided bound $K(X_\varepsilon) \geq -\text{const}$ (see [Grov–Pet], $[\text{Per}]_{\text{SCBB}}$).

Remarks: (*a*) The collapse does not exclude $\dim \lim_{\text{H}} X_\varepsilon \geq \dim X_\varepsilon$. For example, we can first approximate S^n for $n > 3$ by X_ε homeomorphic to S^3 and then take collapsing(!) $X'_\varepsilon = X_\varepsilon \times \varepsilon S^{n-3}$, which also converge to S^n.

(*b*) One may have a partial collapse of X_ε, as for the boundary of the ε-neighborhood of an n-ball in $\mathbb{R}^n$, $n \geq 3$, with a segment sticking out of it (compare (*b*) in $3.33\frac{1}{2}$).



3.36. Controlled contractibility against collapse. Let $\rho = \rho(r) \geq r$ be a (control) function for $0 \leq r < R$ and say that a metric space X is ρ-contractible if every subset in X inside a ball of radius r can be contracted within the concentric ball of radius $\rho(r)$. When this applies for $R = \infty$ (i.e., for all $r \geq 0$), this implies the uniform contractibility of X, which, for noncompact X is strictly stronger than just contractibility, no matter how fast $\rho(r)$ may grow. Here is a contractible but not uniformly contractible $\mathbb{R}^2$:



Exercise: Show that every complete, uniformly contractible surface has infinite area (see $[\text{Gro}]_{\text{FRM}}$ for the case $\dim \geq 3$).

In what follows, we shall be concerned with $R < \infty$ and families X_ε of compact spaces. Such X_ε are called *controllably locally contractible* if they are ρ-contractible for a continuous function on a segment $[0, R]$ vanishing at $r = 0$.

Observation: *A controllably locally contractible family of n-dimensional manifolds does not (filling) collapse.*

Proof. If $X \subset Y$ is ρ-contractible, then every $(n+1)$-dimensional polyhedron lying in the ε-neighborhood $U_\varepsilon(X) \subset Y$ for a small $\varepsilon \leq \varepsilon_0 = \varepsilon_0(R,\rho) > 0$, retracts onto X by a standard argument of the induction by skeletons. It follows that $[X]$ does not vanish in $U_\varepsilon(X)$ and so $\mathrm{FilRad}(X) \geq \varepsilon_0(R,\rho)$.

Remarks: (a) This argument also applies to balls in X_ε and shows that X_ε *nowhere collapses*, i.e., no sequence of balls $B(x_\varepsilon, r_\varepsilon) \subset X_\varepsilon$ with $r_\varepsilon \geq r_0 > 0$ collapses.

(b) The collapse could have been as easily ruled out by *controlled local acyclicity*, which is somewhat weaker than such contractibility.

Non-collapsed convergence $X_\varepsilon \xrightarrow{\mathrm{Hau}} X$ imposes nontrivial restrictions on the topology of X, e.g., $\dim(X)$ cannot be smaller than $n = \dim(X_\varepsilon)$. But, the dimension may jump up quite dramatically.

There exists a family of Riemannian manifolds X_ε which are controllably locally contractible, all diffeomorphic to S^5, and which Hausdorff converges to an infinite-dimensional space X. (See [Fer–Ok]$_{\mathrm{ATM}}$).

The space X above is an instance of a *Dranishnikov manifold*, which looks homologically, locally and globally, like the sphere S^5 but which nevertheless has infinite *Lebesgue* covering dimension.

In general, spaces X appearing in the completion of the space $\mathcal{X}(n,\rho)$ of n-dimensional Riemannian manifolds with local contractibility control by a given $\rho = \rho(r)$, $r \in [0,R]$, has been studied in [Fer]$_{\mathrm{CSHT}}$, [Fer]$_{\mathrm{TFT}}$ and [Dra–Fer]. For example, it is shown in [Fer]$_{\mathrm{TFT}}$ that if $X_i \in \mathcal{X}(n,\rho)$ converge to some X, then for $n \neq 3$, *there are at most finitely many mutually nonhomeomorphic manifolds among the X_i, and moreover if X is finite-dimensional, then all X_i sufficiently close to X (i.e., for $i \geq i_0$) are mutually homeomorphic.* (It is obvious that any two manifolds X_i and $X_{i'}$ close to X are homotopy equivalent).

It would be interesting to extend the results above to the situation where a partial collapse is allowed, as in 3.32 for graphs and surfaces. For example, one can look at n-dimensional manifolds X which are locally ρ-contractible (or just acyclic) up to dimension $(n-2)$, as is the case with the surfaces (see 3.32). Here again, collapse is reflected in cut-points in the limit. Then one wishes to admit some kind of "stratified collapse" including, for example, $X_\varepsilon \times Y \xrightarrow{\mathrm{Hau}} X \times Y$ for the above $X_\varepsilon \xrightarrow{\mathrm{Hau}} X$ with possible collapse at the cut-points in X, and, more generally, the collapse pattern as complicated as for

$$X_\varepsilon \times X'_\varepsilon \times X''_\varepsilon \times \cdots \xrightarrow{\mathrm{Hau}} X \times X' \times X'' \times \cdots$$

In fact, one can enhance convergence by cutting away the collapsing parts of X_ε. Namely, for a given δ and r, one removes from X_ε the δ-collapsed balls B of radius r, where "δ-collapsed" may mean one of the following:

1. $\mathrm{wid}_{n-1} B \leq \delta$,
2. $\mathrm{FilRad}\, B \leq \delta$,
3. B is not contractible in the concentric ball of radius $\delta^{-1}r$,
4. something else in the same spirit.

We denote the remaining *δ-thick* part of X_ε by $X_\varepsilon(\delta, r)$, pass to the limit as $\varepsilon \to 0$, send $\delta \to 0$, and finally let $r \to 0$. The resulting space X^{thick} coincides with $X = \lim_{\mathrm{H}} X_\varepsilon$ in the noncollapsed case, and it may exist in the cases where the original family X_ε admits no Hausdorff convergent subsequence. The most satisfactory picture arises if one can find a thick part $X'_\varepsilon(\delta, r)$ of X with boundary $\partial X'_\varepsilon(\delta, r)$ having $\mathrm{wid}_{n-r} \to 0$ for $\varepsilon, \delta \to 0$, while the interior of $X'_\varepsilon(\delta, r)$ has controlled local contractibility r-far from the boundary (as happened in 3.35 (b)). The question which remains open is that of finding reasonable local conditions which would imply such behavior.

Now, following the ideas of geometric measure theory, we indicate in the following sections another way to cut away undesirable thin parts of X_ε.

Chapter $3\frac{1}{2}{}_+$

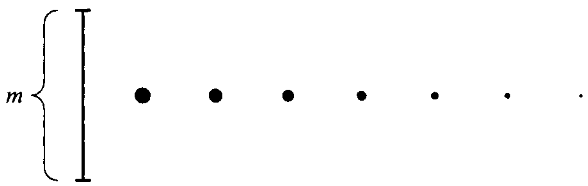
Convergence and Concentration of Metrics and Measures

A. A review of measures and mm spaces

$3\frac{1}{2}$.1. When we speak of measures μ on a metric space X, we always assume that X is a *Polish space*, i.e., *complete with a countable base*, and that "measure" means a *Borel measure*, where all Borel subsets in X are measurable. We are mostly concerned with *finite measures* where X has *finite (total) mass* $\mu(X) < \infty$, but we also allow *σ-finite measure spaces* X, which are the countable unions of X_i with $\mu(X_i) < \infty$. For example, the ordinary n dimensional Hausdorff measure in $\mathbb{R}^n$ is σ-finite, while the k-dimensional Hausdorff measure on $\mathbb{R}^n$ for $k < n$ is not σ-finite. But, we may restrict such a measure to a k-dimensional submanifold $V \subset \mathbb{R}^n$, i.e., we declare $\mu_k(U) = \mu_k(U \cap V)$ for all open $U \subset \mathbb{R}^n$, in which case the measure becomes σ-finite and admissible in our discussion.

Polish metric spaces with σ-finite measures, denoted $X = (X, \mathrm{dist}, \mu)$ are called *mm spaces.*

Reversed definition of mm spaces: One could start with an abstract measure space (X, μ) and then introduce a metric as a measurable function d on $X \times X$ satisfying the metric axioms. If μ appears as a Borel measure on a Polish space, as we will always assume to be the case, then (X, μ) is a *Lebesgue (Rochlin) space*, i.e., it is measure-isomorphic to a real segment of length $m \leq \infty$ with a finite or countable collection of *atoms*, i.e., points x_i with positive measures (masses) $m_i = \mu(x_i)$. Here is a picture of such a space with atoms ordered according to their mass.



If, for example, we restrict ourselves to measures of *mass one without atoms*, then all mm spaces are representable by measurable functions d on the square $[0,1]^2$.

Notice that every finite measure μ can be *normalized* to mass one by setting $\nu(U) = \mu(U)/\mu(X)$ for all $U \subset X$. Such normalized measures are often called *probability measures*, and $\nu(U)$ is interpreted as the probability of the event $x \in U$.

Our basic example is a Riemannian manifold V of dimension n with n-dimensional *Riemannian volume (measure)* vol_n. This can be defined as the *n-dimensional Hausdorff measure* (using countable coverings by balls) or axiomatically by observing that vol_n is *uniquely* characterized by the following two properties:

1. *Monotonicity.* If V_1 admits a bijective 1-Lipschitz map onto V_2, then $\mathrm{vol}_n(V_1) \geq \mathrm{vol}_n(V_2)$.

2. *Normalization.* The Euclidean cube $[0,1]^n$ has $\mathrm{vol}_n([0,1]^n) = 1$.

We are also interested in non-Riemannian measures, such as the *Gaussian* measure on $\mathbb{R}^n$ given by $d\mu = e^{-\|x\|^2} dx_1 dx_2 \cdots dx_n$ confronted with the *ordinary Euclidean* metric on $\mathbb{R}^n$.

Pushforward measures. A continuous (or just Borel measurable) map $f : X \to Y$ *pushes forward* measures μ from X to $\mu_* = f_*(\mu)$ on Y by $\mu_*(U) = \mu(f^{-1}(U))$ for all $U \subset Y$.

Notice that the pushforward of a Riemannian measure is, typically, non-Riemannian, even if the map in question is quite regular. For example, if $f : V_1 \to V_2$ is a *smooth* map between equidimensional manifolds, then $f_*(\mathrm{vol}_n)$ is rarely Riemannian. For example, if V_1 is a closed manifold, then $f_*(\mathrm{vol}_n)$ is not Riemannian unless f is a covering map. On the other hand, there are quite a few (nonsmooth!) Lipschitz maps pushing forward Riemannian measures to Riemannian measures (see Exercise $3\frac{1}{2}.16$).

Examples (Archimedes). (a) The normal projection of the unit sphere

$\mathbb{R}^{n+1} \supset S^n \to \mathbb{R}^{n-1} \subset \mathbb{R}^{n+1}$ pushes forward the spherical measure to $2\pi\cdot$(Lebesgue measure) on the unit ball in $\mathbb{R}^{n-1}$.

(*b*) The "squaring map" from the sphere $S^{2n+1} \subset \mathbb{C}^{n+1}$ to the unit simplex $\Delta^n \subset \mathbb{R}_+^{n+1}$ given by

$$(z_0, \dots, z_n) \mapsto (|z_0|^2, \dots, |z_n|^2),$$

which is an instance of a *moment map*, sends the spherical measure to $(2\pi)^n\cdot$(Lebesgue measure) on Δ^n.

(*c*) The normal projection of the unit ball $B^{n+1} \subset \mathbb{R}^{n+1}$ to $\mathbb{R}$ pushes forward the Lebesgue measure on B^n to $(\sqrt{1-x^2})^n\,dx$ on $[-1,1] \subset \mathbb{R}$.

We present below some standard facts about measures and mm spaces, where the routine proofs are left to the reader (who may also consult [Parth]).

The metric me_λ on maps $X \to Y$. Here, (X,μ) is a measure space and (Y,dist) is a metric space. Given maps $f,g\colon X \to Y$, we define $\mathrm{me}_\lambda(f,g)$ for $\lambda \geq 0$ as the *minimal* or, better, *infimal* number $\varepsilon > 0$ such that f *and* g *are* ε *close in* Y *away from a subset of measure* $\lambda\varepsilon$ *in* X,

$$\mu\{x \in X : \mathrm{dist}(f(x), g(x)) \geq \varepsilon\} \leq \lambda\varepsilon.$$

The metrics me_λ for different $\lambda > 0$ are (obviously) mutually bi-Lipschitz equivalent. Thus, the notion of di_λ-convergence is independent of $\lambda > 0$ and is customarily referred to as *convergence in measure.* Observe that me_λ is not, strictly speaking, a metric, since it can become infinite when $\mu(X) = \infty$ or, for finite $\mu(X)$ if $\lambda = 0$. But everything we say about me_λ still makes sense in these case if we restrict our attention to maps $f\colon X \to Y$ which lie within a finite distance from a fixed $f_0\colon X \to Y$.

Notice that the metric me_0 is quite different from the rest of the me_λ. It is called the *sup-metric* or L_∞ metric on maps:

$$\mathrm{me}_0 = \sup_{x \in X} \mathrm{dist}(f(x), g(x)).$$

An apology: There is something artificial in the way we make the metric out of ε and $\lambda\varepsilon$. It would be more logical to use both numbers independently, say ε and κ (instead of $\lambda\varepsilon$) and say: "f and g are (ε,κ) close if $\mu\{x \in X : \mathrm{dist}(f(x), g(x)) \geq \varepsilon\} \leq \kappa$." To make this consistent, we need a generalized notion of metric space, where the distance is measured by a pair of numbers, or rather by a subset in $\mathbb{R}^2$. Ultimately, the metric should be replaced by a certain partial order on $X \times X$, but the self limitations of the present exposition force us into a procrustean bed of metric spaces.

(1) Exercises: (*a*) Show that if Y is a *complete* metric space, then so is the space of maps $X \to Y$ with the metric me_λ for $\lambda \geq 0$. Furthermore, if Y is Polish, then the space of maps is also Polish for me_λ if $\lambda > 0$.

(*b*) Recall that the L_p-metric between f and g is defined for $1 \leq p < \infty$ by

$$\mathrm{dist}_{L_p}(f,g) = \left(\int_X (\mathrm{dist}(f(x),g(x))^p \, d\mu\right)^{1/p},$$

and observe that me_λ for $\lambda > 0$ can be bounded by dist_{L_p} via the famous (and obvious) *Chebyshev inequality*

$$\int_X (\mathrm{dist}(f,g))^p \, d\mu \geq \varepsilon^p \mu\{x \in X : \mathrm{dist}(f(x),g(x)) \geq \varepsilon\}$$

for all $\varepsilon > 0$ and $p > 0$.

(*c*) Let $d \colon X \times X \to \mathbb{R}_+$ be a symmetric measurable function satisfying the triangle inequality. Invoke the Kuratowski map from X to the space $L_\infty(X)$ of functions $X \to \mathbb{R}$ with the L_∞ metric $K : X \to L_\infty(X)$ for $K \colon x \mapsto d_x(y) = \mathrm{dist}(x,y)$. Denote by μ_* the pushforward of the measure μ from X to $L_\infty(X)$. Recall that μ is (assumed) Lebesgue and show that the L_∞-closure of the image $K(X) \subset L_\infty(X)$ has countable base. Thus, this closure $(Cl(K(X)), \mu_*, \mathrm{dist}_{L_\infty})$ is an mm-space in our sense. Furthermore, if d does not vanish away from the diagonal in $X \times X$, then K is injective (as well as isometric for $d \leftrightarrow \mathrm{dist}_{L_\infty}$), and so $CL(K(X))$ equals the completion of X with respect to d.

$3\frac{1}{2}$.2. The metric $\square_\lambda$ on maps $d \colon X \times X \to Y$. One could use me_λ for $(X \times X, \mu \times \mu)$, but there is another metric, denoted $\square_\lambda(d,d')$ which is defined as the supremal ε such that d and d' are ε-close in Y outside a subset $X_\varepsilon \subset X$ of measure $\leq \lambda\varepsilon$, i.e.,

$$\mathrm{dist}(d(x_1,x_2), d'(x_1,x_2)) \leq \varepsilon$$

for $(x_1,x_2) \in (X \setminus X_\varepsilon) \times (X \setminus X_\varepsilon)$. Clearly $\square_\lambda \geq \mathrm{me}_\lambda$.

In what follows, we will be concerned with $\square_\lambda$ on the space D of symmetric functions $d \colon X \times X \to \mathbb{R}$ satisfying the triangle inequality.$\square_\lambda$

Exercise: Show that $(D, \square_\lambda)$ is a Polish path metric space for all $\lambda > 0$.

B. $\square_\lambda$-convergence of mm-spaces

$3\frac{1}{2}$.3. Every Lebesgue space (X,μ) can be *parametrized* by the segment $[0,m)$ for $m = \mu(X)$, where the parametrization refers to a *measure-preserving* map $\varphi : [0,m) \to X$ (which is necessarily surjective up to measure zero) and where "measure preserving" means $\varphi_*(\nu) = \mu$ for the Lebesgue measure ν on $[0, m)$. Notice that such a φ is by no means unique, and it can be chosen to be bijective if X contains no atoms.

If X comes along with a metric as well as μ, one can pullback this metric $\text{dist}\colon X \times X \to \mathbb{R}$ to a function d on $[0, m)^2$ that is given by

$$d(t, t') = \text{dist}(\varphi(t), \varphi(t')).$$

Definition of $\underline{\square}_\lambda$. Let X and X' be two mm-spaces of equal mass. Parametrize X and X' by $[0, m)$, pullback the metrics dist on X and dist' on X' to functions d and d' on $[0, m)^2$, and set

$$\underline{\square}_\lambda(X, X') = \inf \square_\lambda(d, d'),$$

where the infimum is taken over all parametrizations φ and φ' of X and X' by $[0, m)$.

The case $m \neq m'$. If X and X' have unequal masses $m = \mu(X)$ and $m' = \mu(X')$, then there are several possibilities for defining the distance $\underline{\square}_\lambda(X, X')$. First of all, we set $\underline{\square}_\lambda(X, X') = \infty$ if $m < \infty$ while $m' = \infty$, so we are left with the case where both masses m and m' are finite. Here, we must agree on what is the distance between (X, μ, dist) and $(X, r\mu, \text{dist})$ for a given $r > 0$. It could be $|\log(r)|$, but we will stick to $|r - 1|\mu(X)$. Then we define $\underline{\square}_\lambda(X, X')$ for $m < m'$ as the maximal metric compatible with this convention. This is equivalent (by an easy argument) to

$$\underline{\square}_\lambda(X, X') = m' - m + \underline{\square}_\lambda(X, X'_m),$$

where $X'_m = (X', (m/m')\mu', \text{dist}')$, provided $m' \geq m$. (If we had used the $|\log|$ alternative, we would set

$$\underline{\square}_\lambda(X, X') = |\log(r)| + \underline{\square}_\lambda(X_1, X'_1),$$

where the subscript "1" signifies the normalization of the measure $\mu \rightsquigarrow (1/m)\mu$.)

Later on (see $3\frac{1}{2}.7$), we shall see that $\underline{\square}_\lambda$ is indeed a metric on *the isomorphism classes* of mm-spaces of finite mass, where we use the following

Definition of $\text{Supp}\,\mu$ and of mm-isomorphisms. The *support* of a measure μ on a metric space X is the *minimal closed* subset $Y \subset X$ such that $\mu(X \setminus Y) = 0$. Two mm-spaces X and X' are called *mm-isomorphic* if there exists an isometry between the supports of their respective measures, say $\text{Supp}\,\mu \leftrightarrow \text{Supp}\,\mu'$, such that μ goes to μ' under this isometry. Thus, every (X, μ, dist) is isomorphic to $(\text{Supp}\,\mu, \text{dist})$, and so dealing with $\underline{\square}_\lambda$

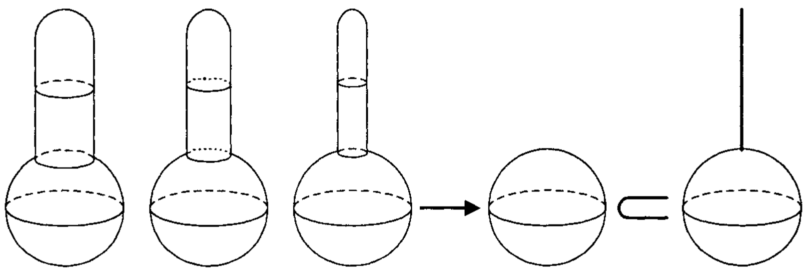
and with isomorphisms between mm spaces in general, we assume that the measures $\mu = \mu_X$ in question have $\operatorname{Supp} \mu = X$.

$\underline{\square}_\lambda$-convergence. We say that a sequence X_i of mm-spaces $\underline{\square}_\lambda$-converges to $X = (X, \mu, \mathrm{dist})$ if

$$\underline{\square}_\lambda(X, X_i) \to 0$$

as $i \to \infty$. Notice that this definition implies that X and almost all X_i have simultaneously finite or infinite masses and $m_i = \mu(X_i) \to m = \mu(X)$. In what follows, we will mainly be concerned with the case of mm-spaces of finite mass, while the case of infinite measure could be treated in the spirit of our approach to the Hausdorff convergence of noncompact spaces.

Example: Consider a sequence of spheres with thin fingers sticking out with areas going to zero.



For each $\lambda > 0$, the $\underline{\square}_\lambda$-limit equals the original sphere, while the Hausdorff limit contains a straight segment as a remnant of the fingers. Also, if $\lambda = 0$, then there is no limit at all.

Exercises: (a) Let a sequence of compact metric spaces X_i Hausdorff-converge to a compact space X. Show that for every sequence of measures μ_i on X_i with $\mu_i(X_i) \leq m < \infty$, there exists a subsequence $(X_{i_j}, \mu_{i_j}, \mathrm{dist}_{i_j})$ which $\underline{\square}_1$-converges to X with some measure μ on it. Moreover, if there is a function $\kappa(\rho) > 0$, $\rho > 0$, such that the ρ-balls in all X_i have $\mu_i(B(\rho)) \geq \kappa(\rho)$, then the μ above has $\operatorname{Supp} \mu = X$, and $(X_{i_j}, \mu_{i_j}, \mathrm{dist}_{i_j})$ converge to (X, μ, dist) with respect to the metric $\underline{\square}_0$.

(b) Let $X = (X, \mathrm{dist}, \mu)$ be an arbitrary mm space of *finite* mass $m = \mu(X)$. Show that $\varepsilon X = (X, \varepsilon\, \mathrm{dist}, \mu)$ $\underline{\square}_1$-converges to a one-point space of mass m as $\varepsilon \to 0$. This can also happen for $m = \infty$ if we allow atoms of infinite mass in the limit. But typically, there is no shade of convergence of εX to an atom for spaces of infinite mass; just consider $X = \mathbb{R}^n$ for instance.

(*c*) Study the convergence $X_i \to X$ where all spaces in question consist of at most N atoms. Notice that in the limit, the number of atoms may drop down for two reasons: either some atoms collide, or some disappear as their mass goes to zero.

(*d*) Let $x_1, x_2, \dots$ be a *uniformly distributed sequence* of points in an mm space X, that is,

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \to \int_X f\, d\mu$$

for every bounded Lipschitz function f on X. Show that the finite spaces

$$(\{x_1, x_2, \dots, x_N\}, \mathrm{dist}_X, \mu_N(x_i) = N^{-1}\mu(X))$$

converge to X in the $\underline{\square}_1$-metric, provided that $\mu(X) < \infty$.

(*e*) Let S^n be the unit sphere with the normalized $O(n+1)$-invariant measure. Show that $\underline{\square}_1(S^n, S^{n-1}) \to 0$ as $n \to \infty$, although the sequence of these spheres $\underline{\square}_1$-diverges as $n \to \infty$.

(*f*) Let X be a Riemannian manifold with a probability measure, say $X = (X, \mathrm{dist}, \mu)$, and let $\lambda_1 \geq \lambda_2 \geq \cdots$ be a sequence of positive numbers. Consider $\lambda_i X = (X, \lambda_i\, \mathrm{dist}, \mu)$ and take the cartesian product

$$X_n = (\lambda_1 X) \times \cdots \times (\lambda_i X) \times \cdots \times (\lambda_n X),$$

where we recall that the distance on X_n is

$$\mathrm{dist}_n((x_1, \dots, x_n), (y_1, \dots, y_n)) = \left(\sum_{i=1}^{n} (\lambda_i\, \mathrm{dist}(x_i, y_i))^2\right)^{1/2}$$

and $\mu_n = \mu \times \mu \times \cdots \times \mu$. Show that the sequence X_n $\underline{\square}_1$-converges to some mm space Y if the series $\sum_{i=1}^{\infty} \lambda_i^2$ converges.[1] Conversely, if the X_n converge, then either $\sum_{i=1}^{\infty} \lambda_i^2 < \infty$ or $\mathrm{diam}(\mathrm{Supp}\,\mu) = 0$ (i.e., μ is Dirac's δ). Show that if $X = (X, \mathrm{dist})$ is isometric to $\mathbb{R}^n$ and $\mathrm{Supp}\,\mu = \mathbb{R}^n$, then $Y = (Y, \mathrm{dist}_\infty)$ is isometric to a Hilbert space. Study specifically the case where (*i*) μ is the Gaussian measure on $\mathbb{R}$, i.e., $\mu = e^{-\pi x^2}\, dx$; (*ii*) X is a complete (possibly noncompact) Riemann surface of constant negative curvature of finite area and μ is the normalized Riemannian volume (area). (*iii*) X is the Heisenberg group with a left-invariant metric and some μ having $\mathrm{Supp}\,\mu = X$.

[1] This is false in general as Yann Ollivier pointed out to me. One needs a decay condition for μ on noncompact X.

(f') Generalize the above to cartesian products of arbitrary mm spaces. Also consider the case where dist_n is defined as $\left(\sum_{i=1}^n \mathrm{dist}_i^p\right)^{1/p}$ for an arbitrary $p \geq 1$.

Encouragement: If the reader finds these exercises difficult, we suggest plodding on through the subsequent sections, where all the problems are eventually solved.

$3\frac{1}{2}$.4. Identification of X by the measures μ_r^X. We recall that the *distance matrix* map from $\underbrace{X \times X \times \cdots \times X}_{r} \to M_r$ which assigns to each r-tuple of points the matrix of their mutual distances in X, and we denote by $\underline{\mu}_r = \underline{\mu}_r^X$ the pushforward of the measure $\mu \times \mu \times \cdots \mu$ to the space M_r of positive, symmetric $(r \times r)$-matrices. Notice that $\mathrm{Supp}\, \underline{\mu}_r \subset M_r$ equals our old friend $K_r(X) \subset M_r$, provided that $\mathrm{Supp}\, \mu = X$ (compare 1.19_+).

We have already seen that the "curvature" sets $K_r(X)$, for $r = 1, 2, \ldots$ give a full set of invariants for compact metric spaces X. Similarly, we have the following

$3\frac{1}{2}$.5. mm–Reconstruction theorem. *If two mm-spaces X and X' with finite measures have $\underline{\mu}_r^X = \underline{\mu}_r^{X'}$ for all $r = 1, 2, \ldots,$ then X is isomorphic to X'.*

Proof. Let us explain how to recapture the full geometry of X in terms of $\underline{\mu}_r = \underline{\mu}_r^X$. To grasp the idea, let us look at the function $\varphi(x) = \mu(B(\rho, x))$, where $B(\rho, x)$ denotes the open ball around x of a fixed radius ρ. We want to show that the pushforward measure $\mu_* = \varphi_*(\mu)$ on $\mathbb{R}$ can be reconstructed from $\underline{\mu}_r$. To see this, observe that the *first moment* of μ_*, i.e.,

$$\int_{\mathbb{R}} t \, d\mu_* = \int_X \varphi(x) \, d\mu$$

is readily available in terms of $\underline{\mu}_2$, since

$$\int_X \varphi(x) \, d\mu = \underline{\mu}_2\{m \in \mathbb{R} = M_2 : m \leq \rho\}.$$

Similarly, we compute the moments $\int_{\mathbb{R}} t^{r-1} \, d\mu_* = \int_X \varphi^{r-1}(x) \, d\mu$ using the obvious formula

$$\int_X \varphi^{r-1}(x) \, d\mu = \underline{\mu}_r\{m_{ij} \in M_r : m_{1j} \leq \rho, \, j = 2, \ldots, r\}$$

and conclude that

$$\underline{\mu}_r^X = \underline{\mu}_r^{X'} \Rightarrow \int_{\mathbb{R}} t^{r-1}(d\mu_* - d\mu'_*) = 0$$

(where we put "d" in front of "μ" following the phonetic tradition accepted by analysts). Now, two measures on $\mathbb{R}$ with compact supports and having equal moments for all r are equal, since polynomials $\sum_{i=0}^{r} a_i t^i$ are dense in the space of continuous functions on compact subsets in $\mathbb{R}$. Therefore, the functions $\varphi(x) = \mu(B(\rho, x))$ on X and $\varphi'(x) = \mu'(B(\rho, x'))$ on X' are equally distributed. It implies that *if X contains an r-ball of μ-measure t, then so does X' for almost all (with respect to μ) $t \in \mathbb{R}$.*

All of this obviously applies to the annuli

$$A(x) = A(x, \rho, R) = \{y \in X : \rho \leq \operatorname{dist}(x, y) \leq R\},$$

where one reconstructs the pushforward of μ under the function $\psi(x) = \mu(A(x))$. Furthermore, one can consider pairs of annuli and take the pair of functions $\psi_i(x) = \mu(A(\rho_i, R_i, x))$, $i = 1, 2, \ldots$. Such a pair pushes the measure μ from X to some μ_* on $\mathbb{R}^2$ where the moments $\int_{\mathbb{R}^2} t_1^{r_1 - 1} t_2^{r_2 - 1}\, d\mu_*$ can be computed as earlier in terms of $\underline{\mu}_r$ for $r = r_1 + r_2$. Then one generalizes to finite systems of annuli $A(\rho_i, \rho_{i+1}, x)$ for some $0 \leq \rho_1 \leq \cdots \leq \rho_i \leq \cdots \leq \rho_n$ and concludes that the measures $\underline{\mu}_r$ recapture the push forward of μ under the map ϕ from X to the space of measures on $\mathbb{R}$, where x goes by ϕ to the pushforward of μ under the function $d : y \mapsto \operatorname{dist}(y, x)$ on X. It follows that *for each $x \in X$, there exists a point $x' \in X$ such that the functions $d(y') = \operatorname{dist}_{X'}(y', x')$ pushes μ' to the measure on $\mathbb{R}$ equal to the pushforward of μ under d.*

Such an x' is constructed as the limit of a sequence of points giving equalities to the measures of annuli, where the convergence is obtained for a subsequence via the following obvious

$3\frac{1}{2}$.6. Subsequence lemma: *Let y_i be a sequence of points in an mm-space Y of finite mass. Then either y_i admits a convergent subsequence, or there exists $\varepsilon > 0$ such that the ε-balls around y_i have $\mu_Y(B(y_i, \varepsilon)) \to 0$ for $i \to \infty$.*

This lemma applies to sequences in X' where the relevant point $x \in X$ must have $\mu_X(B(x, \varepsilon)) > 0$ for all $\varepsilon > 0$. This is indeed the case, since we assume here that $X = \operatorname{Supp} \mu_X$ (as was indicated earlier).

Finally, we generalize the discussion above to finite configurations $(x_1, \ldots, x_k)$ of points in X, where we recapture the pushforward of μ under the map $X \to \mathbb{R}^k$ for $y \mapsto \{\operatorname{dist}(y, x_i)\}_{i=1,\ldots,k}$. Namely, we show in this way that *for each k-tuple of points $x_1, \ldots, x_k$ in X, there exist points $x'_1, \ldots, x'_k$ in X' such that*

$$\operatorname{dist}_X(x_i, x_j) = \operatorname{dist}_{X'}(x'_i, x'_j)$$

and the measures μ and μ' have equal pushforwards in $\mathbb{R}^k$ under the maps

$$y \mapsto \{\text{dist}(y, x_i)\} \quad \text{and} \quad y' \mapsto \{\text{dist}(y', x'_i)\}.$$

We conclude the proof of the theorem by taking a dense sequence $x_1, x_2, \dots, x_i, \dots$ in X and constructing a sequence $x'_1, x'_2, \dots, x'_i, \dots$ in X' with the above such that $\text{dist}_{X'}(x'_i, x'_j) = \text{dist}_X(x_i, x_j)$ for all $i, j = 1, 2, \dots$ and the maps $y'_k : X' \to \mathbb{R}^k$ for $y' \mapsto \{\text{dist}(y', x'_i)\}_{i=1,\dots,k}$ push forward μ' to the measure on $\mathbb{R}^k$ equal to the pushforward μ_* of μ under the corresponding map $X \to \mathbb{R}^k$. Then the map $\{x_i\} \to \{x'_i\} \subset X'$ extends by continuity to an isometric embedding $X \to X'$, where the equalities $\mu'_k = \mu_k$, $k = 1, 2, \dots$, show that this map pushes forward μ to μ'. Thus, it is the required isomorphism $X \simeq X'$.

Corollary: $\underline{\square}_\lambda$ *is a metric.*

Proof. The only point to check is the implication

$$\underline{\square}_\lambda(X, X') = 0 \Rightarrow X \simeq X'.$$

But it is obvious that the measures $\underline{\mu}_r^X$ are independent of parametrization and continuous in functions $d\colon [0, m]^2 \to \mathbb{R}$ with respect to $\square_\lambda$ for each $\lambda \geq 0$. Hence

$$\underline{\square}_\lambda(X, X') = 0 \Rightarrow \underline{\mu}_r^X = \underline{\mu}_r^{X'},$$

and the theorem applies.

This corollary justifies the notion of the $\underline{\square}_\lambda$-limit: if it exists, then it is *unique* in our category of mm-spaces.

$3\frac{1}{2}$.7. Another proof of the reconstruction theorem. The following elegant (albeit less elementary) argument was suggested to me by A. Vershik. Let M_∞ be the space of infinite matrices $\{d_{ij}\}$ for $i, j = 1, 2, \dots$, and let $X^{\mathbb{N}}$ denote the space of infinite sequences $x_i \in X$, $i \in \mathbb{N}$, with the product measure $\mu \times \mu \times \cdots$, where the original measure μ in X is assumed to be normalized. Then we map $X^{\mathbb{N}}$ to M_∞ as before by sending $\{x_i\}$ to $\{d_{ij} = \text{dist}_X(x_i, x_j)\}$, and we denote by $\underline{\mu}_\infty$ the pushforward of the measure $\mu \times \mu \times \cdots$ on M_∞ under this map.

Step A. If $\underline{\mu}_r^X = \underline{\mu}_r^{X'}$ for each $r = 1, 2, \dots$, then

$$\underline{\mu}_\infty^X = \underline{\mu}_\infty^{X'}.$$

In fact, the measure $\underline{\mu}_{r+1}^X$ goes to $\underline{\mu}_r^X$ under the projection $M_{r+1} \to M_r$, forgetting the last column and row, and the measure $\underline{\mu}_\infty^X$ appears as the projective limit of the measures for the tower of measure spaces

$$\cdots \to (M_r, \underline{\mu}_r^X) \to (M_{r-1}, \underline{\mu}_{r1}^X) \to \cdots$$

Step B. A generic sequence $\{x_1, x_2, \dots, \} \in X^{\mathbb{N}}$ is equidistributed in (X, μ), where "generic" means "almost surely" for the product measure on $X^{\mathbb{N}}$ (see $3\frac{1}{2}.22$).

Step C. If $\underline{\mu}_\infty^X = \underline{\mu}_\infty^{X'}$, *then there exist equidistributed sequences* $x_i \in (X, \mu)$ *and* $x_i' \in (X, \mu')$ *such that* $\mathrm{dist}(x_i, x_j) = \mathrm{dist}'(x_i', x_j')$ *for all* $i, j \in \mathbb{N}$. This means that there are points in $X^{\mathbb{N}}$ and $(X')^{\mathbb{N}}$ which correspond to *equidistributed* sequences in X and X' with equal images in M_∞. This is obvious, since the images of sets of generic points from X have full measure in M_∞ and the same is true for X' which correspond to *equidistributed* sequences in X and X'.

Step D. The map $x_i \to x_i'$ is isometric on our sequences, and it extends by continuity to an isometry $X \to X'$. And since these sequences are equidistributed, this isometry sends μ to μ'.

Remark: The reader may feel perplexed by how the seemingly spineless argument above could take care of such technical tools as the Weierstrass approximation theorem used in the first proof. But in fact one knows that this theorem follows from a suitably interpreted version of the law of large numbers, and the reader is invited to find this relation by comparing the two proofs of the reconstruction theorem.

A. Vershik suggested the following questions and remarks (a)-(d) to me.

(a) Positive matrices in M_∞ satisfying the triangle inequality are (essentially) metrics on $\mathbb{N}$, and the probability measures on M_∞ can be interpreted as random metrics on $\mathbb{N}$. The measures of the form $\underline{\mu}_\infty^X$ are invariant under the infinite permutation group S_∞ and are ergodic under this action, since the shift is ergodic on $X^{\mathbb{Z}}$. Can one directly describe these measures in terms of the space M_∞?

(b) One can think of ergodic S_∞-invariant measures on M_∞ as generalized (by how much?) mm-spaces and try to define basic mm invariants for them, such as the covering number, etc. (see $3\frac{1}{2}.\mathrm{F}_+$).

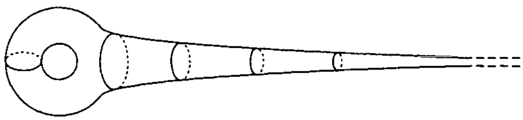
(c) One can define various natural metrics on measures on M_∞, e.g., using the Monge-Kantorovich-Rubenstein transportation metric on measures on M_r (compare $3\frac{1}{2}.10$) and letting $r \to \infty$. (These metrics refine the structure of our space $(\mathcal{X}, \underline{\square}_1)$, compare $3\frac{1}{2}.10$).

(d) The measures with the "maximal entropy" on M_∞ may be related to the Uryson space (see $3.11\frac{2}{3}{}_+$).

C. Geometry of measures in metric spaces

In order to clarify the idea of $\underline{\square}$-convergence of spaces, we want to compare it to the more traditional notion of convergence of measures in a fixed metric space X. Here, we bring together some relevant facts and refer to [Fed]$_{\text{GMT}}$ and [Parth] for missing details and further information.

$3\frac{1}{2}$.8. Thick-thin decomposition and compact exhaustion of mm-spaces. We think of X as being *thin* at a point $x \in X$ on the scale ρ if the ball $B(x, \rho)$ has "small" measure, and we call the point "thick" otherwise. For example, a complete Riemannian manifold of finite volume is getting thinner and thinner as we go to infinity.



Also, the sequence of cubes

$$[0,1], [0,1]^2, [0,1]^3, \dots, [0,1]^n, \dots$$

gets thinner at each point with the growth of dimension but in a geometrically different way.

To make the concept of "thick-thin" precise, we define

$$\operatorname{Tn} X(\rho, \kappa) = \{x \in X : \mu(B(x, \rho)) < \kappa\}$$

and

$$\operatorname{Tk} X(\rho, \kappa) = X \setminus \operatorname{Tn} X = \{x \in X : \mu(B(x, \rho)) \geq \kappa\}.$$

Then for a given positive function $\kappa(\rho) > 0$, we set

$$\operatorname{Tk} X(\kappa(\rho)) = \bigcap_{0<\rho\leq 1} \operatorname{Tk} X(\rho, \kappa(\rho)),$$

and we observe that if $\mu(X) < \infty$, then the *subset* $\operatorname{Tk} X(\kappa(\rho)) \subset X$ *is compact for each function* $\kappa(\rho) > 0$. This follows trivially from the subsequence lemma of the previous section.

Next, we observe that $\mu(\operatorname{Tn} X(\rho, \kappa)) \to 0$ for each fixed $\rho > 0$ and $\kappa \to 0$, again assuming $\mu(X) < \infty$. It follows that there exists a decreasing sequence of positive functions $\kappa_i(\rho) > 0$ (converging to zero) such that the *compact subsets*

$$\operatorname{Tk} X(\kappa_1(\rho)) \subset \operatorname{Tk} X(\kappa_i(\rho)) \subset \cdots$$

exhaust all measure of X, i.e.,

$$\mu(\mathrm{Tk}\, X(\kappa_i(\rho))) \to \mu(X)$$

for $i \to \infty$. Then this trivally implies for all (possibly infinite) mm-spaces the following standard fact.

Compact exhaustion of measures. *Every mm-space* X *admits an increasing sequence of compact subsets* $X_1 \subset X_2, \subset \cdots$ *such that every* $Y \subset X$ *satisfies* $\mu(Y \cap X_i) \to \mu(Y)$ *as* $i \to \infty$.

Almost compact partitions. Every compact subset Y can be covered by δ-small parts, for each $\delta > 0$, and all measure of Y up to ε can be contained in a union of disjoint δ-small subsets. Thus, we arrive at the following

ε-Partition lemma: *For arbitrary* $\delta > 0$ *and* $\varepsilon > 0$, *there exist pairwise disjoint compact subsets* $K_1, \ldots, K_N$ *such that* $\operatorname{diam}(K_i) \leq \delta$ *and their complement* $U = X \setminus \bigcup_{j=1}^{N} K_j$ *has* $\mu(U) < \varepsilon$.

$3\frac{1}{2}$.9. Convergence of measures. A sequence of finite (i.e., of finite mass) measures μ_i is said to *converge* to μ if $\mu_i(f) \to \mu(f)$ for every bounded, positive, continuous function f on X, where $\mu(f)$ stands for $\int_X f\, d\mu$. (If we allow infinite μ_i, then we require that each point $x \in X$ admit a neighborhood $U \subset X$ such that $\mu_i(U) < \infty$ and $\mu_i|_U \to \mu|_U$. This agrees with the previous definition for finite μ_i via partitions of unity).

The metric Lid_b. Set

$$\mathrm{Lid}_b(\mu, \mu') = \sup_f |\mu(f) - \mu'(f)|$$

for $b > 0$, where f runs over all *1-Lipschitz* functions $f \colon X \to [0, b]$. Clearly, these are true metrics on the space $\mathcal{M}(X)$ of finite measures on X, and they are all mutually bi-Lipschitz equivalent. It is also not hard to see that $\mathcal{M}(X)$ is complete for Lid_b. In fact, $\mathcal{M}(X)$ is *Polish*, since measures with finite supports are (obviously) dense in $\mathcal{M}(X)$. Furthermore, *if* X *is compact, then each* $\mathcal{M}_m = \{\mu \in \mathcal{M}(X) : \mu(X) \leq m\}$ *is compact.*

Lid-**convergence**. Obviously, convergence of $\mu_i \to \mu$ implies Lid_b-convergence for each $b < \infty$, provided that $\mu(X) < \infty$. *Conversely,* Lid-*convergence for a given* $b > 0$ *implies convergence.* This follows via the compact exhaustion of μ, since every continuous function is uniformly continuous on each compact subset $Y \subset X$, and so it can be uniformly approximated on Y by a Lipschitz function.

The metric di_λ **on measures.** The distance di_λ for $\lambda \geq 0$ between finite measures μ and μ' on a metric space X is defined (à la Hausdorff) as the infimal ε for which the μ-measure of the ε-neighborhood of each $X_0 \subset X$ is at least $\mu'(X_0) - \lambda\varepsilon$ and vice-versa, i.e.

$$\mu(U_\varepsilon(X_0)) \geq \mu'(X_0) - \lambda\varepsilon \quad \text{and} \quad \mu'(U_\varepsilon(X_0)) \geq \mu(X_0) - \lambda\varepsilon$$

for all Borel subsets $X_0 \subset X$. Here again, the metrics di_λ are all mutually equivalent for $\lambda > 0$. A little thought shows that they are equivalent to Lid_b. (In fact, they could be defined in the spirit of Lid_b with the supremum taken over truncated distance functions to subsets $Y \subset X$.)

The metric di_0 is significantly different from the rest of the di_λ. For example, it obviously majorizes the Hausdorff distance between the supports of the measures. Also, di_0 majorizes Lid_b for all b. In fact,

$$|\mu(f) - \mu'(f)| \leq (\mathrm{Lip}(f))\, \mathrm{di}_0(\mu, \mu'), \tag{$*$}$$

as follows from ($\star$) below.

$3\frac{1}{2}$.10. The transportation metric Tra_λ**.** Let us think of measures as piles of stones in X and try to transport one such pile μ to the location (support) of another pile μ', where we have to replace each μ'-stone by a μ-stone of equal mass. In doing this, we insist that each stone be moved by the distance $\leq \varepsilon$, and we allow some percentage of stones to remain unmoved. We formalize this by defining a (partial) *transportation* as a measure ν on $X \times X$ whose projections (i.e., pushforwards) to X, say $\underline{\nu}$ and $\underline{\nu}'$ satisfy $\underline{\nu} \leq \mu$ and $\underline{\nu}' \leq \mu'$ (where the inequality $\mu_1 \leq \mu_2$ means that $\mu_2 = \mu_1 + \mu_3$ for some measure μ_3). Clearly $\nu(X \times X) \leq \min(\mu(X), \mu'(X))$, and the maximum of the two numbers $\mu(X) - \nu(X \times X)$ and $\mu'(X) - \nu(X \times X)$ is called the *deficiency* of ν. Finally, ν is called an *ε-transportation* if it is supported in the subset

$$Y_\varepsilon = \{(x, x') \in X \times X : \mathrm{dist}_X(x, x') \leq \varepsilon.\}$$

Now we define $\mathrm{Tra}_\lambda(\mu, \mu')$ as the infimal ε for which there exists an ε-transformation ν (from μ to μ') of deficiency $\leq \lambda\varepsilon$. (According to this definition, we are allowed to break stones into smaller pieces and to transport different pieces of the same stone to different locations.) Clearly, if $\mathrm{Tra}_\lambda(\mu, \mu') \leq \varepsilon$, then

$$\mathrm{Lid}_b(\mu, \mu') \leq \varepsilon M + 2b\lambda\varepsilon, \tag{$+$}$$

where $M \overset{\text{def}}{=} \max(\mu(X), \mu(X')) - \lambda\varepsilon$ refers to the mass $M = \nu(Y_\varepsilon)$ of the optimal ε-transformation ν. (To visualize (+), split $\mu = \mu_\varepsilon + \underline{\nu}$ and $\mu' = \mu'_\varepsilon + \underline{\nu}'$, where $\underline{\nu}(X) = \underline{\nu}'(X) = M$, and where ν transports all of $\underline{\nu}$ to $\underline{\nu}'$. This transport moves points by distances $\leq \varepsilon$, and so $|\underline{\nu}(f) - \underline{\nu}'(f)| \leq |\nu|\varepsilon$

for every 1-Lipschitz function f. On the other hand, since $|f| \leq b$, the leftovers $\mu_\varepsilon(f)$ and $\mu'_\varepsilon(f)$ are both bounded by $\lambda\varepsilon$).

It is also easy to see that Tra_λ induces the same topology as di_λ (and hence Lid_b). Namely, di_λ-convergence $\mu_i \to \mu$ implies Tra_λ-convergence. To see this, take small $K_1, \ldots, K_N$ with $\operatorname{diam} K_j \leq \delta$ and $\mu(X \setminus \bigcup_{j=1}^N K_j) \leq \kappa$ for small (and eventually going to zero) δ and κ as in the ε-partition lemma above. When μ_i approaches μ in the di_λ-metric, small disjoint neighborhoods U_j of K_j have $\mu_i(U_j)$ close to that of $\mu(K_j)$. In fact, for large i, we can make the total error $\sum_{j=1}^N |\mu(K_j) - \mu_i(U_j)|$ as small as we want, say $\leq \kappa$. Then we transport as much as we can of the μ_i-measure of each U_i to the μ-measure of K_j, thus transporting all but 3κ of the full $\mu_i(X)$, where we assume (as we may) that $|\mu_i(X) - \mu(X)| \leq \kappa$. This transportation moves points by at most $\operatorname{diam}(U_i)$, which can be taken close to δ, say $\leq 2\delta$. Thus, we bound $\mathrm{Tra}_\lambda(\mu, \mu_i)$ by $\varepsilon \leq 2\delta + 3\kappa\lambda^{-1}$.

Next, we want to obtain a specific bound on Tra_λ by di_λ, and we do this by appealing to the following

König's matching theorem. *Let μ and μ' be measures on X and X', respectively, and let $Y \subset X \times X'$ be a closed subset. Then the following condition is sufficient (and obviously necessary) for the existence of a partial transportation measure ν of a given mass with support in Y:*

Matching property: *For every Borel subset $X_0 \subset X$, the Y-match $X'_0 \subset X'$ defined as the projection of $(X_0 \times X') \cap Y$ to X' satisfies*

$$\mu'(X'_0) - \mu(X_0) \geq m - \mu(X).$$

Remarks: (a) The matching property looks asymmetric for $X \leftrightarrow X'$, but a little thought reveals the actual symmetry.

(b) The classical König theorem refers to finite spaces (see [Lova-Plum] and [Bol-Var]), and the general case then can be derived by an obvious approximation of μ and μ' by measures with finite supports. (We do not need the finer aspect of the classical König theorem which delivers an *integer-valued* tranportation ν for integer valued μ and μ. In fact, the most interesting case refers to measures consisting of atoms of unit mass.)

Corollary: *The metrics* Tra_λ *and* di_λ *are equal for all* $\lambda \geq 0$,

$$\mathrm{Tra}_\lambda(\mu, \mu') = \mathrm{di}_\lambda(\mu, \mu') \qquad (\star)$$

for all measures μ and μ'.

Proof. The inequality $\mathrm{di}_\lambda(\mu,\mu') \le \varepsilon$ is equivalent to the Y_ε-matching condition for transportations from μ to μ' of mass $m = \mu(X) - \lambda\varepsilon$ and from μ' to μ of mass $m' = \mu(X') - \lambda\varepsilon$.

Applications to Lid_b **and** $\underline{\square}_\lambda$**.** (*a*) Bringing together (∗) with (+) above gives an effective bound on Lid_b by di_λ. Namely,

$$\mathrm{di}_\lambda(\mu,\mu') \le \varepsilon \Rightarrow \mathrm{Lid}_b(\mu,\mu') \le \varepsilon M + 2b\lambda\varepsilon$$

for $M = \max(\mu(X),\mu'(X)) - \lambda\varepsilon$. For example, if $\mu(X) \le \mu(X') \le 1$, then

$$\mathrm{Lid}_1 \le 3\,\mathrm{Tra}_1 = 3\,\mathrm{di}_1\,.$$

(*b*) *Convergence* $\mu_i \to \mu$ *implies the* $\underline{\square}_\lambda$*-convergence of the mm-spaces,*

$$(X,\mathrm{dist},\mu_i) \to (X,\mathrm{dist},\mu).$$

In fact, if $\mathrm{Tra}_1(\mu,\mu') \le \varepsilon$, then we have a ν on $Y_\varepsilon \subset X \times X$ which projects to measures $\underline{\nu} \le \mu$ and $\underline{\nu}' \le \mu'$ which ε-approximate μ and μ'. We add the error $\kappa = (\mu - \underline{\nu}) \times (\mu' - \underline{\nu}')$ to ν and even the resulting $\nu_+ = \nu + \kappa$ parametrizes μ and μ', i.e., one has $\underline{\nu}_+ = \mu$ and $\underline{\nu}'_+ = \mu'$. Then we parametrize ν_+ by $[0,m]$ for $m = \nu_+(X \times X)$ and observe that the pullbacks of d and d' of $\mathrm{dist}(X,X')$ satisfy $\underline{\square}_1(d,d') \le 2\varepsilon$.

(*c*) **Exercise:** Consider finite measures μ and μ' with $\mu'(X) = \mu(X)$, and define the *Kantorovich–Rubenstein* distance (compare [Kantor], [Gan–McCann], [Levin])

$$\mathrm{Tran}(\mu,\mu') = \inf_\nu \int \mathrm{dist}(x,x')\,d\nu,$$

where ν runs over all transportations (measures) ν on $X \times X$ of zero deficiency. Show that

$$\mathrm{Tran}(\mu,\mu') = \mathrm{Lid}_\infty(\mu,\mu').$$

Observe that these two are true *everywhere finite* metrics iff $\mathrm{diam}(X) < \infty$.

$3\frac{1}{2}$.11. Induced Minkowski measures on subsets $Y \subset X$**.** A subset Y in an mm space is called *Minkowski regular* if the normalized measures μ_ε on the ε-neighborhoods of Y converge as $\varepsilon \to 0$, where μ_ε denotes the measure $(\mu|_{U_\varepsilon})/\mu(U_\varepsilon)$ and $U_\varepsilon = U_\varepsilon(Y)$. This limit is a probability measure supported on Y, called the *induced Minkowski measure.* This is of particular interest where $\mu(Y) = 0$, and so naively restricting μ to Y gives us the zero measure. Also notice that one can define *an* induced measure on Y by taking the limit of a *subsequence* μ_{ε_i} in the case where Y is *not* Minkowski regular. Such sublimits exist in many cases as the families of measures μ_ε, $\varepsilon \to 0$, are typically di_1-precompact.

Examples and exercises. (*a*) Show that every compact smooth submanifold Y in a Riemannian manifold is Minkowski regular and that the induced Minkowski measure (from the Riemannian measure on X) equals the normalized Riemannian measure on Y.

(*b*) Show that every compact semialgebraic subset $Y \subset X = \mathbb{R}^n$ is Minkowski regular.

D. Basic geometry of the space $\mathcal{X}$

We stick to $\lambda = 1$ (since all $\underline{\square}_\lambda$-metrics for $\lambda > 0$ are equivalent) and denote by $\mathcal{X}$ the space of isomorphism classes of mm-spaces of *finite mass* with the metric $\underline{\square}_1$. We agree that $\mathcal{X}$ contains a unique element of mass zero denoted by $\mathbf{0}$.

$3\frac{1}{2}$.12. Union lemma: *Let $X_i \in \mathcal{X}$ satisfy $\underline{\square}_1(X_{i+1}, X_i) \le 2^{-i}$. Then there exists a Polish space $\overline{X}$ and a sequence of measures $\overline{\mu}_i$ on $\overline{X}$ such that the $\overline{\mu}_i$ converge to some $\overline{\mu}$ on $\overline{X}$, and where each $(X_i, \mathrm{dist}_i, \mu_i)$ is isomorphic to $(\overline{X}, \overline{\mathrm{dist}}, \overline{\mu}_i)$.*

Proof. Consider parametrizations $p_i : [0, m_i] \to X_i$ and $p_i' : [0, m_i' = m_{i+1}] \to X_{i+1}$ responsible for the inequality $\underline{\square}_1(X_{i+1}, X_i) \le 2^{-i}$, and let $T_i \subset [0, m_i] \cap [0, m_i']$ be the subset where the pullbacks d_i and d_i' of the metrics dist_i and dist_{i+1} are 2^{-i}-close. Say that points $x_i \in X_i$ and $x_{i+1}' \in X_{i+1}$ are $\underline{\square}$-neighbors if $\exists t \in T$ such that $p_i(t) = x$ and $p_i'(t) = x'$. Then we consider the disjoint union $X = X_1 \sqcup X_2 \sqcup \cdots$ and take the supremal metric $\overline{\mathrm{dist}}$ on X such that $\mathrm{dist}\,|_{X_i} = X_i$ and $\mathrm{dist}(x_i, x_{i+1}') \le 2^{-i+1}$ for all pairs of neighbors $x_i \in X_i$, $x_{i+1}' \in X_{i+1}$, and all $i = 1, 2, \ldots$. It is obvious that the Tra_1-distance between the measures $\overline{\mu}_i$ and $\overline{\mu}_{i+1}$ on X corresponding to μ_i and μ_{i+1} under the embeddings $X_i, X_{i+1} \subset X$ does not exceed 2^{-i+2}. Thus the $\overline{\mu}_i$ form a Cauchy sequence which converges to some $\overline{\mu}$ on the completion $\overline{X}$ of X.

Corollary: *The space $(\mathcal{X}, \underline{\square}_1)$ is complete and is therefore a Polish metric space.*

$3\frac{1}{2}$.13. A remark on the normalization of measures and subspaces $\mathcal{X}_m$ and $\mathcal{X}_{\le m}$ in $\mathcal{X}$. The group $\mathbb{R}_+^*$ naturally acts on $\mathcal{X}$ by $X = (X, \mathrm{dist}, \mu) \mapsto (X, \mathrm{dist}, r\mu)$ for $r \in \mathbb{R}_+^*$, and the function $X \mapsto \mu(X)$ on $\mathcal{X}$

is homogeneous with respect to this action. Putting this with $r = m^{-1}$ for $m = \mu(X)$ is called the *normalization of X and/or of μ_X*, since the mass of X becomes one in the process. The mm-spaces X with $\mu(X) = 1$ are called *normalized* or *probability spaces*, in which case the measure $\mu\{x \in X_0 \subset X\}$ is interpreted as the probability of the event $x \in X$.

The space $\mathcal{X}$ is naturally sliced into the subspaces $\mathcal{X}_m$, $m \in \mathbb{R}_+$, consisting of X with $\mu(X) = m$. Every such X can be parametrized by the segment $[0, m]$, and the metric $\underline{\square}_1$ on $\mathcal{X}_m$ descends from $\square_1$ on functions $d\colon [0, m]^2 \to \mathbb{R}_+$. The relevant functions d form a convex subset D and $\square_1$ is, obviously, a path metric on D. It follows that $\underline{\square}_1$ on $\mathcal{X}$ is also a path metric. Then it is clear from the definition of $\underline{\square}_1$ on $\mathcal{X}$ that it is a path metric on $\mathcal{X}$.

When dealing with convergence and (pre)compactness in $\mathcal{X}$, it is convenient to restrict to the subspace $\mathcal{X}_{\leq m} = \{X \subset \mathcal{X} : \mu(X) \leq m\}$ (see below). This $\mathcal{X}_{\leq m}$ is, like $\mathcal{X}_m$, *geodesically convex* (i.e. *path* metric) in $\mathcal{X}$. We note in passing that the function $X \mapsto \mu(X)$ is $\underline{\square}_1$-continuous on $\mathcal{X}$, and so $\mathcal{X}_m$ and $\mathcal{X}_{\leq m}$ are closed as well as geodesically convex subspaces in $\mathcal{X}$.

Everything we do with mm-spaces with finite measure can be reduced to the normalized case where the statements are somewhat easier. However, we allow all $m = \mu(X)$ since some natural mm-spaces are not normalized. For example, the mass of the unit ball $B(1) \subset \mathbb{R}^n$ is far from one. In fact, it is about $n^{-n/2}$, while $\mu(B(R)) = 1$ is achieved for $R \approx \sqrt{n}$.

$3\frac{1}{2}$.14. Complexity of mm spaces and (pre-)compactness criteria for subsets $\mathcal{Y} \subset \mathcal{X}$ (compare Ch. 5A). The basic measure of the overall complexity of a finite mm space X, besides its mass $m = \mu(X)$, is reflected in the pair of numbers N and D depending on given $\delta, \varepsilon > 0$, such that X can be covered by N subsets K_i, $i = 1, \dots, N$, of diameter $\leq \delta$ and such that $\mu(X \setminus \bigcup_{i=1}^N K_i) \leq \varepsilon$ and $\operatorname{diam}(\bigcup_{i=1}^N K_i) \leq D$. We express the existence of such an up-to-ε covering by writing

$$\operatorname{co}(X, \delta, \varepsilon) \leq (N, D).$$

Given an up-to-ε covering, we can obviously replace K_i by mutually disjoint compact subsets $K_i^o \subset K_i$ with $\bigcup_{i=1}^N K_i^o \geq \mu(\bigcup_{i=1}^N K_i - \varepsilon_0)$ for a given $\varepsilon_0 > 0$, say for $\varepsilon_0 = \varepsilon$. This allows an approximation of the underlying measure μ by measures with finite supports. In fact, we can replace each K_i^o by a single atom situated in K_i^o of mass $m_i = \mu(K_i^o)$. Thus, we get μ_N with N atoms such that $\operatorname{Tra}_1(\mu, \mu_N) \leq \varepsilon' = \max(\delta, 2\varepsilon)$. In particular, X admits $2\varepsilon'$ approximation by an N-point space in the metric $\underline{\square}_1$. Conversely, if X admits an ε-approximation by some X_N consisting of N atoms and having

$\operatorname{diam}(X_N) = D_N$, then

$$\operatorname{co}(X, 3\varepsilon', 3\varepsilon') \leq (N, D + 3\varepsilon)$$

as an obvious argument shows.

Finally, the complexity of X can be neatly expressed by looking at the thick-thin decomposition of X (see $3\frac{1}{2}.8$). Namely, the pair of numbers $D\,\mathrm{Tk} = \operatorname{diam}\mathrm{Tk}\,X(\rho,\kappa)$ and $\mu\,\mathrm{Tn} = \mu(\mathrm{Tn}\,X(\rho,\kappa))$ carry the same essential information as $\operatorname{co}(X;\delta,\varepsilon)$. In fact, $\mathrm{Tk}\,X(\rho,x)$ can be covered by at most $N \leq \mu(X)/\kappa$ balls of radii 2ρ, which allows a bound on co in terms of $D\,\mathrm{Tk}$ and $\mu\,\mathrm{Tn}$. Conversely, given the K_i^o above, we take those which have $\mu(K_i^o) \geq \kappa_0$ for $\kappa_0 = N^{-2}$ and denote by Tk_0 their union. It is clear that $\mu(\mathrm{Tk}_0) \geq \mu\left(\bigcup_{i=1}^N K_i^o\right) - N^{-1}$ and that $\mathrm{Tk}_0 \subset \mathrm{Tk}\,X(\rho,\kappa)$ for $\rho = \delta$ and $\kappa = N^{-2}$. Thus, we control $D\,\mathrm{Tk}$ and $\mu\,\mathrm{Tn}$ by the covering invariants of X.

Proposition: *Let $\Delta = \Delta(\varepsilon)$ be an arbitrary positive function defined for $\varepsilon > 0$ and consider the following three subsets in $\mathcal{X}$.*

$\mathcal{Y}_1 = \mathcal{Y}_1(\Delta)$ *consisting of those $X \in \mathcal{X}$ such that* $\operatorname{co}(X,\varepsilon,\varepsilon) \leq (\Delta(\varepsilon),\Delta(\varepsilon))$ *for all $\varepsilon > 0$;*

$\mathcal{Y}_2 = \mathcal{Y}_2(\Delta)$ *consisting of those $X \in \mathcal{X}$ which can be ε-approximated for all $\varepsilon > 0$ by spaces X_ε with* $\operatorname{diam}(X_\varepsilon) \leq \Delta = \Delta(\varepsilon)$ *and* $\operatorname{card}(X_\varepsilon) \leq \Delta = \Delta(\varepsilon)$*;*

$\mathcal{Y}_3 = \mathcal{Y}_3(\Delta)$ *consisting of those X where for every ε, there exists $\kappa \in [\varepsilon, \Delta^{-1}]$ such that* $\operatorname{diam}\mathrm{Tk}(\varepsilon,\kappa) \leq \Delta = \Delta(\varepsilon)$ *and* $\mu(\mathrm{Tn}(\varepsilon,\kappa)) \leq \varepsilon$.

Then the intersection $\mathcal{Y}_i \cap \mathcal{X}_{\leq m}$ is precompact in $\mathcal{X}$ for each $i = 1,2,3$ and $m \geq 0$ and the union $\bigcup_\Delta \mathcal{Y}_i(\Delta)$ over all positive functions $\Delta = \Delta(\varepsilon)$ equals all of $\mathcal{X}$, again for each $i = 1,2,3$.

Proof. The space of N-point spaces with mass $\leq m$ and diameter $\leq D$ is, obviously, precompact which implies precompactness of $\mathcal{Y}_2 \cap \mathcal{X}_{\leq m}$. Then the discussion above shows that $\mathcal{Y}_1(\Delta) \subset \mathcal{Y}_2(\Delta')$ for some $\Delta' = \Delta'(\Delta)$ and $\mathcal{Y}_3(\Delta) \cap \mathcal{X}_{\leq m} \subset \mathcal{Y}_2(\Delta'')$ for $\Delta'' = \Delta''(\Delta, m)$. This yields the precompactness of $\mathcal{Y}_i \cap \mathcal{X}_{\leq m}$ for $i = 1,3$. And the equality $\bigcup_\Delta \mathcal{Y}_i(\Delta) = \mathcal{X}$ is obvious for $i = 3$ and hence for $i = 1,2$ as well.

Remark: It is easy to see that the subsets $\mathcal{Y}_i(\Delta)$ are $\underline{\square}_1$-closed in $\mathcal{X}$ for $i = 1,2,3$ and hence the intersection $\mathcal{Y}_i(\Delta) \cap \mathcal{X}_{\leq m}$ are *compact.*

Corollary: Convergence criterion. *The following conditions are necessary and sufficient for $\underline{\square}_1$-convergence of a sequence $X_i \in \mathcal{X}$ to some $X \in \mathcal{X}$.*

I. *The measures* $\underline{\mu}_r^{X_i}$ *(see $3\frac{1}{2}.4$) converge as $i \to \infty$ to some μ_r on M_r for each $r = 1, 2, \ldots$.*

II. *For arbitrary $\rho > 0$ and $\kappa > 0$,*

$$\operatorname{diam}(\operatorname{Tk} X_i(\rho, \kappa)) \leq \text{const} < \infty$$

and there exists a positive function $\rho(\kappa) > 0$ for $\kappa > 0$ such that $\rho(\kappa) \to 0$ for $\kappa > 0$ and

$$\mu(\operatorname{Tn} X_i(\rho(\kappa), \kappa)) \leq \varepsilon(\kappa)$$

for some function $\varepsilon(\kappa) \to 0$ as $\kappa \to 0$ independent of i.

Furthermore, we can replace II *by*

III. *There exists a function $\Delta(\varepsilon)$, $\varepsilon > 0$, such that each X_i admits $N \leq \Delta(\varepsilon)$ subsets $K_{ij} \subset X_i$, $j = 1, \ldots, N$ satisfying*

1. $\operatorname{diam}(K_{ij}) \leq \varepsilon$,
2. $\operatorname{diam} \bigcup_{j=1}^{n} K_{ij} \leq \Delta(\varepsilon)$,
3. $\mu(X_i \setminus \bigcup_{j=1}^{N} K_{ij}) \leq \varepsilon$.

Proof. Properties II and III allow convergent subsequences, and I implies that these have isomorphic limits, according to the reconstruction theorem (see $3\frac{1}{2}.5$).

Remark: The condition I alone does not suffice. For example, the sequence of spaces X_i consisting of i-atoms of mass i^{-1} with unit mutual distances does *not* converge, although the $\underline{\mu}_r^{X_i}$ converge. A more interesting example is provided by round spheres S^i with normalized measures. These violate III, but nevertheless the $\underline{\mu}_r^{S^i}$ converge for each r (see $3\frac{1}{2}.18$). In general, we may have a sequence X_i with $\operatorname{diam}(X_i) \leq D < \infty$ having no convergent subsequence. But there always exists a subsequence X^{i_j} for which the $\underline{\mu}_r^{X_{i_j}}$ converge for every r. (Such subsequences are similar to asymptotic sequences we will introduce in $3\frac{1}{2}$.I but the latter seems to be a better candidate for "ideal mm-spaces".)

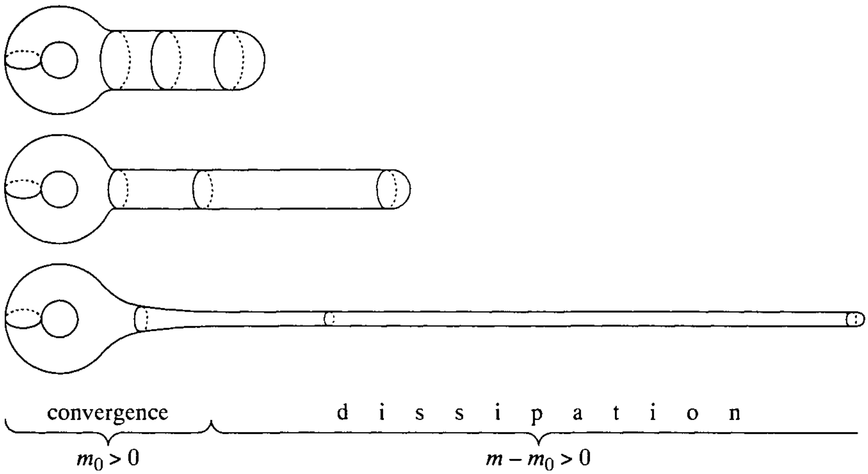
Examples: (a) If $\{X_i\}_{i \in I}$ is a family of compact mm-spaces which is precompact in the Hausdorff topology (where we forget about the measures), then it is also $\underline{\square}_1$-precompact, provided that $\mu_i(X_i) \leq m < \infty$. This is essentially obvious (as it follows, for example, from the metric version of the union lemma) but is particularly useful and can be applied in particular to compact manifolds with Ricci $\geq -$const and diam $\leq D$ (see Ch. 5 and [Che- Col]$_{\mathrm{I--III}}$) and to more general spaces with the doubling property (see Appendix B_+). But nothing essentially new is brought in here along with the measure.

(*b*) A more interesting picture arises if we do not bound the diameter of our manifolds X but somehow prevent collapse and the subsequent dissipation of measure (as is present, for instance, in the product $X = (\varepsilon Y)\times(\varepsilon^{-1}Y)$ for $\varepsilon \to 0$) by imposing suitable geometric and topological conditions on X. For example, it follows trivially from Thurston's theory (see [Gro]$_{\mathrm{HMTJ}}$) that the complete connected Riemannian manifolds X of dimension ≥ 3 with vol$(X) \leq m$ and pinched negative sectional curvatures, $0 \geq -\kappa_1 \geq K(X) \geq \kappa_2 > \infty$ form a $\underline{\Box}_1$-precompact family. (This is nearly true for $n = 2$, where the only catch is a possible nonconnectedness of the thick part of X, which makes the diameter of Tk(X) go to infinity.)

Furthermore, much of the pattern typical for the pinched negative curvature remains valid under the bound Ricci $\geq -$const if we add suitable *topological* assumptions on X. For example, if we assume that each X in question admits an auxiliary complete Riemannian metric with pinched negative curvature and finite volume, we thus impose a nontrivial lower bound on the (volume of) the thick part of (the original metric with Ricci $\geq -$const on) X (see [Gro]$_{\mathrm{FRM}}$, [Gro]$_{\mathrm{VBC}}$, and also [Che-Col], [Col]). This allows a subsequence X_i taken out of the totality of our X's which $\underline{\Box}_1$-converges on substantial parts of X_i, i.e., on some subsets $Y_i \subset X_i$ with $\mu_i(X_i) \geq m_0 > 0$. But some amount of measure may dissipate (in the sense of $3\frac{1}{2}$.J) on some thin, infinitely growing fingers with Ricci ≥ 0.

However, the full beauty and power of the measure becomes visible only when we look at families of manifolds of *growing* dimension, such as the unit spheres S^n, $n = 1, 2, \ldots$, with normalized measures, which $\underline{\Box}_1$-diverge but which nevertheless converge (concentrate) for some weak topology on $\mathcal{X}$ (see $3\frac{1}{2}$.45).

$3\frac{1}{2}$.15. Lipschitz order on mm-spaces. Say that X *(Lipschitz) dominates* Y and write $X \succ Y$ or $X \overset{Lip_1}{\succ} Y$ if there exists a 1-Lipschitz map $X \to Y$ pushing forward the measure μ_X to a measure ν on Y *proportional* to μ_Y, i.e., $\nu = c\mu_Y$ for some positive constant $c \geq 1$.



Here are the basic properties of this order:

(*a*) *If two spaces X and Y of finite mass satisfy $X \succ Y \succ X$, then they are isomorphic.*

(*b*) *The Lipschitz order is $\underline{\square}_1$-continuous, i.e., if $X_i \xrightarrow{\underline{\square}_1} X$ and $Y_i \xrightarrow{\underline{\square}_1} Y$, then $X_i \succ Y_i \Rightarrow X \succ Y$.*

(*c*) *Every bounded subset $\mathcal{X}_0 \subset \mathcal{X}$ is precompact*, where "bounded" means that there exists $X_0 \in \mathcal{X}$ such that $X \prec X_0$ for all $X \in \mathcal{X}_0$.

(*d*) *Every decreasing sequence $X_1 \succ X_2 \succ \cdots \succ X_i \succ \cdots$ in $\mathcal{X}$ converges.*

(*e*) *Every bounded increasing sequence in $\mathcal{X}$ converges.*

(*f*) *Every precompact subset in $\mathcal{X}$ is bounded.*

Proof. To prove (*a*) we introduce the *averaged diameter of an mm space.* This $AvDi(X)$ is defined as

$$m^{-2} \int_{X \times X} \operatorname{dist}(x, x')\, d\mu d\mu$$

for $m = \mu(X)$. Clearly $AvDi$ is monotone for the Lipschitz order, and if $X \succ Y$ while $AvDi(X) = AvDi(Y) < \infty$, then the implied 1-Lipschitz map $X \to Y$ is an isomorphism. This implies (*a*) in the case where $AvDi < \infty$. In general, we replace $AvDi$ by the average of $\varphi(\operatorname{dist}(x, x'))$ for some *bounded*, strictly monotone function $\varphi(d)$. This average is necessarily finite, and the reasoning above yields the proof.

Proof of (b). Start with the following definition. A map from an mm space to a metric space, say $f: X \to Y$ is called *λ-Lipschitz up to ε* if

$$\mathrm{dist}_Y(f(x), f(x')) \leq \lambda\, \mathrm{dist}(x, x') + \varepsilon$$

for all x, x' in a subset $X_0 \subset X$ with $\mu(X \setminus X_0) \leq \varepsilon$.

(3_b) me_1-convergence lemma: *Let X be an mm-space with finite measure, let Y be a Polish space, and let $f_i: X \to Y$ be λ-Lipschitz up to ε_i Borel maps with a fixed λ and $\varepsilon_i \to 0$ for $i \to \infty$. Then the f_i admit an me_1-convergent subsequence, provided that the pushforward measures $(f_i)_*(\mu_X)$ converge to some measure μ on Y.*

Proof. Using the Subsequence Lemma (see $3\frac{1}{2}$.6), one achieves convergence on a countable dense set $A \subset X$. The resulting map $f_A: A \to Y$ is obviously λ-Lipschitz and thus extends by continuity to the desired map $f: X \to Y$.

($3'_\mathrm{b}$) $\square_1$-convergence lemma: *Let $X_i \xrightarrow{\square_1} X$. Then there exist Borel 1-Lipschitz up to ε_i maps $p_i: X \to X_i$ and $q_i: X_i \to X$ such that ε_i as well as $\mathrm{me}_1(p \circ q, \mathrm{id})$ and $\mathrm{me}_1(q_i \circ p_i, \mathrm{id})$ converge to zero for $i \to \infty$, and the pushforward measures $(p_i)_*(\mu_{X_i})$ converge to μ_X.*

Proof. We map each point in X_i to some of its neighbors in X and vice-versa, as we did in the proof of the union lemma (see (1) above). The resulting p_i and q_i are clearly what we want.

Now, for convergent $X_i \to X$ and $Y_i \to Y$, we take $p_i: X \to X_i$ and $q_i: Y_i \to Y$ and then observe that the maps $q_i \circ p_i: X \to Y$ satisfy the assumption of (3_b) with $\lambda = 1$ and $\mu = \mu_Y$.

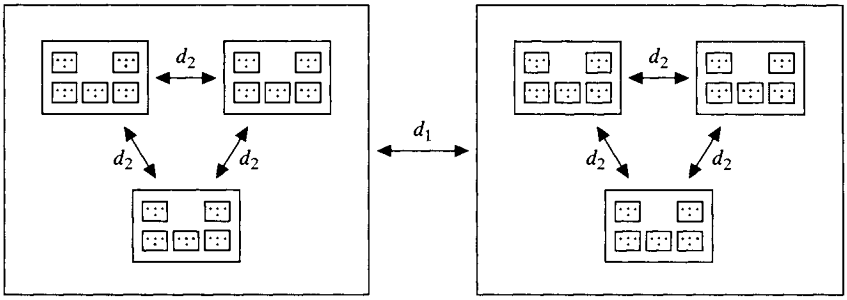
Proof of (c). This follows from the

Monotonicity of basic invariants. Most natural geometric invariants of mm-spaces are monotone for the Lipschitz order. This applies in particular to the covering number $\mathrm{Co}(X, \rho, \varepsilon)$, which is sufficient for (c) via (**2**).

Proof of (d) **and** (e). This is immediate with the above convergence criterion by the following standard (and obvious) lemma applied to the measures $\mu_i = \underline{\mu}_r^{X_i}$ on $M_r \subset \mathbb{R}_+^{r(r-1)/2}$.

Monotone convergence lemma: *Let μ_i be a precompact sequence of finite measures on $\mathbb{R}_+^k$ such that, for every box $B = [0, t_1] \times [0, t_2] \times \cdots [0, t_k]$ in $\mathbb{R}^k$, the numerical sequence $\mu_i(B)$ is increasing. Then, the μ_i converge. Similarly, the μ_i converge if $\mu_i(B)$ are decreasing for all boxes B.*

Proof of (f). Let us describe a class of spaces which bound all precompact subsets in $\mathcal{X}$. First, given a (growing) sequence of positive integers $N_1, N_2, \ldots$ and of reals $d_1, d_2, \ldots$ (slowly) converging to zero, we construct a Cantor set $K = K(N_i, d_i)$ which consists of N_1 equal parts with distances d_1 between the points in different parts; then each part is divided into N_2 pieces with mutual distances d_2, and so on.

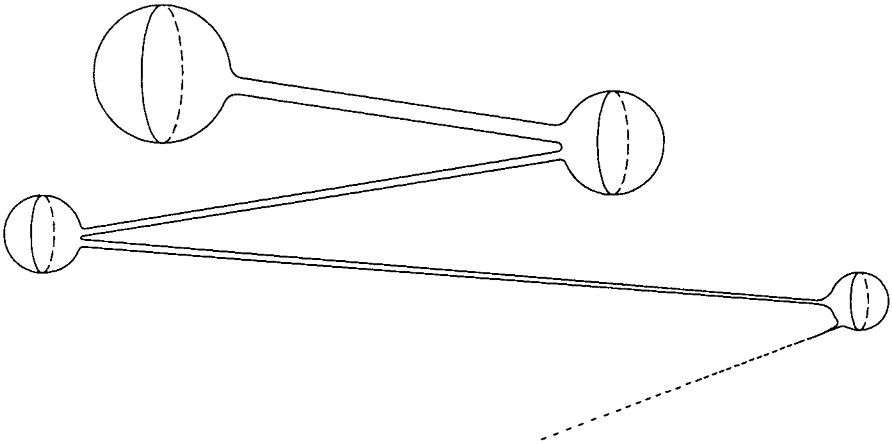


Such K is given a natural measure, where each of the initial parts have masses N_1^{-1}, the parts at the second stage have masses $(N_1N_2)^{-1}$, and so on.

Then we take disjoint unions of K's, say $K_1 \sqcup K_2 \sqcup \cdots$, where the measure in K_i is normalized to have mass 2^{-i} and where the distances between different K_i are made large. If $K_j(N_{ij}, d_{ij})$ have fast-growing N_{ij} and slowly decaying d_{ij}, then the resulting union is good enough to majorize lots of spaces X. This implies (f) when all quantifiers are put into place (which we leave for the reader to do).

$3\frac{1}{2}$.16. Lipschitz order in the Riemannian category. It is obvious that the relation $X \prec Y$ implies $\dim X \leq \dim Y$ for (possibly noncompact) Riemannian manifolds of finite mass. In fact, this is also true and obvious for $m(X) = m(Y) = \infty$ if Y is *thick*, i.e., if each ρ-ball in Y satisfies $\mu(B(y, \rho)) \geq \kappa(\rho) > 0$ for all $\rho > 0$. Yet, there exists a complete Riemannian manifold Y diffeomorphic to $\mathbb{R}^2$ which dominates $\mathbb{R}^n$ for all n. One can take Y, for example, obtained by joining a countable union of spheres by very thin, long tubes.

In fact, one can eventually dominate every connected X of infinite mass using some Y of this kind. What is more interesting, however, is the possibility of dominating X by λY where X and Y are arbitrary compact Riemannian manifolds, possibly with (smooth!) boundaries, which satisfy $\dim Y \geq \dim X$ and where $\lambda Y = (Y, \lambda\,\mathrm{dist}, \lambda^n \mu)$ for the Riemannian distance and measure. So, what we claim is that *there exists* $\lambda = \lambda(X, Y)$ *such that* $X \prec \lambda Y$ *provided that* $\dim X \leq \dim Y$ *and* X *is connected.*



Sketch of the proof. If $\dim X = n$ and Y is diffeomorphic to the n-ball, we cut X^n along some piecewise smooth hypersurface and obtain an n-ball B' with piecewise smooth boundary. For example, S^2 should be cut along a segment, the 2-torus over a figure eight inside it, etc. Then it is not hard to make a (bi)-Lipschitz homeomorphism $Y \to B'$ sending the measure of Y to a multiple of the X-measure on B'.

Next, if Y is an m-ball with $m \geq n$, then we think of Y as $B^n \times B^{m-n}$, where there is no problem of finding our maps $B^n \times B^m \to B^n$, and then we go from $Y \simeq B^n \times B^m$ to every compact $X = X^n$ via $B^n \to B' \subset X^n$.

Finally, every Y of dimension m can be 1-Lipschitz mapped onto a Euclidean simplex Δ^m preserving the measure. This is done by triangulating Y into m-simplices of equal volume and then "folding" across the $(m-1)$-dimensional faces. A typical map of this kind goes from the sphere $S^m = \{x_0^2 + x_1^2 + \cdots + x_m^2 = 1\}$ to Δ^m by the map $(x_0, \ldots, x_m) \mapsto (x_0^2, x_1^2, \ldots, x_m^2)$. We stop at this point and invite the reader to furnish the details. We also suggest that the reader study the possibility of having the implied map $Y \to X$ *smooth* when $\dim Y \geq 2 \dim X$ (as the "curved" Archimedes map $S^2(\pi) \to [-\pi, \pi] \to S^1(1)$).

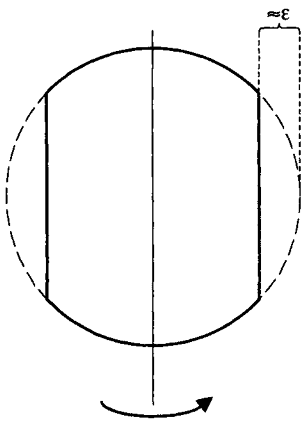
Open questions. It seems hard to evaluate the *domination constant* $\lambda = \lambda(X, Y)$ above with *reasonable precision* for specific manifolds X and Y. For example, let X and Y be the products of round spheres of given radii, $X = S^{i_1}(r_1) \times S^{i_2}(r_2) \times \cdots S^{i_k}(r_k)$ and $Y = S^{j_1}(R_1) \times S^{j_2}(R_2) \times \cdots S^{j_l}(R_l)$. What is, roughly, $\lambda(X, Y)$? Or, suppose that X is a convex hypersurface in $\mathbb{R}^{n+1}$ with principal curvatures ≥ 1. What is the minimal radius of the round sphere in $\mathbb{R}^{n+1}$ dominating this X? One knows in this regard that there is a 1-Lipschitz diffeomorphism f of the unit sphere onto X, but it is unclear if one can make such an f with *constant* Jacobian. A similar question arises when X has $\mathrm{Ricci}\, X \geq n-1$ and we seek the optimal domination of X by the round n-sphere or an n-ball.

A closely related question is as follows. Suppose we have two equidimensional manifolds X and Y which admit a λ-bi-Lipschitz diffeomorphism f between them, which essentially means that $d_L(X,Y) \leq \log(\lambda)$, where d_L is the Lipschitz distance of 3.1. Then we introduce another distance, say $d_L^{\mu}(X,Y)$ corresponding to the best λ' bi-Lipschitz homeomorphism f' between X and Y with *constant* Jacobian. The question is by how much λ' is greater than λ an *generically* (if we only knew what it means!), we expect that the discrepancy between λ and λ' grows exponentially fast with dimension n, since the jacobian of a map f can grow as fast at $(\mathrm{Lip}(f))^n$. Here are specific examples where we want to evaluate this d_L^{μ}.

(a) Consider smooth hypersurfaces X in $\mathbb{R}^{n+1}$ given by the radial functions f on the unit sphere $S^n \subset \mathbb{R}^{n+1}$, i.e.,

$$X = X_f = \{(s,r) : s \in S^n, r = 1 + f(s)\}$$

in the polar coordinates (s,r) on $\mathbb{R}^{n+1}$. Take all those $X = X_f$ where the C^1-norm of f is $\leq \varepsilon$, i.e., $|f(s)| \leq \varepsilon$ and $\|\operatorname{grad} f(x)\| \leq \varepsilon$. We use here some fixed small $\varepsilon > 0$, say $\varepsilon = 0.1$ and let $n \to \infty$. It is not hard to show that the d_L-diameter of the space of these X remains bounded (of order ε), but the d_L^{μ}-diameter grows at least as fast as $\log(\varepsilon\sqrt{n})$. A particular hypersurface X which is about $(\log\sqrt{n\varepsilon})$-far from S^n is obtained as the $(n-1)$-revolution of the curve below (truncated circle).



This X is distinguished from S^k by its *observable diameter* (see $3\frac{1}{2}$.20) which equals $\simeq 1/\sqrt{n}$ for S^n (see $3\frac{1}{2}$.20) and is $\simeq \sqrt{\varepsilon}$ for our X (by a trivial argument. It is also not hard to bound $d_L^\mu(X, S^n) \geq \log\sqrt{n\varepsilon})$. And of course there is much more to this space of surfaces X (defined by $\mathrm{dist}_{C^1}(X, S^n) \leq \varepsilon$) than our rough(?) lower bound on its diameter.

(*b*) Consider the convex surfaces X in $\mathbb{R}^{n+1}$ with a given pinching of the principal curvatures, say with $1 \leq \mathrm{curv}\, X \leq 2$. Here again the d_L-diameter remains bounded as $n \to \infty$, but it is unclear what happens to the d_L^μ-diameter. Also, a similar problem remains very much open for Riemannian manifolds with (sufficiently) pinched positive sectional curvatures.

(*c*) Consider all smooth complex hypersurfaces X of a given degree d in $\mathbb{C}\mathrm{P}^n$. We want to understand the following metrics on the space of these X:

(*i*) the metric d_L,

(*ii*) the metric d_L^μ, where one should note that all complex hypersurfaces of a given degree (and dimension) have equal volumes,

(*iii*) the *symplectic Lipschitz metric* d_L^ω, where we appeal to the *symplectic* λ-Lipschitz maps between our hypersurfaces for the symplectic structure induced by the standard 2-form ω on $\mathbb{C}\mathrm{P}^n$.

Clearly $d_l \leq d_L^\mu \leq d_L^\omega$, and one may think that d_L^ω is more amenable to computation (at least C^1-near a fixed X_0). But we have only a very vague idea of the global geometry of the space of hyperspaces with any of these three metrics.

Finally, we observe that the Lipschitz order can be directly used to define yet another metric, say $d_L^\diamond$, which refers to the minimal λ (or rather $\log(\lambda)$ for which $X \succ \lambda Y$ and $Y \succ \lambda X$. (We suggest that the reader check that this is indeed a metric on the set of isometry classes of compact, equidimensional Riemannian manifolds.) It is interesting to compare $d_L^\diamond$ with d_L^μ, on the one hand, and with $\underline{\Box}_1$ on the other for specific classes of Riemannian manifolds (e.g., for those with some bound on curvature, injectivity radius, diameter, etc.). And once again, we wish to gain some insight into the geometry of simple(?) spaces (such as in (*a*), (*b*), (*c*) above) with the metric $d_L^\diamond$ (along with d_L^μ and $\underline{\Box}_1$, while d_L and d_H look interesting only for (*c*)).

$3\frac{1}{2}$.17. Approximation of mm-spaces by measures in $(\mathbb{R}^N, \|\cdot\|_{\ell_\infty})$. Recall that the ℓ_∞-norm in $\mathbb{R}^N$ is given by $\|b\|_{\ell_\infty} = \sup|b_i|$, $i = 1, \ldots, N$. Thus, every N-tuple of 1-Lipschitz functions $\varphi_1, \ldots, \varphi_N$ on X defines a 1-Lipschitz map $\Phi : X \to (\mathbb{R}^N, \|\cdot\|_{\ell_\infty})$. Now, let X be an mm-space with finite mass, and let $\varphi_1, \varphi_2, \ldots, \varphi_i, \ldots$ be a sequence of 1-Lipschitz functions on X which is me_1-dense in the space of all 1-Lipschitz functions

$X \to \mathbb{R}$. Such a sequence always exists, since the space of functions $X \to \mathbb{R}$ is Polish with respect to the metric me_1 (see $3\frac{1}{2}$.A). We then take $\Phi_N = (\varphi_1, \varphi_2, \dots, \varphi_N) : X \to \mathbb{R}^N$ and let $\underline{X}_N = (\mathbb{R}^N, \underline{\mu}_N, \|\cdot\|_{\ell_\infty})$, where $\underline{\mu}_N$ denotes the Φ_N-pushforward of the measure μ of X. Clearly

$$\underline{X}_1 \prec \underline{X}_2 \prec \cdots \prec \underline{X}_N \cdots$$

and

$$\underline{X}_N \xrightarrow{\underline{\square}_1} X.$$

It follows in particular that there exist maps $\Psi_N : \underline{X}_N \to X$ which are 1-Lipschitz up to $\varepsilon_N \to 0$ and which almost invert Φ_N (compare $(3'_b)$).

E. Concentration phenomenon

$3\frac{1}{2}$.18. Linear geometry of S^n. We start with several examples of geometric behavior of measures on n-dimensional spaces where interesting patterns develop for large n.

(a) Consider the unit Euclidean ball $B^n \subset \mathbb{R}^n$ and observe that most of its measure is contained in the ε-neighborhood of the boundary. In fact, the proportion of the measure lying outside $U_\varepsilon(\partial B^n)$ equals $(j - \varepsilon)^n$ which is small for large ε/n. In fact, if $\varepsilon \sim \lambda/n$, then $(1-\varepsilon)^n \sim e^{-\lambda}$. This implies the following bound on the di_1-distance between the normalized measures $\mu = \mu_{B^n}$ and $\mu' = \mu_{S^n}$ on B^n,

$$di_1(\mu, \mu') \leq (\log(n))/n.$$

Question: Is the $\underline{\square}_1$-distance between B^n and S^{n-1} realized by the standard embedding $S^{n-1} \to \partial B^n \subset B^n$, or equivalently by the radial projection $\varphi \colon B^n \to S^{n-1}$?

(b) An equatorial sphere $S^{n-1} \subset S^n$ contains most of the measure of S^n in its ε-neighborhood $U_\varepsilon(S^{n-1})$. In fact, most of the measure of S^n is, roughly, $(1/\sqrt{n})$-close to S^{n-1}. To see this, observe that the distance function $d(x) = \mathrm{dist}(x_0, x) - \pi/2$ on S^n pushes forward the spherical measure to the segment $[-\pi/2, \pi/2]$ with the density $\gamma_n(\cos(t))^{n-1}$. Thus, the complementary measure $\mu(S^n \setminus U_\varepsilon(S^{n-1}))$ comes from the contribution to the integral

$$\int_{-\pi/2}^{\pi/2} (\cos t)^{n-1}\, dt$$

from the points in $[-\pi/2, \pi/2] \setminus [-\varepsilon, \varepsilon]$. This is denoted

$$\kappa_n = \kappa_n(\varepsilon) = \frac{\int_\varepsilon^{\pi/2} (\cos t)^{n-1}\, dt}{\int_0^{\pi/2} (\cos t)^{n-1}\, dt}$$

and satisfies

$$\kappa_n(\varepsilon) < 2e^{-(n-1)\varepsilon^2/2}$$

(which is $O(1)$ for $\varepsilon \simeq 1/\sqrt{n}$), where the bound on κ_n is derived from:

(*i*) the inequality $\cos t < e^{-t^2/2}$ for $t \in [-\pi/2, \pi/2]$, $t \neq 0$, following from the Taylor expansions

$$\cos t = \sum_i (-1)^i t^{2i}/(2i)! \qquad \text{and} \qquad e^{-t^2/2} = \sum_i (-1)^i t^{2i}/i!2^i;$$

(*ii*) the relations

$$\int_\varepsilon^{\pi/2} (e^{-t^2/2})^n\, dt = \int_{\varepsilon\sqrt{n}}^{\sqrt{n}\pi/2} e^{-s^2/2}\, ds/\sqrt{n},$$

and

$$\begin{aligned}\int_a^b e^{-s^2/2}\, ds &\leq e^{-a^2/2} \int_0^{b-a} e^{-s^2/2}\, ds \\ &\leq e^{-a^2/2} \int_0^\infty e^{-s^2/2}\, ds \\ &= \sqrt{\pi/2}e^{-a^2/2} \\ &\leq 2e^{-a^2/2},\end{aligned}$$

for $0 \leq a \leq b$.

(*iii*) the estimate

$$\int_0^{\pi/2} (\cos t)^n\, dt \geq 1/\sqrt{n}$$

(see §2 in [Mil–Sch] for details).

Exercise: Evaluate the distance $\underline{\square}_1(S^n, S^{n-1})$ and, more generally, the distance $\underline{\square}_1(S^n, S^m)$. Observe that the sequence of unit spheres S^n with normalized Riemannian measures *does not* $\underline{\square}_1$-converge for $n \to \infty$, nor does it admit a convergent subsequence.

The concentration of the mass of S^n near S^{n-1} implies that most pairs of points $x_1, x_2 \in S^n$ have $\mathrm{dist}_{S^n}(x_1, x_2)$ close to $\pi/2$ for large n (and $\mathrm{dist}_{\mathbb{R}^{n+1}}(x_1, x_2)$ close to $\sqrt{2}$). This brings us to the following concept.

Characteristic size of X. This terminology applies to an X where the distance function dist : $X \times X \to \mathbb{R}$ is me_1-close to a constant function, say d_0, and d_0 is called the *(approximate) characteristic size* of X. To be precise, we have to indicate an ε such that $\mathrm{me}_1(\mathrm{dist}, d_0) \le \varepsilon$ or to apply the notion of the characteristic size to a sequence of spaces X^n, $n = 1, 2, \ldots$, where $\mathrm{me}_1(\mathrm{dist}_{X^n}, d_0) \to 0$ for $n \to \infty$. Notice that the existence of this (asymptotic) characteristic size d_0 for X^n implies that the average diameters of the X^n also converge to d_0 and that the M_r-valued pairwise distance function K_r on $X \times \cdots \times X$ (r factors) me_1 converges to the constant matrix M_{d_0} with all off-diagonal entries equal to d_0. In other words, each measure $\underline{\mu}_r^{X^n}$ converges to a single atom of mass equal to $(\lim_{n\to\infty} \mu(X))^r$ situated at $M_{d_0} \in M_r$.

$3\frac{1}{2}$.19. Coming back to S^n and B^n we can say that these have characteristic sizes about $\mu/2$ and $\sqrt{2}$ respectively for large $n \to \infty$. However, the following striking theorem shows that S^n (and consequently B^n) converges, in some sense, to a single point since every 1 Lipschitz function on S^n for large n becomes me_1-close to a constant.

Levy concentration theorem: *An arbitrary 1-Lipschitz function f on S^n concentrates near a single value $a_0 \in \mathbb{R}$ as strongly as the distance function does. Namely,*

$$\mu\{x \in S^n : |f(x) - a_0| \ge \varepsilon\} < \kappa_n(\varepsilon) \le 2e^{-(n-1)\varepsilon^2/2}, \qquad (*)$$

where μ refers to the normalized spherical measure, and κ_n is the above

$$\kappa_n(\varepsilon) = \frac{\int_\varepsilon^{\pi/2} (\cos t)^{n-1}\, dt}{\int_0^{\pi/2} (\cos t)^{n-1}\, dt}.$$

Remarks on the average and Levy mean. The specific value a_0 which makes $(*)$ sharp is the so-called *Levy mean* of f, and has the property that the level set $f^{-1}(a_0) \subset S^n$ divides the sphere into *equal halves.* More rigorously, this a_0 is uniquely characterized by the inequalities

$$\mu(f^{-1}(-\infty, a_0]) \ge \frac{1}{2} \qquad \text{and} \qquad \mu(f^{-1}[a_0, \infty)) \ge \frac{1}{2},$$

(where we should keep in mind the possibility of $\mu(f^{-1}(a_0)) > 0$). And we would not lose much by using the ordinary *mean*, i.e., *the average* $m^{-1} \int_X f(x)\, dx$ for $m = \mu(X)$ (which equals one for our normalized μ_{S^n}). Namely, if a function concentrates near some value a_0, then this a_0 must be close to the average a of f. In fact, $(*)$ obviously implies that

$$\mu\{x \in S^n : |f(x) - a| \ge 2\varepsilon + \pi\kappa_n\} \le \kappa_n \qquad (\circ)$$

(compare ObsCRad in $3\frac{1}{2}$.31).

The Levy mean of a function f on an arbitrary measure space (X, μ) with $\mu(X) < \infty$ can be adequately expressed entirely in terms of the push-forward measure $\mu_* = f_*(\mu)$ on $\mathbb{R}$. Namely, the Levy mean equals the $a_0 \in \mathbb{R}$ such that

$$\mu_*(-\infty, a_0] \geq \frac{1}{2}\mu_*(\mathbb{R}) \qquad \text{and} \qquad \mu_*[a_0, \infty) \geq \frac{1}{2}\mu_*(\mathbb{R}).$$

In fact, the concentration property of f claimed by Levy's theorem amounts to the concentration of μ_* in the ε-interval around a_0,

$$\mu_*\{y \in \mathbb{R} : |y - a_0| \geq \varepsilon\} \leq \kappa_n \leq e^{-(n-1)\varepsilon^2/2}.$$

On the uniqueness of the Levy mean. If the measure $\mu_* = f_*(\mu)$ in $\mathbb{R}$ has disconnected support, then the Levy mean may be non unique. In fact, the values $a \in \mathbb{R}$ dividing μ into equal halves may range over some interval in $\mathbb{R}$. If this is the case, we *define* the Levy mean to be the *center* of this interval.

Idea of the proof. Fix two positive numbers $\mu_0 \leq 1$ and ε. Consider subsets $X \subset S^n$ with $\mu(X) = \mu_0$ and minimize $\mu(U_\varepsilon(X))$ over all such X, where $U_\varepsilon(X)$ denotes the ε-neighborhood of X. The *spherical isoperimetric inequality* (easily) implies that the minimum is achieved by some ball $B(r) \subset S^n$ (spherical cap), namely the one which has $\mu(B(r)) = \mu_0$.

Now we return to our f on S^n and consider the level $\Sigma = f^{-1}(a_0) \subset S^n$ for the Levy mean a_0 of f. The discussion above shows that

$$\mu(U_\varepsilon(\Sigma)) \geq \mu(U_\varepsilon(S^{n-1})$$

for all $\varepsilon > 0$ and since f is 1-Lipschitz, all of $U_\varepsilon(\Sigma = f^1(a_0))$ goes to the segment $[a_0 - \varepsilon, a_0 + \varepsilon]$.

(Notice that the only technically nontrivial ingredient of this proof is the isoperimetric inequality, which claims that *among all domains in S^n with a given volume, the minimal volume of the boundary is assumed by some ball in S^n.*)

$3\frac{1}{2}$.20. Observable diameter. $\mathrm{ObsDiam}_Y(X) = \mathrm{diam}(X \xrightarrow{\mathrm{Lip}_1} Y, m - \kappa)$. We start by defining the (partial) diameter $\mathrm{diam}(\nu, m - \kappa)$ for a measure ν on a metric space Y as the infimal D such that Y contains a subset Y_0 of diameter $\mathrm{diam}(Y_0) \leq D$ and measure $\nu(Y_0) \geq m - \kappa$, where m stands for the total mass $\nu(Y)$. Clearly, this diameter is monotone for the *Lipschitz* *ordering*: if $\nu = f_*(\mu)$ for a 1-Lipschitz map $f\colon X \to Y$, then $\operatorname{diam}(\nu, m - \kappa) \le \operatorname{diam}(\mu, m - \kappa)$ for all κ. What is not obvious at all is that this diameter may dramatically decrease under all 1-Lipschitz maps f from X to certain Y, such as $Y = \mathbb{R}$ for instance. We define, keeping this in mind, the *observable diameter* $\operatorname{diam}(X \xrightarrow{\mathrm{Lip}_1} Y, m - \kappa)$ as the infimal D such that the measure $\mu_* = f_*(\mu)$ on Y has $\operatorname{diam}(\mu_*, m - \kappa) \le D$ for all 1-Lipschitz maps $f\colon X \to Y$. Here, we think of μ as a *state* on the configuration space X, and f is interpreted as an *observable*, i.e., an observation device giving us the visual (tomographic) image μ_* on Y.

We watch μ_* in Y with a naked eye which has certain sensitivity σ and which cannot distinguish a part of Y of measure (luminicity) less than κ for $\sigma\kappa \ll 1$. (Of course, we could use an observable f with a higher resolution $\lambda = \mathrm{Lip}(f)$, but this can be achieved by just scaling $Y \mapsto \lambda^{-1} Y \overset{\mathrm{def}}{=} (Y, \lambda^{-1}\operatorname{dist}_Y)$, so we can stick to $\lambda = 1$ without losing the freedom of notation.) The observable diameter is usually rather unsensitive to κ when κ is small, and we may pretend that κ is a certain given small number, say $\kappa = 10^{-10}$. Then we abbreviate $\operatorname{diam}(X \xrightarrow{\mathrm{Lip}_1} Y, m - \kappa)$ with $\operatorname{ObsDiam}_Y(X) = \operatorname{ObsDiam}_Y(X, -\kappa)$ and write $\operatorname{ObsDiam}(X)$ for $\operatorname{ObsDiam}_{\mathbb{R}}(X)$. Now, the Levy inequality $(*)$ can be loosely stated as

$$\operatorname{ObsDiam}(S^n) = O(1/\sqrt{n})$$

(and, in fact, $\operatorname{ObsDiam}(S^n) \simeq 1/\sqrt{n}$ since the inequality $(*)$ becomes an equality for the distance function $f(x) = \operatorname{dist}(x_0, x)$ on S^n). More explicitly, $(*)$ implies that

$$\operatorname{diam}(S^n \xrightarrow{\mathrm{Lip}_1} \mathbb{R}, 1 - \kappa) \le \frac{2\sqrt{2}}{\sqrt{n-1}} \sqrt{-\log(\kappa/2)} \approx \frac{1}{\sqrt{n}} \sqrt{-\log \kappa}. \qquad (\star)$$

This should be compared with our earlier evaluation of the characteristic size and/or the average diameter of S^n, which both converge to $\pi/2$ as $n \to \infty$.

$3\frac{1}{2}$.21. Spheres, cubes, and the law of large numbers. The concentration phenomenon generalizes *the law of large numbers*, where X^n is the cartesian power of a given space X, e.g., the n-cube $[0,1]^n$, and then the relevant observable function is the averaged sum of the n coordinate projections $[0,1]^n \to [0,1]$, namely $(1/n)\sum_{i=1}^n t_i$, which concentrates near the single value $1/2$. In geometric language, this says that the normal projection of the cube $[0,1]^n$ to the principal diagonal identified with the segment $[0, \sqrt{n}]$ sends most of (the measure of) the cube to the subsegment

$$[\sqrt{n}/2 - \varepsilon\sqrt{n}/2, \sqrt{n}/2 + \varepsilon\sqrt{n}/2] \subset [0, \sqrt{n}]$$

with small ε. In fact, much of the pushforward measure μ_* is concentrated in the segment $[\sqrt{n}/2 - c, \sqrt{n}/2 + c]$ for some (large) c independent of n, as follows from the central limit theorem. In other words ,

$$\operatorname{diam}(\mu_*, 1 - \kappa) \leq C(\kappa)$$

when $\kappa > 0$, where C is some positive function.

We will generalize this later on to all Lipschitz functions f on $[0,1]^n$, thus showing that

$$\operatorname{diam}([0,1]^n \xrightarrow{\mathrm{Lip}_1} \mathbb{R}, 1 - \kappa) \leq C(\kappa).$$

This is similar to the behavior of the $(n-1)$-sphere of radius $\sqrt{n}$, which has according to $(*)$

$$\operatorname{diam}(S^{n-1}(\sqrt{n}) \xrightarrow{\mathrm{Lip}_1} \mathbb{R}, 1 - \kappa) \simeq \sqrt{-\log(\kappa)}$$

for $n \to \infty$. Actually, this sphere

$$S^{n-1}(\sqrt{n}) = \{x_1, x_2, \dots, x_n : \sum_{i=1}^{n} x_i^2 = n\}$$

is quite similar to the unit cube in many respects. For example, the two have comparable volume, since $\operatorname{vol} S^{n-1}(\sqrt{n}) \approx \mathrm{const}^n$, which is like the volume of the scaled cube $[0, \mathrm{const}]^n$ (while the *unit* sphere has much smaller volume, something of order $n^{-n/2}$). We also know that the average diameter of $S^{n-1}(\sqrt{n})$ (and the characteristic size as well) is about $\sqrt{n}$ (more precisely, it is $\sim \pi\sqrt{n}/2$ for $n \to \infty$), and now we shall see that the same is true for the n-cube. Actually, it is much easier to integrate $(\operatorname{dist}(x,y))^2$ over $[0,1]^n \times [0,1]^n$ rather than $\operatorname{dist}(x,y)$ itself. The former integrates as follows

$$\begin{aligned}
\int\int_{[0,1]^{2n}} \operatorname{dist}^2(x,y)\, dx\, dy &= \int\int_{[0,1]^{2n}} \sum_{i=1}^{n} (x_i - y_i)^2 \\
&= n \int_0^1 \int_0^1 (t-s)^2\, dt\, ds \\
&= n/3.
\end{aligned}$$

On the other hand, $\sup \operatorname{dist}(x,y) = \sqrt{n}$, and so,

$$\sqrt{n} \geq \int\int_{[0,1]^{2n}} \operatorname{dist}(x,y)\, dx\, dy \geq \frac{\sqrt{n}}{3}.$$

Then, by the concentration in $[0,1]^n$ mentioned above, the characteristic size of $[0,1]^n$ is also about $\sqrt{n}$.

Question: What is the $\underline{\square}_\lambda$-distance between $[0,1]^n$ and $S^n(R)$ for given R? (where we refer to the *normalized* measure in S^n)? What are possible Lipschitz maximizations of $S^n(R)$ by $[0,1]^n$ and vice-versa (where it may be reasonable to replace S^n by the ball B^n)? (See $3\frac{1}{2}.59$ for a continuation of this discussion, in which an appropriate distance between $[0,1]^n$ and $S^n(\sqrt{n})$ is estimated as $\log\log(n)$.)

$3\frac{1}{2}$.22. The Maxwell–Boltzmann distribution. We can look at the sphere $S^{3n-1}(\sqrt{n}) \subset \mathbb{R}^{3n} = (\mathbb{R}^3)^n$ as the constant (kinetic) energy surface in the (velocity) space of the system of n (noninteracting) particles in (a box in) $\mathbb{R}^3$, where the sphere equation $\sum_{i=1}^n \|x_i\|^2 = n$ for $x_i \in \mathbb{R}^3$ says that the *energy per particle* equals one independently of n. A man/woman in the street observation consists in measuring the pressure exerted by the particles on one of the walls of the box, which roughly amounts to projecting the velocity vectors of the particles to the normal direction to the wall and then summing up the squares of the positive (i.e., directed towards the wall) projections. A more general (macroscopic) observable associated with a domain Ω in the (velocities) space $\mathbb{R}^3$ is given by the number $N(\Omega)$ of particles (whose velocity vectors are) contained in Ω. Actually, one measures instead the average $N(\Omega)/n$; furthermore, one may consider some function $\varphi\colon \mathbb{R}^3 \to \mathbb{R}$ and take the average $(1/n)\sum_{i=1}^n \varphi(x_i)$, where $x_i \in \mathbb{R}^3$ denotes the (velocity of the) i-th particle, i.e., the i-th component of $x = (x_1, \dots, x_n) \in S^{3n-1}(\sqrt{n})$. Notice that if φ is a λ-Lipschitz function on $\mathbb{R}^3$, then $\Phi(x) = \Phi(x_1, \dots, x_n)$ is $(\lambda/\sqrt{n})$-Lipschitz on $\mathbb{R}^{3n}$ and, in particular, on $S^{3n-1}(\sqrt{n}) \subset \mathbb{R}^{3n}$. Thus, it becomes λ-Lipschitz as we scale $S^{3n-1}(\sqrt{n})$ to the unit sphere $S^{3n-1} = S^{3n1}(1)$. It follows from Levy's theorem, and was known to Maxwell long before Levy, that for most points $x \in S^{3n-1}(\sqrt{n})$, the value $\Phi(x)$ is very close to the average of $\Phi(x)$ over $x \in S^{3n-1}(\sqrt{n})$. This average obviously equals the integral of φ by the measure μ_* on $\mathbb{R}^3$ which is the pushforward of the normalized spherical measure on $S^{3n-1}(\sqrt{n}) \subset (\mathbb{R}^3)^n$ under the projection $x = (x_1, x_2, \dots, x_n) \mapsto x_1$. Now, this measure is given by the density function

$$C_n(\sqrt{1 - \|x_1\|^2/n})^{3n-4}\, dx_1$$

on $\mathbb{R}^3$, which converges as $n \to \infty$ to the Gaussian measure with density

$$Ce^{-3\|x_1\|^2/2}\, dx_1$$

where the (normalization) constant C equals

$$\left(\int_{\mathbb{R}^3} e^{-3\|x_1\|^2/2}\, dx_1\right)^{-1} = \left(\frac{2\pi}{3}\right)^{-1/2}.$$

The asymptotic formula $\mu_* \simeq Ce^{-3\|x_1\|^2/2}$ is known under the name of the *Maxwell (-Boltzmann) distribution law.* It gives a fairly sharp quantitative estimate of concentration of the measure of S^{3n-1} near its equator of codimension 3, but, apparently it tells us precious little about the actual (not the averaged!) distribution of the coordinates (particles) $x_1, \dots, x_n$ in $\mathbb{R}^3$ of an individual configuration $x = (x_1, x_2, \dots, x_n) \in S^{3n-1}(\sqrt{n})$. Yet, by the Levy (-Maxwell) theorem, the distribution of the points $x_1, \dots, x_n$ in $\mathbb{R}^n$ closely follows $Ce^{-3\|x\|^2/2}$ for most $x \in S^{3n-1}(\sqrt{n})$. Namely, the average of the atomic measures concentrated at x_i, denoted $(1/n)\sum_{i=1}^n \delta(x_i)$ stays di_1-close to the Gaussian measure with the density $Ce^{-3\|x\|^2/2}$ for most points $x \in S^{3n-1}(\sqrt{n})$. Here, we think of $x \mapsto (1/n)\sum_{i=1}^n \delta(x_i)$ as a map f_n from $S^{3n-1}(\sqrt{n})$ to the space M of probability measures on $\mathbb{R}^3$ with the metric di_1 and claim that these $f_n\colon S^{3n-1}(\sqrt{n}) \to M$ concentrate to the constant map sending the sphere to the Gaussian measure. (We shall discuss on several occasions the concentration of maps $X \to Y$ for general Y, see, e.g., $3\frac{1}{2}.41$.)

The above is similar to the *law of large numbers for mm spaces* X, where we look at the map f_n from $X^n = X \times X \times \cdots \times X$ (n factors) to the space $P = P(X)$ of probability measures on X which assigns the measure $(1/n)\sum_{i=1}^n \delta(x_i)$ to each $(x_1, \dots, x_n) \in X^n$. The law of large numbers claims that the concentration of f_n for large n to the constant map sending most of X^n to the normalized measure $\mu/\mu(X)$ on X. Moreover, if μ is normalized, (i.e., $\mu(X) = 1$), one has the *strong law of large numbers.* This applies to the infinite product $X^\infty = (X, \mu) \times (X, \mu) \times \cdots$ and claims that the sequence of measures $f_n(\overline{x}) = (1/n)\sum_{i=1}^n \delta(x_i)$ on X, with $\overline{x}$ standing for $(x_1, x_2, \dots, x_n, \dots) \in X^\infty$, converges for almost all $\overline{x} \in X^\infty$ to the original measure $\mu \in P(X)$. (This trivially follows, for example, from the individual ergodic theorem applied to the shift $(x_1, x_2, x_3, \dots) \mapsto (x_2, x_3, \dots)$ on X^∞.) In other words, almost every sequence $x_1, x_2, \dots$ is *uniformly distributed* on X. Consequently, every *mm*-space of finite mass admits a uniformly distributed sequence of points (see [Ellis]).

$3\frac{1}{2}$.23. Normal law à la Levy. The projection of $S^n(\sqrt{n})$ to $\mathbb{R}$ pushes forward the normalized measure μ on $S^n(\sqrt{n})$ to some measures μ_n which converge as $n \to \infty$ to the Gaussian measure $\nu = Ce^{-x^2/2}\,dx$ on $\mathbb{R}$ for $C = (2\pi)^{-1/2}$ by the Maxwell-Boltzmann theorem. And Levy's theorem (or rather the underlying spherical isoperimetric inequality) implies that *this ν majorizes every measure ν' on $\mathbb{R}$ which appears as a limit of pushforwards of μ under 1-Lipschitz maps $f_n : S^n(\sqrt{n}) \to \mathbb{R}$.* Namely, if x and x' are points in $\mathbb{R}$ such that $\nu(-\infty, x] \le \nu'(-\infty, x']$ and $\nu[x, \infty) \le \nu'(x', \infty]$ (we could have said $\nu(-\infty, x] = \nu'(-\infty, x']$ if we knew that x' was not an atom for ν'), then the density of ν' at x' is greater than $Ce^{-x^2/2}$. It follows that *there exists a unique monotone increasing* 1*-Lipschitz map* $\alpha\colon \mathbb{R} \to \mathbb{R}$ *pushing forward* ν *to* ν'.

Now, we turn to 1-Lipschitz maps $f_n : S^n(\sqrt{n}) \to \mathbb{R}^k$ for a given $k \geq 2$ and look at the measures ν' arising as limits of $(f_n)_*(\mu)$ on $\mathbb{R}^k$. There is a distinguished measure among these ν', namely the Gaussian $\nu = C_k e^{-\|x\|^2/2}\,dx$ for $C_k = (2\pi)^{-k/2}$. One may expect that this ν majorizes all ν' in a reasonable way, i.e., there exists a "nice" (1-Lipschitz?) map $\alpha\colon \mathbb{R}^k \to \mathbb{R}^k$ with $\alpha_*(\nu) = \nu'$ for all ν'.

Observe that *all measures* ν' *above obviously satisfy the isoperimetric inequality induced by that in* $S^n(\sqrt{n})$ *for* $n \to \infty$. Namely, if Ω is an arbitrary domain and $B \subset \mathbb{R}^k$ is a halfspace such that $\nu(B) = \nu'(\Omega)$, then *the* ν'*-measure of the* ε*-neighborhood of the boundary* $\partial\Omega$ *of* Ω *is bounded from below by the* ν *measure of such a neighborhood of* ∂B*, i.e.,* $\nu'(U_\varepsilon(\partial\Omega)) > \nu(U_\varepsilon(\partial B))$ *for all* $\varepsilon \geq 0$.

Actually, this result is quite nontrivial for the Gaussian measure ν itself, i.e., for $\nu' = \nu.$, where it is highly appreciated by the probability people and is called *the Gaussian isoperimetric inequality* (see [Tal]$_{\text{NII}}$ and the references therein).

Exercise: Identify the extremal domains Ω and measures ν' where

$$\lim_{\varepsilon \to 0} \nu'(U_\varepsilon(\partial\Omega))/\nu(U_\varepsilon(\partial B)) \to 1$$

by adopting the concavity argument from $3\frac{1}{2}$.27.

One could replace the spheres $S^n(\sqrt{n})$ above by another natural sequence of mm spaces X^n with $\mathrm{ObsDiam}(X^n) = O(1)$ for $n \to \infty$ and try to characterize the (space of) limits of the pushforward measures ν' under 1-Lipschitz maps $X^n \to \mathbb{R}^k$. However, I doubt that even the case $X^n = [0,1]^n$ has been thoroughly investigated, not to mention many other sequences, such as $\sqrt{n}\,\mathbb{C}P^n$, $\sqrt{n}\,\mathbb{H}P^n$, $S^{n_1}(\sqrt{n_1}) \times S^{n_2}(\sqrt{n_2}) \times \cdots \times S^{n_m}(\sqrt{n_m})$, ℓ_p-products, etc., which we shall meet in the following sections. Sometimes one may try the first eigenfunction (or a combination of these) of the Laplace operator on X^n instead of the linear projection $S^n \to \mathbb{R}$ in order to generate the corresponding "normal law."

$3\frac{1}{2}$.24. Approximating the spherical means on $S^n(\sqrt{n})$ by the Gaussian $C_{n+1}e^{-\|x\|^2/2}\,dx$. The Gaussian density is constant on the concentric spheres in $\mathbb{R}^{n+1}$ and equals $C'R^n e^{R^2/2}\,dR\,ds$ in polar coordinates. The density function $R^n e^{-R^2/2}$ achieves its maximum at $R = \sqrt{n}$ and is rather sharply concentrated near this $R = \sqrt{n}$. Namely, the measure outside the segment $[\sqrt{n}-r, \sqrt{n}+r]$ has mass $\leq 2e^{-r^2/2}$. This follows trivially from the Gaussian isoperimetric inequality above or can be shown by a direct, elementary computation appealing to the strict log-*concavity* of $R^n e^{R^2/2}$, i.e., to the inequality $(\log R^n e^{-R^2/2})'' \leq -1$. Thus, almost all the mass of the Gaussian measure $C_{n+1}e^{-\|x\|^2/2}$ is located near the sphere $S^n(\sqrt{n}) \subset \mathbb{R}^{n+1}$. In fact, the Gaussian measure outside the r-neighborhood $A_r(\sqrt{n})$ of this sphere is $\leq 2e^{-r^2/2}$.

Exercise: Majorize the sphere by the above r-neighborhood $A_r(\sqrt{n})$ with Gauss measure via the normal (radial) projection $A_r(\sqrt{n}) \to S^n(\sqrt{n}-r)$ and thus derive the spherical isoperimetric inequality (with an almost sharp constant) from the (sharp!) Gaussian one (mentioned earlier).

Notice that the Gauss measure $C_{n+1}e^{-\|x\|^2/2}\,dx$ equals the product of $n+1$ copies of $e^{-x^2/2}\,dx$ on $\mathbb{R}$. Thus, the spherical measure may be seen as an approximate product measure which makes the similarity between $S^n(\sqrt{n})$ and $[0,1]^n$ less surprising (compare Appendix V in [Mil–Sch]).

$3\frac{1}{2}$.25. Canonical and microcanonical. The above illustrates the physical idea of the *equivalence between the canonical and microcanonical measures* in statistical mechanics which applies to certain (physically meaningful) functions' "energies" $E(x) = E_n(x)$ on $\mathbb{R}^n$. Recall that one introduces the *microcanonical measure* $\mu_{\mathrm{mic}} = \mu_{\mathrm{mic}}(\rho)$ supported on the energy surface

$$S(\rho) = \{x \in \mathbb{R}^n : E(x) = \rho\}$$

by first restricting the Lebesgue measure μ to the "annuli"

$$S_r(\rho) = \{x \in \mathbb{R}^{n+1} : \rho - r \leq E(x) \leq \rho + r\},$$

then by normalizing this $\mu|_{S_r(\rho)}$ and by finally letting $r \to 0$. Actually, one does not need $r \to 0$ but rather $r/\rho \to 0$ as $n \to \infty$. In fact, one typically has $\rho = \rho_n \simeq n$ for $n \to \infty$, i.e., having bounded energy per particle, while $r = r_n = o(n)$.

Next, one introduces the *canonical measure* $\mu_{\mathrm{ca}} = \mu_{\mathrm{ca}}(\beta)$ with the normalized density $e^{-\beta E(x)}$,

$$\mu_{\mathrm{ca}} = \mathrm{const}\, e^{-\beta E(x)}\mu,$$

where $\mu = dx$ is Lebesgue measure and

$$\mathrm{const} = \mu(e^{-\beta E(x)})^{-1} = \left(\int_{\mathbb{R}^n} e^{-\beta E(x)}\,dx\right)^{-1}.$$

The above-mentioned equivalence $\mu_{\mathrm{ca}} \sim \mu_{\mathrm{mic}}$ refers to a certain correspondence $\rho \mapsto \beta = \beta(\rho)$ so that the canonical measure $\mu_{\mathrm{ca}}(\beta)$ remains rather close to $\mu_{\mathrm{mic}}(\rho)$ for suitable $\rho = \rho_n$, $\beta = \beta(\rho_n)$ and $n \to \infty$. This closeness can be expressed in terms of our metrics di_λ, Tra_λ, and/or Lid_b, or, more generally, in terms of

$$|\mu_{\mathrm{ca}}(\Phi) - \mu_{\mathrm{mic}}(\Phi)|$$

for a suitable class of functions (observables) Φ on $\mathbb{R}^n$. In particular, we can insist that the E-pushforwards of μ_{ca} and μ_{mic} be mutually close in $\mathbb{R}$. Since $\mu_{\mathrm{mic}}(\rho)$ goes to the δ-measure $\delta(\rho)$ on $\mathbb{R}$, this amounts to a sufficient concentration of $\underline{\mu}_{\mathrm{ca}} = E_*(\mu_{\mathrm{ca}})$ near $\rho = \rho_n$ for $\beta = \beta_n = \beta(\rho_n)$. In other words, the $\underline{\mu}_{\mathrm{ca}}$-measure of the segment $[\rho - r, \rho + r]$ must be close to one for relatively small r (e.g., $\rho = n$ and $r = o(n)$).

Here is the common physicists' recipe for determining $\beta = \beta(\rho)$. Let $\underline{\mu} = E_*(\mu)$ be the pushforward of the Lebesgue measure of $\mathbb{R}^n$, and write

$$\underline{\mu}_{\mathrm{ca}} = Ce^{-\beta\rho}\underline{\mu} = \varphi(\rho)e^{-\beta\rho}\,d\rho$$

for the constant

$$C = C_\beta = \left(\mu(e^{-\beta E(x)})\right)^{-1} = (\underline{\mu}(e^{-\beta\rho}))^{-1},$$

and the density function φ of the measure $C\underline{\mu}$. We choose β by requiring that the derivative of the function $\psi(\rho) = \psi_\beta(\rho) = \varphi(\rho)e^{-\beta\rho}$ vanishes at given $\rho = \rho_n$ (hoping that this is a maximum of $\psi(\rho)$ with sharp drop of ψ away from this $\rho = \rho_n$). Thus,

$$\psi'_\beta = \varphi'(\rho)e^{-\beta\rho} - \beta\varphi(\rho)e^{-\beta\rho} = 0,$$

and so

$$\beta = \beta(\rho) = (\log\varphi(\rho))'.$$

Everything here works fine if φ is smooth and $(\log\varphi)'' \leq 0$, but this is rarely observed in nontrivial physical situations unless we pass to a (suitably renormalized) limit for $n \to \infty$. Nevertheless, there are (artificial from a physicist's point of view) examples where these conditions are met before passing to the limit. For instance, if $E(x)$ is a positive homogeneous function on $\mathbb{R}^n$ of degree $p \geq 1$, i.e., $E(\lambda x) = \lambda^p E(x)$, then clearly $\varphi(\rho) = \mathrm{const}\,\rho^{n/p-1}$ and for $\rho = n$ we obtain $\beta_n = \frac{1}{p} - \frac{1}{n}$, which can be replaced for our purposes by $\frac{1}{p}$ for large n. Then, the measure $\underline{\mu}_{\mathrm{ca}} = C\rho^{n/p-1}e^{-\rho/p}$ concentrates near the value $\rho = n$ in the sense that

$$\underline{\mu}_{\mathrm{ca}}[n - \lambda\sqrt{n}, n + \lambda\sqrt{n}] \to 1$$

as $\lambda \to \infty$, where the convergence is uniform in n as an elementary computation shows (substitute $\rho = t^2$ and appeal to the strict log-concavity of $t^{2n/p1}e^{-t^2/p}$).

Finally, let us estimate the distance between μ_{ca} and μ_{mic} on $\mathbb{R}^n$ as follows. Denote by δ the infimum of the radial derivatives of $E(x)$ on the unit level $S(1) = \{x \in \mathbb{R}^n : E(x) = 1\}$ and observe that the "annulus"

$$S_r(n) = \{x \in \mathbb{R}^n : \rho - r \leq E(x) \leq \rho + r\}$$

is contained in the s-neighborhood $U_s(S(n))$ with s equal to about $r/\delta n^{(p-1)/p}$. It follows that most of the μ_{ca}-measure is contained in $U_s(S(n))$ with $s = \lambda\delta^{-1}n^{\frac{1}{p}-\frac{1}{2}}$, provided that λ is large (independently of n). This agrees with our previous computation for $E(x) = \sum_{i=1}^n x_i^2$ where $p = 2$ and $\delta = 2$. Similarly, if $E(x) = \sum_{p=1}^n x_i^p$, then by a straighforward computation

$$\delta = \begin{cases} p & \text{for } p \leq 2 \\ pn^{\frac{1}{p}-\frac{1}{2}} & \text{for } p \geq 2 \end{cases}$$

and we get the (well-known) estimates

$$s = O(n^{\frac{1}{p}-\frac{1}{2}})$$

for $p \leq 2$ and

$$s = O(1)$$

for $p \geq 2$ (where the special role of $p = 2$ is due to the quadratic nature of the Euclidean metric defining our s-neighborhood $U_s(S(r))$.

Later on, we shall see, following G. Schechtman, how the above leads to the following bound on the ℓ_p-ball

$$B_p(n^{1/p}) = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^p \leq n\} :$$

$$\operatorname{ObsDiam} B_p(n^{1/p}) = O(1),$$

i.e., ObsDiam remains bounded for $n \to \infty$.

$3\frac{1}{2}$.26. Concentration as ergodicity. The ergodicity property of a measurable transformation can be expressed by saying that the orbit (saturation) of every subset of *positive* measure has *full* measure. Similarly, we may loosely say that an mm space X is "infinitely concentrated" if the ε-neighborhood of each subset $X_0 \subset X$ with positive measure has full measure, where ε is an "infinitesimally small" number. Here, the relevant dynamics is the growth of subsets $X_0 \mapsto X_\varepsilon = U_\varepsilon(X_0)$, which in some cases comes from an ordinary flow. For example, if X is a Riemannian manifold, then $U_\varepsilon(X_0)$ is obtained by first lifting (i.e., pulling back) X_0 to the unit tangent bundle, say to $\tilde{X}_0 \subset T_1(X)$ over X, then applying the geodesic flow for the time $t \in [0, \varepsilon]$, and finally projecting the resulting $[0, \varepsilon]$-orbit of $\tilde{X}_0$ back to X.

It is unclear if the ergodic analogy goes anywhere far on the technical level, but it certainly suggests many interesting questions. For example, one may ask what happens for more general, nonmetric "ε-saturations" of subsets X_0 in a measure space X with a suitable structure, and many results here have been obtained by M. Talagrand (see $[\mathrm{Tal}]_{\mathrm{CMII}}$). We also shall see later on how the ideas of *ergodic decomposition* and *dissipation* fare in the category of mm spaces.

$3\frac{1}{2}$.27. Generalized Levy's inequalities. Levy himself generalized the spherical isoperimetric inequality to convex hypersurfaces $\Sigma^n \subset \mathbb{R}^{n+1}$ with principal curvatures ≥ 1. Since then, his argument has been extended to all closed n-dimensional Riemannian manifolds V with $\operatorname{Ricci} V \geq n-1 = \operatorname{Ricci} S^n$ (see Appendix C_+ and $[\mathrm{Croke}]_{\mathrm{IIEE}}$, $[\mathrm{Gal}]_{\mathrm{Iso}}$, $[\mathrm{Gal}]_{\mathrm{IIA}}$, $[\mathrm{Gal}]_{\mathrm{ES}}$, [Ber–Gal]).

This extended Levy inequality implies that every such V (e.g., the above Σ^n) satisfies $(*)$ in (1) with the same constant κ_n. In particular, $\operatorname{ObsDiam}_{\mathbb{R}}(V; -\kappa) \leq \operatorname{ObsDiam}_{\mathbb{R}}(S^n, -\kappa)$ for all $\kappa > 0$.

Another generalization concerns certain path metrics and measures supported in convex domains $\Omega \subset S^n$ (we also allow $\Omega = S^n$), where the geodesics of the metric coincide with the arcs of great circles, and where the measures μ in question must satisfy some convexity condition somewhat similar to that implied by $\operatorname{Ricci} \geq \varepsilon > 0$. Namely, assuming that μ has a continuous density function, we define a *convex descendent* ν of μ on each geodesic arc $I \subset \Omega$ as the weak (i.e., di_1) limit of the sequence $\mu_i = (\mu|_{U_i})/\mu(U_i)$, where U_i is a decreasing sequence of *convex* neighborhoods of $I \subset \Omega$ with $\bigcap_i U_i = I$. (Notice that each arc I contained in a hemisphere I carries at least one convex descendent ν of μ, since we always can find U_i for which μ_i converge and, in general, there may be many ν on the same I). Then the convexity of μ refers to the convexity properties of the (density functions) of the measures ν on the segments. In particular, one can measure the convexity (or rather concavity) of μ by the "supremum of the *separation distances*" (see (1) $3\frac{1}{2}$.30)

$$C_\lambda(\mu; \kappa) = \sup_\nu \operatorname{Sep}(\nu; \kappa, \lambda\kappa),$$

where the sup is taken over all geodesic segments $I \subset \Omega$ and all convex descendents ν on I.

Separation inequality ([Gro–Mil]$_{\mathrm{GSII}}$). *The separation distance of a normalized measure μ is bounded by*

$$\mathrm{Sep}(\mu, \kappa, \lambda\kappa) \le C_\lambda(\mu; \kappa) \qquad (+)$$

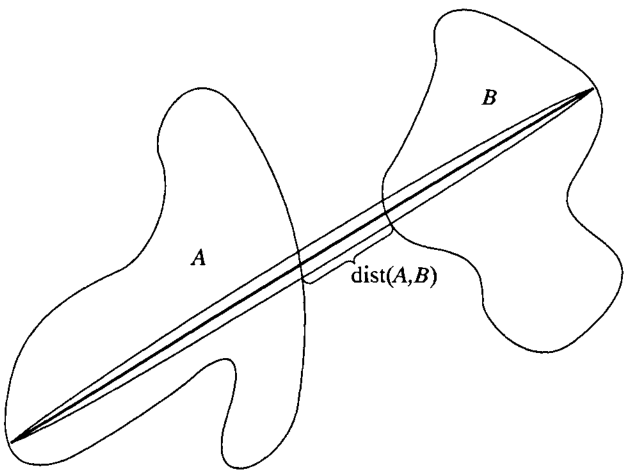
for all $\kappa, \lambda > 0$, where we assume that μ is normalized by $\mu(\Omega) = 1$.

We shall indicate the proof below. Here, we observe that (+) rather easily yields the sharp spherical isoperimetric inequality when applied to the spherical measure μ on S^n. Also, (+) gives the following bound on the observable diameter of the unit ball in the ℓ_p-space $(\mathbb{R}^n, \|\cdot\|_{\ell_p})$ for $\|x\|_{\ell_p} = (\sum_i |x_i|^p)^{1/p}$,

$$\mathrm{ObsDiam}_{\mathbb{R}}(B(1), -\kappa) \le \mathrm{const}_p(\kappa)/\sqrt{n}$$

for all $1 < p < \infty$, where the measure on $B(1)$ in question is just the restriction of the Lebesgue measure from $\nu \supset B(1)$ (see [Gro–Mil]$_{\mathrm{GSII}}$ for a more precise statement applicable to all uniformly convex Banach spaces).

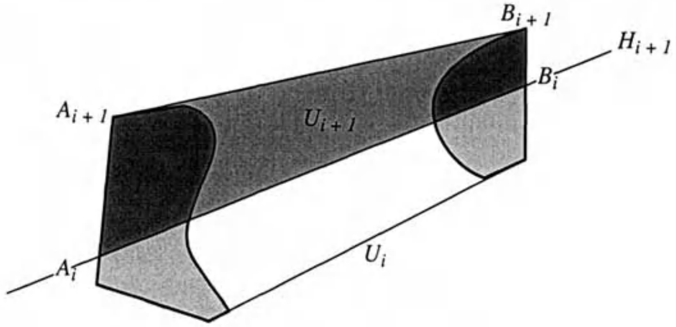
Sketch of the proof of (+). We want to bound the distance between two subsets A and B in $\Omega \subset S^n$ with $\mu(A) \ge \kappa$ and $\mu(B) \ge \lambda\kappa$ by finding a segment I in Ω with an "infinitely thin" convex neighborhood U (called a *needle* in [Kan–Lov–Sim]), such that $\mu(U \cap A) \ge \kappa\mu(U)$ and $\mu(U \cap B) \ge \lambda\kappa\mu(U)$.



To put it another way, we need a descending sequence of convex domains $U_1 \supset U_2 \supset \cdots \supset U_i \supset \cdots$ in S^n such that $\mu(U_i \cap A) \ge \kappa\mu(U_i)$, $\mu(U_i \cap B) \ge \lambda\kappa\mu(U_i)$, and the intersection $\bigcap_{i=1}^{\infty} U_i$ is 1-dimensional (i.e., a segment). Our construction of U_i is based on the (trivial special case of the) Borsuk -Ulam theorem. This provides a hyperplane (i.e., an equatorial hypersurface) H_1 in S^n dividing both A and B into equal halves, i.e., of measures $\mu(A)/2$ and $\mu(B)/2$, respectively. We take the smaller of the two halves of S^n for U_1. Thus, U_1 is a hemisphere with

$$\begin{aligned} \mu(U_1) &\leq \frac{1}{2} \\ \mu(U_1 \cap A) &= \mu(A)/2 \geq \kappa\mu(U_1) \\ \mu(U_2 \cap B) &= \mu(B)/2 \geq \lambda\kappa\mu(U_2). \end{aligned}$$

Then we apply the same procedure to $A_1 = U_1 \cap A$ and $B_1 = U_1 \cap B$ in place of A and B. Namely, we divide A_1 and B_1 into equal halves by a hyperplane H_2 and take the smallest of the halves of U_1 for U_2. Next, we pass to U_3 by equi-dividing $A_2 = U_2 \cap A_1$ and $B_2 = U_2 \cap B_1$, and so on.



It is clear that the limit $I = \bigcap_{i=1}^{\infty} U_i$ must be one dimensional (otherwise we just keep dividing) and the proof follows.

Remarks: (a) The Borsuk–Ulam subdivision argument can also be applied to k subsets in S^n for $k \geq 3$, which leads to a bound on $\mathrm{Sep}(\mu, \kappa_1, \kappa_2, \dots, \kappa_k)$ in terms of $(k-1)$-dimensional convex descendents of μ, but the merit of the resulting inequalities remains uncertain (compare (j)).

(b) There are other interesting classes of (nonconcave) measures μ on S^n showing definite patterns of concentration, e.g., those given by polynomial densities of low degree, but little has been understood about them so far (compare [Bourg]).

(c) **Log-concavity.** Recall that a function φ on an affine space X is called log-*concave* (or *multiplicatively concave*) if

$$\varphi\left(\frac{x+y}{2}\right) \geq \sqrt{\varphi(x)\varphi(y)},$$

or equivalently, if $\log(\varphi)$ is concave. If $X = \mathbb{R}$, then this is equivalent to $(\log\varphi)'' \leq 0$, and the *δ-sharp* log-concavity refers to the inequality $(\log\varphi)'' \leq -\delta < 0$. The archtypical example is $\varphi(x) = e^{-x^2/2}$, which is 1-sharp log-concave, since $(\log e^{-x^2/2})'' = -1$. Similarly, one defines δ-sharpness for log-concavity on $\mathbb{R}^n$ by requiring this property for the restrictions of φ to all lines.

Exercises: (d) Show that every probability measure μ on $\mathbb{R}$ (or on a segment in $\mathbb{R}$) with δ-sharp log-concave density function is more concentrated than the Gaussian measure $\mu_\delta = \sqrt{2\delta/\pi}e^{-\delta x^2/2}$ on $\mathbb{R}$, namely,

$$\operatorname{Sep}(\mu; x, \lambda x) \leq \operatorname{Sep}(\mu_\delta, x, \lambda x)$$

for all $x, \lambda \geq 0$.

(e) Show that the convex descendents of a log-concave measure μ on $\mathbb{R}^n$ are log-concave and δ-sharpness is also inherited by the convex descendents of μ.

(f) Prove an isoperimetric inequality for δ-sharp log-concave functions generalizing the Gaussian isoperimetric inequality above.

(g) Show that the pushforwards of log-concave measures in convex domains in $\mathbb{R}^n$ under linear projections $\mathbb{R}^n \to \mathbb{R}^m$ are log-concave and that δ-sharpness also persists under projections (compare the proof of the Brunn-Minkowski theorem in [Gro–Mil]$_{\text{GSII}}$).

(h) Say that φ on $\mathbb{R}^n$ is *δ-sharp in codimension k* if it is log-concave, and for each x, there is an $(n-k)$-plane $P \subset \mathbb{R}^n$ passing through x such that $\varphi|_P$ is δ-sharply log-concave on P at x, i.e., $\partial^2 \log\varphi \leq -\delta$ for all unit vectors ∂ at x tangent to P. Show that a probability measure μ on (a convex domain in) $\mathbb{R}^m$, with density φ^n, where φ is δ-sharp in codimension k, concentrates near some k-plane $F \subset \mathbb{R}^m$ in the following sense:

the μ-measure of the complement of the $\varepsilon\sqrt{m}$ neighborhood of F is $\leq 2e^{-n\delta\varepsilon^2/2}$.

(i) Prove that "δ-sharp in codimension k" is stable under linear projections of measures, as well as under passing to convex descendents of *all* dimensions ≥ 1 (as in (a)).

Combining (h) and (i) shows that linear projections of sufficiently sharp in codimension k measures from $\mathbb{R}^n \to \mathbb{R}^m$ are concentrated near some k-dimensional linear subspaces. It is unclear how to prove a similar result for nonlinear 1-Lipschitz maps $\mathbb{R}^n \to \mathbb{R}^m$ where one wants the pushforward measures to concentrate near something of dimension $\leq k$. Nevertheless, we have the following alternative.

(j) Define the "k-dimensional distance" $\Delta_k(A_0, \ldots, A_k)$ for $A_i \subset \mathbb{R}^n$ as the infimum of the k-volumes of the k-simplices in $\mathbb{R}^n$ with vertices $a_i \in A_i$. Then give a lower bound on $\Delta_k(A_0, \ldots, A_k)$ in terms of $\mu(A_i)$, $i = 0, \ldots, k$, for a probability measure $\mu = \varphi\, dx$, where the density φ is δ-sharp in codimension $k-1$ in some convex domain in $\mathbb{R}^n$. (*Hint:* See (a).)

(k) **Convexity and log-convexity.** The Brunn theorem states that the pushforward of the Lebesgue measure dx on a convex subset X in $\mathbb{R}^n$ becomes $\sqrt[k]{\ }$-concave under the projection $\mathbb{R}^n \to \mathbb{R}^{n-k}$, where the $\sqrt[k]{\ }$-concavity of a measure $\varphi\, dx$ means concavity of $(\varphi)^{1/k}$ at the points where $\varphi > 0$. More generally, $\sqrt[\alpha]{\ }$ concave measures in (convex domains in) $\mathbb{R}^n$ go to $\sqrt[\beta]{\ }$-concave ones in (convex domains in) $\mathbb{R}^m$ for $\alpha - \beta = n - m$, as a simple induction on $n - m$ shows (see [Gro- Mil]$_{\mathrm{Brunn}}$). Furthermore, a suitably defined strict (or sharp) $\sqrt[\alpha]{\ }$-concavity should descend to the corresponding strict (or sharp) $\sqrt[\beta]{\ }$-concavity of the projection, where we want to model such a descent on what happens to the projection of dx on the round ball $B \subset \mathbb{R}^n$ as it projects to $\mathbb{R}^{n-k}$ and then to $\mathbb{R}^{n-\ell}$, $\ell > k$. We suggest that the reader investigate this more carefully (perhaps by looking through the literature).

Here, we indicate how to derive possible corollaries concerning the concentration of the projections by approximating the Lebesgue measure dx on a convex body $X \subset \mathbb{R}^n$ by a log-concave measure. To do this, we assume that X contains the origin and then consider the associated "norm" $\|x\|_0$ on $\mathbb{R}^n$, i.e., the homogeneous positive function equal to one on the boundary of X. Now we take the measure $e^{-\|x\|_0^2}\, dx$ and observe that the sharpness of the log concavity of this measure reflects the strictness of the convexity of X. On the other hand, most of this measure is concentrated in a narrow band of width about $1/\sqrt{n}$ around ∂X, and so the concentration properties of functions on X (or ∂X) with respect to dx, as well as those of projections $X \to \mathbb{R}^m$, can be derived from the corresponding properties of $e^{-\|x\|_0^2}$ (compare $3\frac{1}{2}.24$).

(k') Suppose that X is strictly convex in codimension k, which signifies here that at each boundary point $x \in X$, $k - n - 1$ (out of $n-1$) principal curvatures of X are ≥ 1, where ∂X is assumed C^2-smooth for the sake of simplicity. Then every orthogonal projection $X \to \mathbb{R}^m$ concentrates near some k plane F, i.e., of the measure dx on X pushed forward to $\mathbb{R}^m$ in the

What would be more interesting is to prove similar concentration of every 1-Lipschitz map $X \to \mathbb{R}^m$ with F replaced by a more general (nonflat) k-dimensional subset.

(k'') Use $e^{-\|x\|_0^2}\,dx$ to prove the isoperimetric inequality from [Gro–Mil]$_{\text{GSII}}$ for uniformly convex Banach spaces.

$3\frac{1}{2}$.28. Subvarieties in S^n, $\mathbb{R}\mathrm{P}^n$, and $\mathbb{C}\mathrm{P}^n$. Consider a submanifold $X \subset S^n$ of codimension k which satisfies the following two conditions:

(*i*) X meets each k-dimensional equatorial sphere in S^n in at most d points.

(*ii*) $\operatorname{vol}_{n-k} X \geq c \operatorname{vol} S^{n-k}$.

(Notice that necessarily $c \leq d/2$, since the average number of intersection points of X with equatorial k-spheres equals $2\operatorname{vol}_{n-k}(X)/\operatorname{vol} S^{n-k}$ by the Crofton formula. Also observe that (*i*) is satisfied with $d = d(\delta)$ for semialgebraic subsets in S^n of degree δ.) Take a subset $X_0 \subset X$ of relative measure κ (i.e., $\operatorname{vol}_{n-k}(X_0) = \kappa \operatorname{vol}_{n-k}(X)$) and observe that the ε neighborhood of X_0 in S^n satisfies

$$\mu(U_\varepsilon(X_0)) \geq \kappa_\varepsilon = 2cd^{-1}b(k,\varepsilon)\kappa, \tag{$*$}$$

where $\mu = \mu_n$ denotes the normalized $O(n)$-invariant measure (volume) in S^n and $b(k,\varepsilon)$ is the μ_k-measure of the ε-ball in S^k, where μ_k is also normalized. This follows from Crofton's formula, since every equatorial S^k meeting X_0 contains an ε-ball inside $U_\varepsilon(Y)$ (compare [Milm]). Then we can estimate the distance between two such subsets, say X_0 and X_1, in X by the distance between $U_\varepsilon(X_0)$ and $U_\varepsilon(X_1)$ plus 2ε, which allows a bound on $\operatorname{ObsDiam} X$ by $\operatorname{ObsDiam} S^n$ related to the *separation distance* (see $3\frac{1}{2}$.3D). Namely,

$$\begin{aligned}
\operatorname{ObsDiam}(X;-2\kappa) &\leq \operatorname{Sep}(X;\kappa,\kappa)\\
&\leq \operatorname{Sep}(S^n;\kappa_\varepsilon,\kappa_\varepsilon) + 2\varepsilon\\
&\leq \operatorname{ObsDiam}(S^n,-\kappa_\varepsilon) + 2\varepsilon\\
&\leq \frac{2\sqrt{2}}{\sqrt{n}}\sqrt{-\log(\kappa_\varepsilon/2)} + 2\varepsilon\\
&\leq \frac{2\sqrt{2}}{\sqrt{n}}\sqrt{-\log(2cd^{-1}b(k,\varepsilon)\kappa)} + 2\varepsilon.
\end{aligned}$$

To make use of this, we have to bound $b(k,\varepsilon)$ from below and then judiciously choose ε. We roughly estimate $b(k,\varepsilon) \geq (\varepsilon/3)^k$ (by looking at the measure $(\cos t)^k\,dt$ on $[-\pi/2, \pi/2]$) and take $\varepsilon = \sqrt{k/n}$. Then we get

$$\mathrm{ObsDiam}(X, -\kappa) \leq \sqrt{\frac{k}{n}}\left(4 + \sqrt{-\log\left(\frac{k}{n}cd^{-1}\kappa\right)}\right), \qquad (+)$$

which is only slightly more than $\sqrt{k/n}$ for each given $\kappa > 0$ and $cd^{-1} \geq$ const. Actually, this "more" seems excessive and the best possible (but maybe impossible?) inequality should read

$$\mathrm{ObsDiam}(X, -\kappa) \leq \sqrt{\frac{k}{n}}(C_1 + C_2\sqrt{-\log(cd^{-1}\kappa)}).$$

Exercise: Consider the union $X = S_1^{n-k} \cup S_2^{n-k}$ of two mutually orthogonal equators S_1^{n-k} and S_2^{n-k} in S^n and show that the observable diameter of this X is about $\sqrt{k/n}$ (compare (e) below).

The case of $\mathbb{C}\mathrm{P}^n$. The Hopf fibration $S^{2n+1} \to \mathbb{C}\mathrm{P}^n$ tells us that $\mathbb{C}\mathrm{P}^n \prec S^{2n+1}$ (in the sense of $3\frac{1}{2}.16$), and so

$$\mathrm{ObsDiam}\,\mathbb{C}\mathrm{P}^n \leq \mathrm{ObsDiam}\,S^{2n+1} \leq 1/\sqrt{n}.$$

Furthermore, the above argument carries over to subsets $X \subset \mathbb{C}\mathrm{P}^n$ and applies in particular to *complex algebraic* X where (properly redefined) d/c equals 1. *It follows that*

$$\mathrm{ObsDiam}(X, -\kappa) \leq 10\sqrt{\frac{k}{n}}\sqrt{-\log\frac{k}{n}\kappa} \qquad (+)_{\mathbb{C}}$$

for all complex algebraic subvarieties of complex codimension k with the normalized $(2n-2k)$-dimensional measure (volume) and the distance induced from the standard ($U(n+1)$-invariant) Fubini–Study metric on $\mathbb{C}\mathrm{P}^n$. (Again, we would prefer a slightly better bound with $\log\kappa$ instead of $\log\left(\frac{k}{n}\kappa\right)$.)

Remarks: For small k/n, these concentrated spaces do not display an essentially new geometry compared to the spheres and/or projective spaces, since these X are $\underline{\Box}_1$-close to spheres and to the complex projective spaces. In fact, the proof above shows that the normalized Riemannian measure on such $X \subset S^n$ (or $\subset \mathbb{C}\mathrm{P}^n$) is only slightly farther than $\sqrt{k/n}$ from the ambient spherical (Fubini–Study in $\mathbb{C}\mathrm{P}^n$) measure for our di_1-metric on measures. (One may observe here the obvious continuity of the function $X \mapsto \mathrm{ObsDiam}(X, -\kappa)$ with respect to the metric $\underline{\Box}_1$.) Furthermore, our argument suggests the following "ε-order" on measures on a metric space S

$$\mu \geq_\varepsilon \nu \Leftrightarrow \mu(U_\varepsilon(Y) \geq b(\varepsilon)\nu(Y)$$

for all $Y \subset S$ and a given function $b(\varepsilon)$ specifying this "order."

We suggest that the reader pursue this train of thought by studying the formal properties of the ε-order and by bringing forth further examples.

Question: When can one strengthen the above inequalities, e.g., (+), by allowing the induced *path* metric in subsets X of S^n (or $\mathbb{C}\mathrm{P}^n$) rather than the induced distance function?

Examples: Let X be the union of two $(n-1)$-spheres in S^n. Since these S_1^{n-1} and S_2^{n-1} meet along S^{n-2} whose small neighborhoods in S_1^{n-1} and S_2^{n-1} have nearly the full volume, this $X = S_1^{n-1} \cup S_2^{n-1}$ has essentially the same concentration as a single S^{n-1}. Similarly, the union X of d equatorial spheres of codimension k in S^n is highly concentrated for large n if k and d are small compared to n. What is somewhat more amusing, the concentration remains high even if $d \to \infty$ as fast as we want (with k bounded). To see this from another angle, let X be obtained by attaching d copies of the hemisphere S_+^n to S^n along equators S_i^{n-1} in S^n identified with ∂S_+^n. Then $\mathrm{ObsDiam}\, X \leq 1/\sqrt{n}$ *independently* of d, where we can use the obvious path metric and the normalized measure in X. In fact, every subset $X_0 \subset X$ of measure κ contains at least κ's part of measure of some of S_{+i}^n. Then the ε-neighborhood $U_\varepsilon(X_0)$ bites away about κ relative measure of the boundary (equator) S_i^{n-1}, and then $U_{2\varepsilon}(X_0) = U_\varepsilon(U_\varepsilon(X_0))$ κ-penetrates into the central sphere S^n. It follows that $U_{3\varepsilon}(X_0)$ and $U_{3\varepsilon}(X_1)$ necessarily meet in S^n (and also in each S_{+i}^n) if κ is not terribly small.

Corollary to the proof: *Finite spherical buildings X of dimension n are concentrated at least as strongly as S^n, thus having*

$$\mathrm{ObsDiam}(X, -\kappa) \leq \frac{5}{\sqrt{n}}\sqrt{-\log \kappa}.$$

Here we include the 5 to play it safe.

This may look impressive (especially if you have never heard of Tits' buildings before), but, alas, it reveals precious little new measure geometry compared to the basic Levy theorem.

Exercises: (*a*) Evaluate more carefully the observable diameter of the union of d spheres in S^n,

$$X = S_1^{n-k_1} \cup S_2^{n-k_2} \cup \cdots \cup S_d^{n-k_d} \subset S^n.$$

(*b*) Do the same for the union of planes $\mathbb{C}\mathrm{P}_i^{n-k_i}$ in $\mathbb{C}\mathrm{P}^n$ and for quaternionic planes in $\mathbb{H}\mathrm{P}^n$.

(*c*) Estimate the observable diameter of the product of spheres, say $S^m \times S^m \times \cdots S^m$ for large m and small numbers of factors by embedding this product into S^{km+k-1}. Do the same for the Stiefel manifold of orthogonal k frames in $\mathbb{R}^n$.

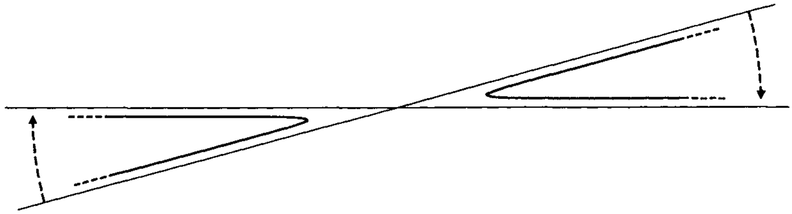
(*d*) Estimate the observable diameter of the Grassmannian space of k planes in $\mathbb{R}^n$ via the natural fibration $\mathrm{St}_k \to \mathrm{Gr}_k$ giving the Lipschitz majorization $\mathrm{Gr}_k \prec \mathrm{St}_k$.

Remarks: The estimates suggested in (*b*) and (*d*) appear in $[\mathrm{Milm}]_{APF}$, $[\mathrm{Milm}]_{HPL}$, and [Mil–Sch]. These extend to more general homogeneous submanifolds $X \subset S^n$ of *low codimension* and also to some non homogeneous submanifolds X in S^n with bounded relative curvatures (i.e., the second fundamental form) such as *isoparametric hypersurfaces* dividing S^n into equal halves. But for many compact symmetric spaces X, better estimates follow from the generalized Levy's isoperimetric inequality (depending on $\inf \mathrm{Ricci}\, X$, see $[\mathrm{Gro–Mil}]_{\mathrm{GSII}}$). Probably asymptotically sharp bounds on $\mathrm{ObsDiam}\, X$ for symmetric spaces of rank $r \to \infty$ *cannot* be derived from similar inequalities for S^n, since one can expect a lower bound on $\underline{\Box}_1(X_1, X_2)$ for symmetric spaces X_1 and X_2 in terms of $\mathrm{rank}\, X_1 / \mathrm{rank}\, X_2$. One also does not think of a good(?) majorization $X_1 \prec X_2$ for $\mathrm{rank}\, X_1 >> \mathrm{rank}\, X_2$. Nevertheless, one hopes to eventually find asymptotically sharp values of $\mathrm{ObsDiam}\, X^n$, $n = 1, 2, \ldots$, for the classical sequence of compact symmetric spaces as well as for the isoperimetric profiles of these spaces. One may even aspire to do the same for all compact homogeneous spaces, but the only known systematic approach to these problems comes from spectral geometry (see $3\frac{1}{2}$.G).

$3\frac{1}{2}$.29. Concentration for $(X, \mathrm{DIST}) \subset \mathbb{C}\mathrm{P}^k$. Now we turn to (the concentration problem for) complex algebraic varieties X in $\mathbb{C}\mathrm{P}^n$ of low codimension k and degree d. We take the induced *path metric* DIST on X and the normalized Riemannian volume. This DIST may be significantly greater than the induced distance function $\mathrm{dist} = \mathrm{dist}_{\mathbb{C}\mathrm{P}^n}|_X$. In fact, the ratio DIST / dist blows up for some degenerating families of algebraic varieties, e.g., for the quadrics $X_\varepsilon = y(y + \varepsilon x) = 1$ as $\varepsilon \to 0$, in the plane $\mathbb{C}^2 \subset \mathbb{C}\mathrm{P}^2$.

However, we shall see below that this blowup is (essentially) limited to a subset of relatively small measure in X. Consequently, we shall prove in (G‴) that $\mathrm{ObsDiam}(X; -\kappa)_{\mathrm{DIST}}$ *decays at least as fast as* $(\log(n)/n)^{1/2d}$ *for* $n \to \infty$ *if* d, k, *and* κ *are kept fixed or grow much slower than* n.

Our analysis of the relevant geometry of X depends on locating a subvariety $\Sigma \subset X$ of codimension $d - 1$ such that



(i) Almost all measure of X lies ε-close to Σ with respect to DIST, where $\varepsilon \to 0$ as $n \to \infty$;

(ii) The metric DIST near Σ can be bounded in terms of dist in a manner independent of $m = \dim X$ (but dependent on d and k).

This Σ appears as a certain singularity of a suitable projection of X from $\mathbb{C}\mathrm{P}^n$ to a subspace $\mathbb{C}\mathrm{P}^m$ in $\mathbb{C}\mathrm{P}^n$ as follows.

(A) Projections $X \to \mathbb{C}\mathrm{P}^m$ and the loci $\Sigma \subset X$ of maximal multiplicity. Consider two mutually *opposite* subspaces in $\mathbb{C}\mathrm{P}^n$ of dimensions m and $k - 1 = n - m - 1$, where "opposite" corresponds to orthogonality between the corresponding linear subspaces of dimensions $m + 1$ and k in $\mathbb{C}^{n+1}$. Then we *project* $\mathbb{C}\mathrm{P}^n$ to our $\mathbb{C}\mathrm{P}^m$ from the opposite $\mathbb{C}\mathrm{P}^k$, called the *center* of the projection. This means that every $x \in \mathbb{C}\mathrm{P}^n$ goes to the intersection of $\mathbb{C}\mathrm{P}^m$ with the k-plane P in $\mathbb{C}\mathrm{P}^n$ passing through $\mathbb{C}\mathrm{P}^{k-1}$ and x. This projection is well-defined and regular on the complement $\mathbb{C}\mathrm{P}^n \backslash \mathbb{C}\mathrm{P}^{k-1}$; we shall apply it to an m dimensional subvariety $X \subset \mathbb{C}\mathrm{P}^n$ which misses $\mathbb{C}\mathrm{P}^{k-1}$.

Bézout theorem. *The projection $p \colon X \to \mathbb{C}\mathrm{P}^m$ is finite-to-one. In fact, there is an integer d such that* $\mathrm{card}(p^{-1}(y)) = d$ *for $y \in \mathbb{C}\mathrm{P}^m \backslash \underline{\Sigma}^1$, where $\underline{\Sigma}^1$ is a hypersurface in $\mathbb{C}\mathrm{P}^m$ (which is empty for $d = 1$). This d is independent of the choice of the $(k-1)$-plane $\mathbb{C}\mathrm{P}^{k-1} \subset \mathbb{C}\mathrm{P}^n$ from where we project and is called the degree* $\deg X$.

Remark: Judging by the name, this result was probably first proved by anybody but Bézout. A novice reader may enjoy finding a proof on his/her own.

Local and global degrees. Let $p : X \to Y$ be a holomorphic map between complex analytic varieties, where Y is assumed to be nonsingular. The *(local) degree* of p at a point $x \in X$ is defined as the integer d_x with the following property. There exist neighborhoods $U_x \subset X$ of x and $\underline{U}_* \subset Y$ of $p(x) \in Y$ and a complex hypersurface $\underline{\Sigma}_x^1 \subset Y$ such that

$$\mathrm{card}(p^{-1}(y) \cap U_x) = d_x$$

for all $y \in \underline{U}_* \setminus \underline{\Sigma}_x^1$.

The existence of these $U_x, \underline{U}_x$, and $\underline{\Sigma}_x^1$ follows by an easy (and standard) variation of the Bezout tune (where we observe that one can choose $\underline{\Sigma}_x^1$ independent of x in our case where X, Y, and p are algebraic). Furthermore, *the global degree d of our $p\colon X \to Y = \mathbb{C}\mathrm{P}^m$ satisfies*

$$d = \sum_{x \in p^{-1}(y)} d_x \tag{$*$}$$

for all $y \in \mathbb{C}\mathrm{P}^m$. In particular, $d_x \leq d$ for all $x \in X$.

Actually, this is true ("well-known and easy to prove," as the poor reader might have learned by now) for arbitrary proper holomorphic maps (and even for continuous maps with suitable definitions of d and d_x). And $(*)$ obviously implies the following

(A′) One-to-One Corollary: *Let $\Sigma \subset X$ consist of those $x \in X$ where $d_x = d$ and set $\underline{\Sigma} = p(\Sigma) \subset \mathbb{C}\mathrm{P}^m$. Then p is one-to-one over $\underline{\Sigma}$, i.e.,* $\mathrm{card}(p^{-1}(y)) = 1$ *for all $y \in \underline{\Sigma}$.*

Now we ask ourselves: how large are Σ and $\underline{\Sigma}$? Here is our

Answer: These $\Sigma \subset X$ and $\underline{\Sigma} \subset \mathbb{C}\mathrm{P}^m$ are algebraic subvarieties of codimension $\leq k(d-1)$.

Proof. The fiber $p^{-1}(y)$, $y \in \mathbb{C}\mathrm{P}^m$, equals the intersection of X with the k-plane P_y in $\mathbb{C}\mathrm{P}^n$ passing through y and containing the center $\mathbb{C}\mathrm{P}^{k-1}$. This intersection $X \cap P_y$ consists of d points $x_i = x_i(y)$ if we count with multiplicities, and y belongs to $\underline{\Sigma}$ if and only if all d points collide, i.e., $\mathrm{card}(X \cap P_y) = 1$. This collision can be expressed by $d-1$ equations

$$x_1(y) = x_2(y) = \cdots = x_d(y),$$

where each of these represents a system of k numerical equations as x_i sum over $P_y = \mathbb{C}\mathrm{P}^k$. So, we have $k(d-1)$ equations altogether, which *suggests* the equalities $\mathrm{codim}\,\Sigma = \mathrm{codim}\,\underline{\Sigma} = k(d-1)$.

Now, to give an actual proof, we observe that our projection $\mathbb{C}\mathrm{P}^n \setminus \mathbb{C}\mathrm{P}^{k-1} \to \mathbb{C}\mathrm{P}^m$ makes a vector bundle over $\mathbb{C}\mathrm{P}^m$ with fiber $P_y \setminus \mathbb{C}\mathrm{P}^{k-1}$, $y \in \mathbb{C}\mathrm{P}^m$. This vector bundle (obviously) splits into a sum of k line bundles, each being isomorphic to the normal bundle of $\mathbb{C}\mathrm{P}^m \subset \mathbb{C}\mathrm{P}^{m+1}$. First let $k = 1$, where we are left with a single line bundle $L \to \mathbb{C}\mathrm{P}^m$. The equations $x_1 = x_2 = \cdots = x_d$ for the *symmetric* configuration of points $\{x_i = x_i(y)\}$ in each fiber $L_y \subset L$ can be rewritten in terms of the symmetric functions as follows

$$\sum_{i=1}^{d} x_i^r = \frac{1}{d^{r-1}} \left(\sum_{i=1}^{d} x_i \right)^r,$$

$r = 2, \ldots, d$. Here the first equation with $r = 2$ refers to the tensor square $L^2 = L \otimes L$, the second to $L^3 = L \otimes L \otimes L$, etc., where we use the canonical maps $L \to L^r$ sending each x to $x^r \in L^r$. Thus, the solutions of the r-th equation correspond to the zeros $Z_r \subset \mathbb{C}\mathrm{P}^m$ of the section $s_r = \sum_{i=1}^r x_i^r - \left(\sum_{i=1}^d x_i\right)^r / d^{r-1}$ of L^r. Every such Z_r is a hypersurface in $\mathbb{C}\mathrm{P}^m$ of degree r (unless it coincides with $\mathbb{C}\mathrm{P}^m$), and so their intersection

$$\underline{\Sigma} = Z_2 \cap Z_3 \cap \cdots \cap Z_d \subset \mathbb{C}\mathrm{P}^m$$

has codimension $\leq d-1$ (and degree $d!$). Thus, the case $k = 1$ is concluded.

Finally, let $k \geq 2$ and look at $x_1, \ldots, x_d$ in the sum of k line bundles, say in $L_1 \oplus \cdots \oplus L_j \oplus \cdots \oplus L_k$. Then, clearly, $\underline{\Sigma}$ *contains* the intersection

$$\bigcap_{\substack{r=2,\ldots,d \\ j=1,\ldots,k}} Z_{r,j},$$

and so both $\underline{\Sigma}$ and Σ have codimension $\leq k(d-1)$ (where $\deg \underline{\Sigma} \leq (d!)^k$ and $\deg \Sigma \leq d(d!)^k$).

Remarks: (a) One can think of $\Sigma \subset X$ as a cycle representing the dual to a certain characteristic class $\delta_d(X) \in H^*(X)$, at least when X is smooth. Actually, this δ_d must lie in $H^{2m-2d+2}(X)$ and whenever $\delta_d \neq 0$, every holomorphic map $p : X \to \mathbb{C}\mathrm{P}^m$ of degree d will have $\operatorname{codim} \Sigma \geq d-1$ (where Σ is defined as the set of points $x \in X$ where $d_x = d$). This suggests that the codimension of our Σ and $\underline{\Sigma}$ for $X = \mathbb{C}\mathrm{P}^n$ must be $\leq d-1$ for all codimensions $k = 1, 2, \ldots$. This is not true, however, since our bound codim $\leq k(d-1)$ is sharp for X equal to the union of several m planes in $\mathbb{C}\mathrm{P}^n$ meeting transversally. But it is true for all smooth X and for complete intersections. In fact, every holomorphic map p of a normal (e.g., nonsingular) variety X into $\mathbb{C}\mathrm{P}^n$ has d points coming together along a subvariety of codimension $\leq d$ for $d = \deg_{\mathrm{top}} p$ (see [Ga–La]; this was pointed out to me by Fedia Zak).

(b) One can easily arrange a hypersurface $X \in \mathbb{C}\mathrm{P}^{m+1}$ where $\operatorname{codim} \Sigma < d-1$; just take $d \geq 3$ hyperplanes meeting along an $(m-2)$-plane. One can also produce examples with nonsingular X and special p, such as a plane cubic curve X projected from a center lying on a line tangent to X at an inflection point in X. We suggest working out further examples to the reader, who may also try to figure out if $\operatorname{codim} \Sigma = d-1$ for *every nonsingular* hypersurface X and a *generic* projection p (which seems easy to me). For example, if $X \subset \mathbb{C}\mathrm{P}^{m+1}$ is a nonsingular quadric, then $\underline{\Sigma} \subset \mathbb{C}^m$

is also a nonsingular quadric hypersurface for every center $c_0 = \mathbb{C}\mathrm{P}^0 \subset \mathbb{C}\mathrm{P}^{m+1} \setminus X$. And $p\colon X \to \mathbb{C}\mathrm{P}^m$ is a ramified double cover which branches exactly along $\underline{\Sigma}$, while $\Sigma = p^{-1}(\underline{\Sigma})$ equals the singular set of p where Jacobian$(p) = 0$.

(c) Persistence (or disappearance) of our Σ with $\operatorname{codim}\Sigma = d-1$ implies nonvanishing (or vanishing) of the corresponding characteristic class $\delta(X) \in H^{2m-2d+2}(X)$ under suitable algebraic geometric assumptions. The key point is the *non-vanishing* of the $2r$-dimensional homology classes in $H_{2r}(X;\mathbb{R})$ represented by r-dimensional algebraic subvarieties Y in X (with $r = d-1$ in the case of $Y = \Sigma$). This nonvanishing is due to the *positivity* of the intersection of each Y with a generic $(N-r)$-plane P in some ambient projective space $\mathbb{C}\mathrm{P}^N \supset X \supset Y$, which makes $[Y]$ nonzero in $H_{2r}(\mathbb{C}\mathrm{P}^N)$ and thus in $H_{2r}(X)$. (Notice that complex (sub)varieties come along with natural orientations, and so intersection indices are well-defined. Also, the intersection with P gives us the degree of Y by the Bezout theorem.)

Here is another (well-known) instance of this phenomenon, where the topology of a map is strongly influenced by its algebraic geometric origin.

Let p be a surjective holomorphic map between equidimensional complex algebraic manifolds, say $p: X \to A$, (where, actually, one can admit an arbitrary complex manifold for A). If $p^*: H^2(A) \to H^2(X)$ sends the first Chern class $c_1(A)$ to $c_1(X)$, then p is a covering map. In particular, if A is simply connected, then p is an isomorphism.

In fact, the Jacobian (determinant) of a noncovering map p must vanish on a subvariety $\Sigma_0 \subset X$ with $\operatorname{codim}\Sigma_0 \leq 1$, and the cohomological dual to $[\Sigma_0]$ corresponds to the difference $p^*(c_1(Y)) - c_1(X)$.

One can say slightly more if the receiving space A is complex parallelizable, i.e., admits a full frame of holomorphic vector fields, or, equivalently, of holomorphic 1-forms. In this case, the condition $c_1(X) = 0 (= c_1(A))$ implies that a holomorphic map $p: X \to A$, which is not assumed to be surjective, necessarily has *constant rank*, thus decomposing into a holomorphic fibration $X \to \underline{X}$ followed by an immersion $\underline{X} \hookrightarrow A$, where this immersion is *parallel* with respect to the parallelization of A. This means that the isomorphism between the tangent spaces of A given by the frames sends $T_a(\underline{X}) \to T_b(\underline{X})$ for all $a, b \in X$. We leave the proof of this to the reader, and we also indicate the following classical corollary.

Let X be an algebraic manifold with $c_1(X) = 0$ and $\operatorname{rank} H^1(X) \geq 2m$ *for $m = \dim_{\mathbb{C}} X$. Then X is isomorphic to a complex torus, i.e., $\mathbb{C}^{2m}/$*lattice.

Proof. One knows that X admits a holomorphic map to some torus $p\colon X \to A = \mathbb{C}^{2k}/\text{lattice}$ for $2k = \operatorname{rank} H^1(X)$ (where A is called the *Albanese variety* and p is the *Abel–Jacobi–Albanese map*), such that $p^*\ H^1(A) = \mathbb{Z}^{2k} \to H^1(X)$ is an isomorphism. This p factors as $X \to \underline{X} \hookrightarrow A$ with an immersed *subtorus* $\underline{X} \hookrightarrow A$ which must necessarily have real dimension $2k$ to make $p*$ an isomorphism. Since $2k \geq 2m$, we are left only with the possibility of $2k = 2m$ and $p\colon X \to A$ being a finite covering map.

(d) The above can be sharpened with the notion of *positivity* on cohomology $H^2(X)$. For example, the conclusion that X is a torus holds whenever $\operatorname{rank} H^1(X) \geq 2 \dim X$ and $-c_1(X)$ is *non positive*, i.e., there is *no* complex algebraic hypersurface $Y \subset X$ with the class $[Y] \in H_{2m-2}(X)$ being dual to a *negative* multiple of $c_1(X)$. On the other hand, there are deeper relations between Chern classes, such as the Bogomolov–Miyoaka–Yau inequality $c_2 \geq 3c_1^2$ for surfaces of general type, or Yau's theorem: $c_1 = c_2 = 0 \Rightarrow X = \text{torus}$, which go beyond mere positivity.

(B) Controlling the center $\mathbb{C}\mathrm{P}^k \subset \mathbb{C}\mathrm{P}^n \setminus X$. We now plunge into the muddy waters of mixed algebraic and metric geometry, and we invite the courageous reader to swim along through a dozen pages until the finish in (G‴).

Our first objective is to position $\mathbb{C}\mathrm{P}^k$ far from X in order to protect the projection $p\colon X \to \mathbb{C}\mathrm{P}^m$ from exploding. Here is the relevant

(B′) Lemma: *For every $\ell < \operatorname{codim} X$, there exists an ℓ-plane $\mathbb{C}\mathrm{P}^\ell \subset \mathbb{C}\mathrm{P}^n$ lying ε-far from X (i.e., in the complement of the ε-neighborhood of X) for $\varepsilon \geq \varepsilon_{\ell,d} \geq (10d)^{-(\ell+1)}$ and $d = \deg X$.*

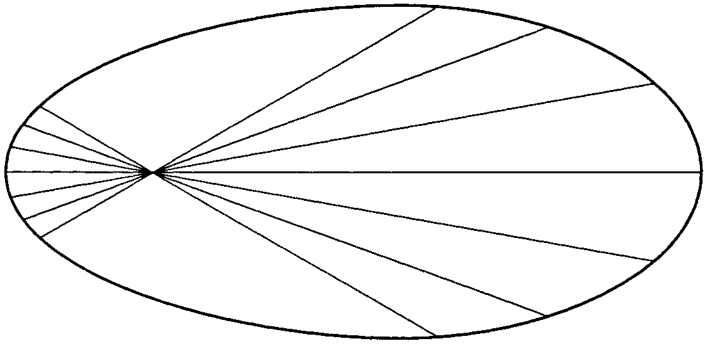
Proof. Everything is based on the

Bernstein inequality. *Let $Q(x)$ be a complex polynomial of degree d on $\mathbb{C}^n$. Then the supremum of the gradient of Q over an R ball $B \subset \mathbb{C}^n$ satisfies*

$$\sup_{x \in B} \| \operatorname{grad} Q(x) \| \leq 3d \sup_{x \in B} |Q(x)|/R. \tag{+}$$

This is classical for $n = 1$ and without any 3. Then the case $n \geq 2$ follows by restricting Q to suitable disks in B (which bring along this unfortunate, and probably unnecessary 3). In fact, a similar inequality holds on every (convex) subset $B \subset \mathbb{C}^n$ which contains "sufficiently many" "large" disks through every point.

And, (+) holds without 3 on the polydisk $D(R) \times D(R) \times \cdots$ (n factors) as is seen by applying the 1-dimensional Bernstein lemma to circles in the



n torus $S^1(R) \times S^1(R) \times \cdots \times S^1(R)$, making the *Shilov boundary* of the polydisk.

(B″) Corollary: *There is a point $x \in B$ which is ε-far from the zero set Z of Q in B for some $\varepsilon \geq R/3d$, unless Q is identically zero.*

Proof. If the zero set Z is ε-dense in B, then clearly,

$$\sup_{x \in B} |Q(x)| \leq \varepsilon \sup_{x \in B} \| \operatorname{grad} Q \|,$$

which makes $\sup_{x \in B} |Q(x)| = 0$ for $\varepsilon < R/3d$, as follows from (+).

Remarks: (a) The above may stay true with $\varepsilon \geq R/c\sqrt{d}$ (say for $c = 1$). Of course, this is easy to prove with $\varepsilon \geq R/c_n\sqrt{d}$ by evaluating the volume of the ε-neighborhood of $Z \subset B$; the whole point, however, is to make the constants independent of the dimension! Actually the dimension free bound $\varepsilon \geq R/c\sqrt{d}$ seems nontrivial even when Z is the union of d complex hyperplanes in $\mathbb{C}$ intersected with B. (This special case gains in significance if we look at d hyperplanes in the Hilbert space $\mathbb{C}^\infty$ and ask how dense they can be in $B \subset \mathbb{C}^\infty$).

(b) Real polynomials Q also admit bounds on $\| \operatorname{grad} Q \|$ by $|Q|$, but now with the constant d^2 (rather than d) according to the *Markov inequality.* This gives a version of (B″) with $\varepsilon \geq R/cd^2$ for real algebraic hypersurfaces in $B = B(R)$ instead of the expected $\varepsilon \geq R/cd$. Getting this R/cd bound does not appear hard, since Markov's inequality is linear in d (deeply) inside B. In fact, every real algebraic subset Z in B of degree d and codimension k probably admits an ε-far point, with $\varepsilon \geq R/cd^{1/k}$ (where $c = 1$ is a possibility). One can slightly vary this conjecture by using different notions of degree. I would prefer the projective degree of the complexification of Z. But for the moment, I did not check this conjecture, even for unions of d

planes intersected with B, where the case of real hyperplanes is relatively easy (see [Bang]). (One may ask in this regard how dense a union of linear subspaces and other simple subvarieties of finite codimension, such as spheres, can be in a given Banach space.) On the other hand, everything simplifies if one allows c to depend on the dimension (how boring!), where the bound with $\varepsilon \geq R/c_n d^{1/k}$ is achieved by intersecting A with the standard grid of ε-cubes and applying the Petrovski–Thom–Milnor inequality (compare [Iva] and $[\mathrm{Gro}]_{\mathrm{Yom}}$).

Finally, we point out a hard generalization of the nondensity properties of Z above to the nodal sets of eigenfunctions of differential operators, due to Donnely and Feferman.

Proof of (B′). If $k = \operatorname{codim} X = 1$ and $\ell = 0$, then everything follows directly from (B'') applied to suitable balls in affine charts in $\mathbb{C}\mathrm{P}^n$. Then we also get the case $\ell = 0$ and *arbitrary* k, since every subvariety X of degree d is contained in a hypersurface X^* of degree d. Namely, one may take the cone X^* over X from some $(k-2)$-dimensional center, i.e., the pullback of the projection of X to some $\mathbb{C}\mathrm{P}^{m+1}$, $m = \dim X$. Finally, one handles $\ell = 1, 2, \ldots$, arguing by induction on ℓ as follows. Take a point $x_0 \in \mathbb{C}\mathrm{P}^0$ ε_0 far from X with $\varepsilon_0 \approx d^{-1}$, and project X to the opposite $\mathbb{C}\mathrm{P}^{n-1}$. The image $X' \subset \mathbb{C}\mathrm{P}^{n-1}$ has degree d and so admits by the inductive hypothesis an $(\ell-1)$-plane $\mathbb{C}\mathrm{P}^{\ell-1}$ lying ε'-far from it with $\varepsilon' \geq (10d)^{-(\ell+1)}$. Then the cone $\mathbb{C}\mathrm{P}^{\ell}$ from x_0 over $\mathbb{C}\mathrm{P}^{\ell-1}$ lies as far from X as required if the implied constant in the relation $\varepsilon \approx d^{-1}$ was suitably chosen (where the reader is left with the task of bringing the constants into line).

Questions: (a) What is the true dependence of $\varepsilon_{\ell,d}$ on ℓ and d? Is $\varepsilon_{\ell,d} \approx 1/d^{1/2(k-\ell)}$?

(b) What happens in the real algebraic case? What can one gain by looking for subvarieties of degree $\delta \geq 2$ far from X instead of planes (where $\delta = 1$)?

(C) Bound on $\underline{\text{dist}}$ by dist. We denote by dist the standard ($U(n+1)$-invariant) metric on $\mathbb{C}\mathrm{P}^n$ and by $\underline{\text{dist}}$ the metric on $\mathbb{C}\mathrm{P}^m$ where we are interested in the behavior of the (degenerate) metrics on $X \subset \mathbb{C}\mathrm{P}^n$, where $\underline{\text{dist}}$ comes to X via the projection p. Since the center of p is ε-far from X, we trivially have

$$\underline{\mathrm{dist}}|_X \leq c_\varepsilon \operatorname{dist}|_X,$$

say for $c_\varepsilon = 2\varepsilon^{-1}$. (What is more interesting, one can also bound $\operatorname{dist}(x, x')$ in terms of $\underline{\text{dist}}(p(x), p(x'))$ for all $x \in X$ and $x' \in \Sigma \subset X$, independently of m, as will become clear later on.)

(D) Comparison between $\underline{\text{dist}}$ and $\underline{\text{DIST}}$. We denote by $\underline{\text{DIST}}$ the path metric on X induced from $\underline{\text{dist}}$ by our $p\colon X \to \mathbb{C}\mathrm{P}^m$. This is indeed a true metric, since the map p is finite-to-one, as follows from the fact that the center of the projection does not meet X. It is obvious that

$$\underline{\text{dist}} \leq \underline{\text{DIST}},$$

where $\underline{\text{dist}}(x,x') \stackrel{\text{def}}{=} \underline{\text{dist}}(p(x),p(x'))$. Somewhat surprisingly, we also have the following

(D′) Partial equality between $\underline{\text{DIST}}$ and $\underline{\text{dist}}$. *We have*

$$\underline{\text{DIST}}(x,x') = \underline{\text{dist}}(x,x')$$

for all $x \in X$ and $x' \in \Sigma = \Sigma(p)$.

Proof. Join $p(x)$ with $p(x')$ by the shortest geodesic arc a in $\mathbb{C}\mathrm{P}^m$ and observe that there is a *covering arc* A in X issuing from x. Indeed, our arc a lies in some complex line $\mathbb{C}\mathrm{P}^1 \subset \mathbb{C}\mathrm{P}^n$, and the covering A should be constructed in the algebraic curve $X^1 = X \cap p^{-1}(\mathbb{C}\mathrm{P}^1)$, which goes to $\mathbb{C}\mathrm{P}^1$ by the map $p|_{X^1} : X^1 \to \mathbb{C}\mathrm{P}^1$. Since $\dim_{\mathbb{C}} X^1 = 1$ and since p is finite-to-one, the map $p|_{X^1} : X^1 \to \mathbb{C}\mathrm{P}^1$ is a covering away from finitely many ramification points, where p is locally modeled by $z \mapsto z^r$ in the complex plane $\mathbb{C}$ (with possible values $r = 2,3,\dots,d$), provided that the curve X^1 is irreducible. If it is reducible, then this (ramification) behavior refers to each irreducible branch of X^1. It is obvious that the map $z \mapsto z^r$ has the path lifting property (at least for piece-wise smooth paths in $\mathbb{C}$), which yields the desired lifts for $X^1 \to \mathbb{C}\mathrm{P}^1$.

The path A starting from x terminates at some point in X over $p(x')$, and since p is one-to-one over $\Sigma \ni x'$, the path A terminates at x', as this is the only point over $p(x')$. It follows that $\underline{\text{DIST}}(x,x') \leq \text{length}(A) = \text{length}(a) = \underline{\text{dist}}(x,x')$.

Remarks. (a) The path lifting property is shared by all proper surjective holomorphic maps between irreducible complex analytic varieties. In fact, we can lift (real analytic) arcs also for real analytic maps which have *no folds* as a simple (and well-known, I guess) argument shows. The reader is invited to think the matter through for him/herself.
(b) If Σ is nonempty, then the above implies that

$$\text{diam}(X, \underline{\text{DIST}}) \leq 2\,\text{diam}(\mathbb{C}\mathrm{P}^m) = 2\pi.$$

(This was pointed out to me by Carlos Simpson.)

(E) Bound on DIST **by** $\underline{\text{DIST}}$**.** Here DIST stands for the induced path metric in $X \subset \mathbb{C}\mathrm{P}^n$ which obviously bounds $\underline{\text{DIST}}$ in terms of our $\varepsilon = \varepsilon_{k-1,d} > 0$ by

$$\underline{\text{DIST}} \leq 2\varepsilon^{-1}\,\text{DIST}\,.$$

It follows by a simple compactness argument that

$$\text{DIST} \leq \rho(\underline{\text{DIST}}) \tag{$*$}$$

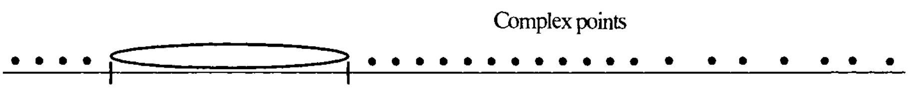
for some function $\rho = \rho_{d,k,m}$ satisfying $\rho(\delta) \to 0$ for $\delta \to 0$. Since $\underline{\text{DIST}}$ is determined by (the length of) arcs contained in projective lines in $\mathbb{C}\mathrm{P}^m$, we can bound $\rho_{d,k,m} \leq \rho_{d,k,t} = \rho_{d,k}$, thus making $(*)$ independent of m.

Here is a more specific (rather rough) bound on ρ.

$$\rho(\delta) \leq \text{const}\, d^{2k+2}\delta^{1/d} \tag{$\star$}$$

for some const $\leq 1,000,000$.

Idea of the proof. We must show that paths do not become much longer as they are lifted from $\mathbb{C}\mathrm{P}^m$ to $X \subset \mathbb{C}\mathrm{P}^n$ for the projection $p\colon X \to \mathbb{C}\mathrm{P}^m$. As we mentioned before, this lift takes place over complex projective lines in $\mathbb{C}\mathrm{P}^m$, and so we may assume that $m = 1$ to start with. If a δ-short arc a in $\mathbb{C}\mathrm{P}^1$ becomes long in X, where we assume for the moment that $n = 2$, then X lies δ-close to some projective line $L \subset \mathbb{C}\mathrm{P}^2$ issuing from the center (which is ε-far away from X). In fact, X and L are mutually close along some real algebraic arc A of degree d and length ρ, and by an interpolation argument (see below), L must be everywhere close to X if ρ is large compared to ε. Notice that this is a purely *complex* algebraic phenomenon: an ellipse can be close to a line along a segment without approaching this line outside this segment in the *real* domain.



This happens because a *real* polynomial may be small at some point without vanishing near this point. In the complex domain, however, we have the following standard

(E′) Vanishing lemma: *Let Q be a nonconstant polynomial of degree d on the R-ball $B = B(x_0, R) \subset \mathbb{C}^n$ with* $\sup_{x\in B}|Q(x)| \geq b$*. If* $|Q(x_0)| \leq \delta_0$

at the center x_0 of B, then the distance $\varepsilon = \mathrm{dist}(x_0, Z)$ from x_0 to the zero set Z of Q is bounded by

$$\frac{\varepsilon}{1 + R^{-1}\varepsilon} \leq R\left(\frac{\delta_0}{b}\right)^{1/d}.$$

Proof. Everything trivially reduces to the case $n = 1$, $x_0 = 0 \in \mathbb{C}$, $R = 1$, and where the polynomial $Q(x)$ decomposes as the product over its roots $z_i \in \mathbb{C}$. Thus,

$$\left|\frac{b}{\delta_0}\right| = \sup_{|x|\leq 1}\left|\prod_{i=1}^{d}\left(1 - \frac{x}{z_i}\right)\right| \leq \left(1 + \frac{1}{\delta}\right)^d,$$

for $\delta = \min_{i=1,\dots,d}|z_i|$.

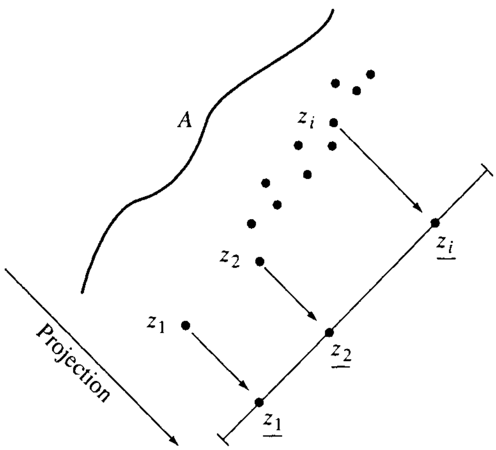
Next, we need another elementary and well-known

(E″) Interpolation lemma: *Let $A \subset B = B(x_0, R) \subset \mathbb{C}$ be a connected set (arc) of diameter $\geq \rho_0$. Then*

$$b \overset{\text{def}}{=} \sup_{x\in B}|Q(x)| \leq b_0(CR/\rho_0)^d$$

for $b_0 \overset{\text{def}}{=} \sup_{x\in A}|Q(x)|$, where $Q(x)$ is as before an arbitrary polynomial on $\mathbb{C}$ of degree $\leq d$ and C is a numerical constant ≤ 20.

Proof. Assume $B = B(0, 1)$ and $Q(x) = \prod_{i=1}^{d}(x - z_i)$. Then normally project A *onto* some segment $I \subset B$ of length ρ_0 and integrate $\sum_{i=1}^{d} \log|\underline{x} - \underline{z}_i|$ over $\underline{x} \in I$, where $\underline{z}_i \in I$ are the projections of z_i.



Since $\log x$ is integrable at zero, we have

$$\int_I \sum_{i=1}^{d} \log |\underline{x} - \underline{z}_i| \, d\underline{x} \leq d \,\mathrm{const},$$

and consequently,

$$\sup_{x \in A} \prod_{i=1}^{d} |x - z_i| \geq (\rho_0 / \,\mathrm{const})^d.$$

If $|z_i| \leq 2$ for all $i = 1, \ldots, d$, then obviously $\sup_{x \in B} |Q(x)| \leq 4^d$ and the proof follows. Otherwise, apply the above to the partial product $Q_\bullet$ as the remaining factor $Q^\bullet = Q/Q_\bullet = \prod_{|z_i|>2}(x - z_i)$ obviously satisfies

$$\sup_{x \in B} |Q^\bullet(x)| \leq 4^d \inf_{x \in A} |Q^\bullet(x)|.$$

Proof of ($\star$) in (E). Start with the case $k = 1$ and look at our lifted arc $A \subset X \subset \mathbb{C}\mathrm{P}^2$ issuing from a point $x_0 \in X$ and projecting to a δ-segment in $\mathbb{C}\mathrm{P}^1$. We shall work in a suitable affine chart $\mathbb{C}^2 \subset \mathbb{C}\mathrm{P}^2$ with the origin at the center of the projection $c_0 \in \mathbb{C}\mathrm{P}^2$, such that A lies inside the unit ball $B = B(0, 1) \subset \mathbb{C}^2 \subset \mathbb{C}\mathrm{P}^2$, and let $B' = B \cap L$ for the (projecting) line L between c_0 and x_0. We normally project A to L, call the image $A' \subset L$ and observe that $A' \subset B'$ and that the diameter of A' is $\geq \rho_0 \geq \rho/10d$ for $p = \mathrm{length}\, A$. In fact, the length of a piece of a real algebraic curve of degree d in $\mathbb{C}^2 = \mathbb{R}^4$ is bounded by its diameter

$$\mathrm{length} \leq d \,\mathrm{const} \cdot \mathrm{diameter}$$

according to the Buffon–Crofton formula expressing the length as the integrated intersection number of the curve with the hyperplanes in $\mathbb{R}^4$.

We observe that A' lies $(5\varepsilon^{-1}\delta)$-close to X and, therefore,

$$\sup_{x \in A'} |\mathbb{Q}(x)| \leq \mathrm{const} \cdot \varepsilon^{-1} \delta$$

by the Bernstein inequality, where Q is the polynomial defining X in $\mathbb{C}^2$ normalized to have the sup-norm equal 1 on the ball $B \subset \mathbb{C}^2$. Then the interpolation lemma applies to $A' \subset B' \subset L = \mathbb{C}$ and yields the following bound on Q on B',

$$\sup_{x \in B'} |Q(x)| \leq C_1^d d \varepsilon^{-1} \delta / \rho_0^d,$$

and the vanishing bounds the distance ε from c_0 to X by

$$\varepsilon \leq C_2 (d/\varepsilon^{-1}\delta)\varepsilon)^{1/d} / \rho_0.$$

Since $\varepsilon \geq 1/10d$ by (B') applied to $\ell = 0$, we conclude that

$$\rho_0 \leq C_3 d^{1+2/d} \delta^{1/d}$$

and thus

$$\rho \leq C_4 d^{2+2/d} \delta^{1/d}$$

for $C_4 \leq 100000$. This settles $(\star)$ for $k = 1$, and now we turn to the general case of $k \geq 2$. We project X to k mutually perpendicular planes, say $\mathbb{C}P^2_i$, $i = 1, \dots, k$, passing through $\mathbb{C}P^1$, and apply the above to each curve $X_i \subset \mathbb{C}P^2_i$ projected to $\mathbb{C}P^1 = \bigcap_{i=1}^k \mathbb{C}P^2_i$ from the center $c_i = \mathbb{C}P^{k-1} \cap \mathbb{C}P^2_i$, where $\mathbb{C}P^{k-1}$ is the center of the projection $p\colon X \to \mathbb{C}P^1$. Clearly,

$$\varepsilon_i \stackrel{\text{def}}{=} \operatorname{dist}(c_i, X_i) \geq \varepsilon = \operatorname{dist}(\mathbb{C}P^{k-1}, X),$$

and so we bound each DIST_i of X_i in $\mathbb{C}P^2$ by the formula above, with $\varepsilon = (10d)^{-k}$ (see (B)), which makes $\rho_i \leq C_4 d^{((k+1)d+1)/d} \delta^{1/d}$. On the other hand, the metric DIST on X is easily bounded by $\varepsilon^{-1} \sum_{i=1}^k \mathrm{DIST}_i$ on $X \to X_i \subset \mathbb{C}P^2_i$, and so we get the desired bound

$$\rho(\delta) \leq \operatorname{const} d^{2k+1+d^{-1}} \delta^{1/d},$$

which yields $(\star)$.

Remark: The $\delta^{1/d}$-term in $(\star)$ is right, as seen for the map $x \mapsto x^d$, but $d^{2k+\cdots}$ looks excessive. A better estimate may be expected with a suitable *Schwartz lemma* for the embedding $(X, \mathrm{DIST}) \hookrightarrow \mathbb{C}P^n$. Actually, the worst example I see right now is the curve given in affine coordinates $x_0, \dots, x_k$ by the equations $x_i^d = x_0$, $i = 1, \dots, k$, projected by $(x_0, \dots, x_k) \mapsto x_0$, where $\rho(\delta) \approx \sqrt{k}\delta^{1/d}$ near $x_0 = x_1 = \cdots = x_k = 0$.

(F) Partial bound on the path metric DIST **on** $X \subset \mathbb{C}P^n$ **by** $\operatorname{dist}_{\mathbb{C}P^n}$**.** We just combine (B)-(E) and arrive at the estimate

$$\mathrm{DIST}(x, x') \leq C_{k,d} (\operatorname{dist}(x, x'))^{1/d} \qquad (\star\star)$$

for all $x \in X$ and $x' \in \Sigma$, where Σ with $\operatorname{codim} \Sigma = d - 1$ is the maximal multiplicity locus of the projection $X \to \mathbb{C}P^m$, $m = \dim X$, from a suitable (ε-far away from X with $\varepsilon \geq (10d)^{-k}$) center $\mathbb{C}P^{k1} \subset \mathbb{C}P^n$, and

$$C_{k,d} \leq \operatorname{const} d^{3k+2},$$

$\operatorname{const} \leq 10^9$.

(G) Evaluation of $\operatorname{ObsDiam}(X)$ **with respect to** DIST**.** Take an ε-neighborhood $U_\varepsilon(\Sigma) \subset \mathbb{C}P^n$ and show that $U_\varepsilon(\Sigma) \cap X$ contains much of the measure of X (see below). Then the concentration for (X, dist) applied to the part of X in $U_\varepsilon(\Sigma)$, yields via $(\star)$ a similar (albeit weaker) concentration for (X, DIST). Here are the details.

(G′) Lemma: *Let X and $Y \subset X$ be complex algebraic subvarieties in $\mathbb{C}\mathrm{P}^n$ of codimensions k and $\ell \geq k$ correspondingly, with normalized Riemannian measures (volumes). Then*

$$\mu_X(X \cap U_\varepsilon(Y)) \geq 1 - \frac{8}{\varepsilon} e^{-\varepsilon^2 n/72\ell}.$$

Proof. Let κ denote the μ_X-measure of the complement $X_0 = X \setminus U_\varepsilon(Y)$ and observe that the neighborhoods $U_{\varepsilon_0}(X_0)$ and $U_{\varepsilon_1}(Y)$ in $\mathbb{C}\mathrm{P}^n$ may intersect only at their boundaries for $\varepsilon_0 + \varepsilon_1 = \varepsilon$, and so the sum of their measures $\mu = \mu_{\mathbb{C}\mathrm{P}^n}$ does not exceed $1 = \mu(\mathbb{C}\mathrm{P}^n)$. On the other hand, the μ-measures of these neighborhoods can be bounded from below according to $(*)$ in $3\frac{1}{2}.28$ or rather by the $\mathbb{C}\mathrm{P}^n$-version of that inequality. Namely,

$$\mu(U_{\varepsilon_0}(X_0)) \geq b_{\mathbb{C}}(k, \varepsilon_0)\kappa \qquad (*)_{\mathbb{C}}$$

and

$$\mu(U_{\varepsilon_1}(Y) \geq b_{\mathbb{C}}(\ell, \varepsilon_1) \qquad (*)'_{\mathbb{C}}$$

where $b_{\mathbb{C}}(k, \varepsilon_0)$ denotes the measure of the ε-ball in $\mathbb{C}\mathrm{P}^k$ with the normalized measure. These inequalities are rather weak as they stand, but their power is enhanced by the concentration property of $\mathbb{C}\mathrm{P}^n$, which is at least as good as that of S^{2n-1}, which Lipschitz dominates $\mathbb{C}\mathrm{P}^n$ via the Hopf map $S^{2n-1} \to \mathbb{C}\mathrm{P}^n$ (compare [Milm]$_{\mathrm{APF}}$, [Milm]$_{\mathrm{HPL}}$, and [Mil–Sch]). This concentration tells us that the δ-neighborhood of a subset $U \subset \mathbb{C}\mathrm{P}^n$ of measure μ_0 dramatically (even exponentially) increases with δ. This suggests looking at $U_{\varepsilon_0}(X_0)$ as $U_{\varepsilon''_0}(U_{\varepsilon'_0}(X_0))$ for suitable ε'_0 and ε''_0 with $\varepsilon'_0 + \varepsilon''_0 = \varepsilon_0$, and similarly to think of $U_{\varepsilon_1}(Y)$. More technically, we use the bound on the (separation) distance between two subsets U_0 and U_1 of measure $\geq \kappa'$, $\mathrm{dist}(U_0, U_1) \leq d(\kappa') = -\mathrm{ObsDiam}(\mathbb{C}\mathrm{P}^n, -\kappa') \leq \mathrm{ObsDiam}(S^{2n-1}, \kappa')$ (see $3\frac{1}{2}$.F). We apply this to the neighborhoods $U_0 = U_{\varepsilon'_0}(X_0)$ and $U_1 = U_{\varepsilon'_1}(Y)$, which satisfy

$$\mathrm{dist}(U_0, U_1) \geq d' = \varepsilon''_0 + \varepsilon''_1 = \varepsilon - \varepsilon'_0 - \varepsilon'_1,$$

and

$$\begin{aligned} \mu(U_0) \geq \mu_0 &= b_{\mathbb{C}}(k, \varepsilon'_0)\kappa, \\ \mu(U_1) \geq \mu_1 &= b_{\mathbb{C}}(\ell, \varepsilon'_1). \end{aligned}$$

We invoke the rough (and obvious) lower bound

$$b_{\mathbb{C}}(k,\varepsilon) \geq \left(\frac{\varepsilon}{4}\right)^{2k}$$

for $\varepsilon \leq 1$, set $\kappa' = \min((\kappa\varepsilon_0'/4)^{2k}, (\varepsilon_1'/4)^{2\ell})$, and combine the above with the Levy inequality (see $3\frac{1}{2}$.19)

$$d' \leq d(\kappa') \leq \frac{3}{\sqrt{n}}\sqrt{-\log(\kappa'/2)},$$

with $3/\sqrt{n} \geq 2\sqrt{2}/\sqrt{2n-1}$ since we care little for exact numbers). We choose $d' = \varepsilon/2$ and obtain

$$\frac{\varepsilon}{2} \leq \frac{3}{\sqrt{n}} \min\left(\sqrt{-2k\log(\varepsilon_0'/4)\kappa}, \sqrt{-2\ell\log(\varepsilon_1'/4)}\right),$$

where we are free to choose any ε_0' and ε_1' satisfying $\varepsilon_0' + \varepsilon_1' = \varepsilon/2$. Being lazy, we stick to $\varepsilon_0' = \varepsilon_1' = \varepsilon/4$ and conclude that

$$\frac{\varepsilon}{2} \leq \frac{3}{\sqrt{n}}\sqrt{-2\ell\log(\varepsilon\kappa/8)},$$

keeping in mind that $\ell \geq k$ and $\kappa \leq 1$. Thus,

$$\kappa \leq \frac{8}{\varepsilon} e^{-\varepsilon^2 n/72\ell}.$$

(G″) Evaluation of ObsDiam(Σ) **for** DIST$|_\Sigma$**.** We know (see $3\frac{1}{2}$.28) that

$$\mathrm{ObsDiam}(\Sigma, -\kappa)_{\mathrm{dist}} \leq 10\sqrt{\frac{\ell}{n}}\sqrt{-\log\frac{\ell}{n}\kappa}$$

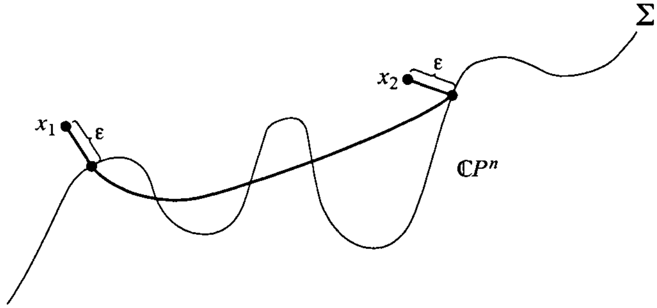
for $\ell = \mathrm{codim}(\Sigma = \mathbb{C}\mathrm{P}^n \leq k + d$ (where $k = \dim X$ and $d = \deg X$). This combines with the ($\star\star$) above and yields

$$\begin{aligned} \mathrm{ObsDiam}(\Sigma)_{\mathrm{DIST}} &\leq C_{k,d}(\mathrm{ObsDiam}(\Sigma)_{\mathrm{dist}})^{1/d} \\ &\leq 10^{11} d^{3k+2}\left(\frac{\ell}{n}\right)^{1/2d} (-\log\frac{\ell}{n}\kappa))^{1/2d}. \end{aligned}$$

In particular, for fixed k, d, and $n \to \infty$,

$$\mathrm{ObsDiam}(\Sigma)_{\mathrm{DIST}} \leq n^{-1/2d+\alpha_n}$$

for some $\alpha_n \to 0$ as $n \to \infty$.



(G‴) At last we bring everything together and evaluate $\mathrm{ObsDiam}(X)_{\mathrm{DIST}}$ by looking at $U_\varepsilon(\Sigma) \cap X \subset X$ and arguing as above. First, we observe with $(\star)$ that the metric DIST on $U_\varepsilon(\Sigma)$ is bounded by

$$\mathrm{DIST}(x_1, x_2) \leq C_{k,d}((\mathrm{dist} + 2\varepsilon)^{1/d} + 2\varepsilon^{1/d})$$

via the triangle inequality.

It follows that

$$\mathrm{ObsDiam}(U_\varepsilon(\Sigma), -\kappa)_{\mathrm{DIST}} \leq 2C_{k,d}((\mathrm{ObsDiam}(U_\varepsilon(\Sigma); -\kappa)_{\mathrm{dist}} + 2\varepsilon)^{1/d} + 2\varepsilon^{1/d},$$

which should be combined with the obvious inequalities

$$\mathrm{ObsDiam}(U_\varepsilon(\Sigma), -\kappa)_{\mathrm{dist}} \leq \mathrm{ObsDiam}(X, -\kappa)_{\mathrm{dist}}$$

and

$$\mathrm{ObsDiam}(X, -\kappa - \kappa_\varepsilon)_{\mathrm{DIST}} \leq \mathrm{ObsDiam}(U_\varepsilon(\Sigma), \kappa)_{\mathrm{DIST}}$$

for $\kappa_\varepsilon = \mu_X(X \setminus U_\varepsilon(\Sigma))$. We then use our bound

$$\mathrm{ObsDiam}(X; -\kappa)_{\mathrm{dist}} \leq 10\left(-\frac{k}{n}\log\left(\frac{k}{n}\kappa\right)\right)^{1/2} \leq 10\left(-\frac{\ell}{n}\log\left(\frac{k}{n}\kappa\right)\right)^{1/2}$$

from $3\frac{1}{2}.28$ and the inequality $\kappa_\varepsilon \leq 8/\varepsilon e^{-\varepsilon^2 n/72\ell}$ in (G'). Next, we apply all these inequalities to $\varepsilon = \varepsilon_x = 10(-(\ell/n)\log(k\kappa/n))^{1/2}$. Then we have

$$\mathrm{ObsDiam}(X, -\kappa - \kappa_\varepsilon)_{\mathrm{DIST}} \leq 100 C_{k,d}\left(\frac{\ell}{n}\right)^{1/2d}\left(-\log\left(\frac{k}{n}\kappa\right)\right)^{1/2d}$$

for

$$\begin{aligned}\kappa_\varepsilon &\leq \left(-\frac{\ell}{n}\log\left(\frac{k}{n}\kappa\right)\right)^{-1/2}\exp\frac{n}{\ell}\left(\frac{\ell}{n}\right)^{1/2d}\log\left(\frac{k}{n}\kappa\right)\\ &= \kappa(k/\ell n)^{1/2}/(-\log((k/n)\kappa))^{1/2}.\end{aligned}$$

This implies $\kappa_\varepsilon \le \kappa$ for $\kappa \ge ne^{-n}$ and consequently,

$$\mathrm{ObsDiam}(X, -\kappa)_{\mathrm{DIST}} \le 100C_{k,d}\left(\frac{\ell}{n}\right)^{1/2d}(-\log(k\kappa/n))^{1/2d}$$

for all n and $\kappa \ge 2ne^{-n}$, where $\ell \le k(d-1)$ and $C_{k,d} \le 10^9 d^{3k+2}$.

So, *the final bound on the observable diameter of a complex algebraic submanifold $X \subset \mathbb{C}\mathrm{P}^n$ of codimension k and degree d and normalized Riemannian volume with the induced path metric reads*

$$\mathrm{ObsDiam}(X, -\kappa) \le 10^{11} d^{3k+2}\left(\frac{k(d-1)\log(2n/k\kappa)}{n}\right)^{1/2d} \qquad (**)$$

provided that $\kappa \ge 2ne^{-n}$. This makes

$$\mathrm{ObsDiam}(X) \le \mathrm{const}\cdot\left(\frac{\log n}{n}\right)^{1/2d}$$

for fixed k, d, κ, and $n \to \infty$.

(H) Remarks, exercises, and open problems. (a) Our argument can probably be improved to eliminate $\log n$ from the above estimate, but it is less clear whether the exponent $1/2d$ is to the point and what the right coefficient instead of d^{3k+2} should be. (It is highly unlikely that one can get a true insight into the asymptotic measure-theoretic geometry of algebraic varieties X with an argument similar to ours, since all we do is compare X with $\mathbb{C}\mathrm{P}^n$ based on a partial majorization of X by $\mathbb{C}\mathrm{P}^n$ without studying X at a closer range. But it may be useful to axiomatize our approach, which we leave to the pleasure of the reader.)

(b) *The ordinary diameter of* (X, DIST) *is bounded by*

$$\mathrm{diam}(X, \mathrm{DIST}) \le 10d$$

for every connected complex algebraic subvariety in $\mathbb{C}\mathrm{P}^n$ *of degree* d.

Proof. If $\dim_{\mathbb{C}} X = 1$, then the area of X equals d if we normalize the Kähler form (metric) ω in $\mathbb{C}\mathrm{P}^n$ by $\int_{\mathbb{C}\mathrm{P}^1}\omega = 1$. On the other hand, the unit disks in (X, DIST) have areas $\ge$ const $> 1/5$ (see (e) below). This settles the matter for curves $X \subset \mathbb{C}\mathrm{P}^n$ and the general case follows from the following obvious

(b_1) **Lemma:** *For every irreducible* $X \subset \mathbb{C}\mathrm{P}^n$ *of codimension* k *and every pair of points* $x_0, x \in X$, *there exists a* $(k+1)$*-plane in* $\mathbb{C}\mathrm{P}^n$ *passing through*

x_0 such that the irreducible component of the intersection $X \cap P$ containing x_0 also contains x.

(b_2) **Exercise.** Render the above argument *real* algebraic.

(b_3) What is the actual range of the function $X \mapsto \mathrm{diam}(X, \mathrm{DIST})$ for X running over the subvarieties in $\mathbb{C}P^n$ of given degree and dimension? Sid Frankel suggested that these diameters of curves in $\mathbb{C}P^2$ are bounded by a constant *independent of the degree*, but F. Bogomolov constructed a counterexample, i.e., a sequence of complex algebraic curves $X_i \subset \mathbb{C}P^2$ with $\mathrm{diam}(X_i, \mathrm{DIST}) \to \infty$. Nevertheless, it is still conceivable that a universal bound exists for hypersurfaces in $\mathbb{C}P^n$ for $n \geq 3$. Similarly, one wonders what the range of the function $X \mapsto \mathrm{ObsDiam}(X, -\kappa)_{\mathrm{DIST}}$ is for the X above. Does it truly depend on $\deg X$?

(c) The examples of functions diam and ObsDiam on the spaces of algebraic subvarieties in $\mathbb{C}P^n$ (or in any given ambient algebraic space for this matter) suggest a more general view on the structure of various geometric functions on the space of X's. To be specific, stick to hypersurfaces X in $\mathbb{C}P^n$ of degree d, where eventually $n, d \to \infty$. Such hypersurfaces themselves constitute a projective space $\mathcal{X} = \mathbb{C}P^N$, $N = (n+d)!/n!d! - 1$, which is highly concentrated for large d and/or n, and each "natural geometric function" $X \mapsto \mathrm{Inv}(X)$ should be strongly localized near a single value.

Among such functions (maps), one may distinguish certain universal ones such as the tautological map from $\mathcal{X} = \mathbb{C}P^N$ to the space of subsets $X \subset \mathbb{C}P^n$ with the Hausdorff metric, or to the space of measures on $\mathbb{C}P^n$, where each X goes to the (normalized) Riemannian measure supported on $X \subset \mathbb{C}P^n$.

(d) The "natural functions" are equally interesting in the real case, say on the space of real algebraic hypersurfaces $X \subset \mathbb{R}P^n$ of degree d. Here, one has several new possibilities, such as $X \mapsto \mathrm{vol}_{n-1}(X)$ and $X \mapsto$ (number of components of X), where the latter is accompanied by further topological invariants of X, such as the Betti numbers, the characteristic numbers, etc. For example, if $n = 1$ and $\dim X = 0$, then the basic invariant is the cardinality of X, and the question is to determine the measures of $\mathcal{X}_i = \{X \subset \mathcal{X} = \mathbb{R}P^d : \mathrm{card}\, X = i\}_{i=0,\dots,d}$ or rather the asymptotic behavior for $d \to \infty$ of some functions of $\mu(\mathcal{X}_i)$.

(d_1) **Exercise:** Consider the *Veronese curve* $V_1 \subset \mathbb{R}P^d$ that is the image of the *Veronese embedding* $\mathbb{R}P^1 \to \mathbb{R}P^d$ for $(t_1, t_2) \mapsto (t_1^d, t_1^{d-1}t_2, \dots, t_2^d)$ in homogeneous coordinates, and observe that the 0-dimensional subvarieties X in $\mathbb{R}P^1$ given by the homogeneous equations

$$c_0 t_1^d + c_1 t_1^{d-1} t_2 + \cdots + c_d t_2^d = 0$$

correspond to the hyperplanes H in $\mathbb{C}\mathrm{P}^d$ given by the equations $c_0 x_0 + c_1 x_1 + \cdots + c_d x_d = 0$, where $X = V^1 \cap H \subset V^1 = \mathbb{C}\mathrm{P}^1$. Evaluate length$(V)$ in $\mathbb{C}\mathrm{P}^d$, identify this using Crofton's formula with the average cardinality of X, that is,

$$\int_{\mathcal{X}} \operatorname{card} X \, dX = \int_{\mathcal{H}} (V^1 \cap H) \, dH$$

for $\mathcal{X} = \mathcal{H} = \mathbb{C}\mathrm{P}^d$ (where $\mathcal{H}$ appears as the projective space of the hyperplanes H in the space $\mathbb{C}\mathrm{P}^d$ containing V^1) and study the asymptotic behavior of this length$(V^1) = \int \operatorname{card}(X) \, dX$ for $d \to \infty$. Then show that the average number of roots of a random real polynomial of degree d grows approximately as $\log d$ (compare Ch. 1 in [Kac]).

Next, define the Veronese map $\mathbb{R}\mathrm{P}^n \to \mathbb{R}P^N$, $N = (n+d)!/n!d! - 1$, try to evaluate the n-dimensional volume and the total curvature of the Veronese image $V^n = \mathbb{R}P^n$ in $\mathbb{R}P^N$, and estimate the average Euler characteristic

$$\int_{\mathcal{X}} \chi(X) \, dX = \int_{\mathcal{H}} \chi(V^n \cap H) \, dH,$$

where $\mathcal{X} = \mathcal{H} = \mathbb{R}P^N$. (I admit, I did not try it myself; the reader is encouraged to study the work by Smale and Shub on the complexity of solutions of algebraic equations, especially their papers Bezout I. – Bezout V).

Question: How strongly does the Euler characteristic (function) $X \mapsto \chi(X)$ concentrate near its mean value? (Also one wonders how strongly concentrated the average estimates in [Sma–Shu]$_{\mathrm{BI-BV}}$ are.)

(*e*) Let $X \subset \mathbb{C}\mathrm{P}^n$ and $Y \subset X$ be complex algebraic varieties of dimensions m and n, respectively. Is the $2m$-volume of the intersection $X \cap U_\varepsilon(Y) \subset X$ bounded from below by $\operatorname{vol} U_\varepsilon(\mathbb{C}\mathrm{P}^r)$ in $\mathbb{C}\mathrm{P}^m$? This question makes sense for the ε-neighborhoods in X with respect to the metrics dist (induced from $\mathbb{C}\mathrm{P}^n$) and DIST (corresponding to the induced path metric in X), where the latter question seems significantly harder. For example, if $r = \dim Y = 0$, then the *sharp* lower bound on the volume of $U_\varepsilon(Y) = B(\varepsilon) \subset X$ for dist follows from the standard monotonicity formula for minimal subvarieties in $\mathbb{C}\mathrm{P}^n$ (Lelong inequality), but the similar sharp lower bound for the volumes of DIST-balls is probably still unsettled. In fact, one expects that every small domain $U \subset X$ satisfies the *sharp isoperimetric inequality*

$$\mathrm{vol}_{2m}(U) \le C_m I_m(\mathrm{vol}_{2m-1}\,\partial U),$$

where $I_m(s)$ denotes the volume of the ball B in $\mathbb{C}\mathrm{P}^m$ with $\mathrm{vol}_{2m-1}(B) = s$. Such an inequality (which rather easily implies the lower bounds on the volume of DIST-balls $B(\varepsilon)$ in X when applied to the concentric balls $B(t\varepsilon)$, $t \in [0,1]$) is known to hold for (domains in) *all minimal subvarieties* in $\mathbb{C}^n = \mathbb{R}^{2n}$ (this is highly nontrivial, see [Alm]$_{\mathrm{Iso}}$), and it yields some information on small domains $U \subset X \subset \mathbb{C}\mathrm{P}^n$ via suitable weakly distorting holomorphic diffeomorphisms of small Euclidean balls $B(\varepsilon) \subset \mathbb{C}^n$ to $\mathbb{C}\mathrm{P}^n$ (local Kähler coordinates in $\mathbb{C}\mathrm{P}^n$).

One can specify the question above to $Y = X \cap H$ for a hyperplane $H \subset \mathbb{C}\mathrm{P}^n$ and then inquire further into the measure-theoretic shape of the function $H \mapsto \mathrm{vol}_{2m}(U_\varepsilon(X \cap H))$ for H running over the space $\mathcal{H} = \mathbb{C}\mathrm{P}^n$. Actually, one may move Y in a wider class of subvarieties in X, i.e., one still expects the sharp lower bound on $\mathrm{vol}_m(U_\varepsilon(Y))$ in X for every $(2m-2)$-cycle Y in X homologous to $X \cap H$ in X without any holomorphicity condition on Y. But this question remains open, even for $\mathbb{C}\mathrm{P}^n$!

Finally, we observe that one has (sharp) lower bounds on the volumes of certain subvarieties in some (often homogeneous or locally homogeneous) spaces (e.g., obtained with *calibrating forms* of Harvey and Lawson) one expects similar bounds on the ε-neighborhoods of these submanifolds. For example, one expects that the ε-neighborhood of every r-dimensional subvariety Y nonhomologous to zero in a real, complex, or quaternionic projective space X has $\mathrm{vol}(U_\varepsilon(Y)) \ge J(\varepsilon)$, where $J(\varepsilon)$ denotes the volume of the ε-neighborhood of the projective subspace in X of (real) dimension r.

Exercise: Give a lower bound on $\mathrm{vol}\, U_\varepsilon(Y)$ for large $m = \dim X$ and small $\operatorname{codim} Y$ using the concentration in X and Milman's argument that we employed in (G$'$) above and earlier in $3\frac{1}{2}.28$.

(f) Large domains U in algebraic subvarieties $X \subset \mathbb{C}\mathrm{P}^n$ do not satisfy (uniform) isoperimetric inequalities unless $\deg X = 1$. For example, as a smooth quadric X in $\mathbb{C}\mathrm{P}^n$ degenerates to a union of two hyperplanes, it can be cut into two equal pieces along a (real) hypersurface S of arbitrarily small $(2n-3)$-volume. (Yet the ε-neighborhood of this S must have a significant $(2n-2)$-volume in X in accordance with the concentration property for X that we just spent 10 pages proving!) However, one expects that X can be decomposed into d (or slightly more than that) pieces, where each piece does satisfy an isoperimetric inequality. In fact, one also expects this to hold for real algebraic varieties X, where the mere existence of an isoperimetric inequality is unknown for singular X. Namely, one expects $\mathrm{vol}_n(U) \leq C_X(\mathrm{vol}\,\partial U)^\alpha$ for all small domains U in X with some $\alpha = \alpha(\dim X, \deg X)$. This underlies the difficulty in finding an appropriate lower bound on the spectrum of the Laplace operator on X (see [Gro]$_{\mathrm{SGSS}}$ on the state of affairs), which does not appear in the complex case (see [Li–Tian]).

On the development and ramifications of the idea of concentration. Paul Levy discovered the geometric concentration sometime between 1919 and 1951 and explained it in the second edition of his book on geometric functional analysis (see [Levy]). Levy's fascinating results and ideas had remained largely unknown for 20 years until Vitali Milman realized their importance and ubiquity. In particular, he used Levy's theorem for a short new proof of *Dvoretzky's theorem* (see [Milm]$_{71}$.

This Dvoretzky theorem is every bit as amazing and beautiful as Levy's result. It says that *every N-dimensional convex body $X \subset \mathbb{R}^N$ admits an ε-round section by a k-plane L passing through a given point $x_0 \in X$*, where "ε-round" means that $X \cap L$ is pinched between two concentric k-balls in L of radii R and $(1+\varepsilon)R$, and where $\varepsilon \leq \varepsilon(k, N) \to 0$ for $N \to \infty$ and every fixed k). Milman extended Levy's result to some nonspherical spaces (e.g., Grassmannian and Stiefel manifolds) and then pushed forward the idea of concentration as a general unifying principle. Ever since, Milman vigorously promoted *the concentration phenomenon*, as he called it, and the idea was reluctantly accepted by the community of functional analysts, but only after they located in the literature an alternative proof of the spherical isoperimetric inequality based on symmetrization (see [Fie–Lin–Mil]).

A couple of years later, Milman persuaded me to read Levy's book, and after deciphering Levy's terminology, I realized that his proof could easily be rendered rigorous using the geometric measure theory of the 1970's and then applied to all Riemannian manifold with a lower bound on the Ricci curvature (see Appendix C_+). In fact, I still don't know what to do with the original Levy idea of minimizing the volume of $U_\varepsilon(\Sigma) \subset X^n$ rather than $\mathrm{vol}_{n-1}(\Sigma)$ among all hypersurfaces $\Sigma \subset X$ dividing the volume of X in a given proportion.) And when Ricci $\geq c > 0$, the generalized Levy isoperimetric inequality turned out to yield new concrete concentration results (see [Gro–Mil]$_{\mathrm{TAI}}$), while the case Ricci $\geq c$ for $c < 0$ did not bring similar fruits, although it had deserved some merit in estimating the spectrum of the Laplace operator. But even this is questionable, since one does not need the full nonlinear variational power of Levy's argument to bound $\lambda_i(X)$ from below. This can also be achieved by elementary geometric and/or analytic means, where we do not seem to lose much precision, since Levy's inequality is never sharp for closed manifolds with Ricci ≤ 0 anyway (see [Ber–Gal], Appendix C_+, [Li -Yau]). Also see [Ledoux] for a simple proof of an asymptotically sharp inequality for Ricci > 0.)

Nowadays, the concentration phenomenon blends with geometric branches of probability theory, as is best witnessed by $[\mathrm{Tal}]_{\mathrm{CFII}}$ and $[\mathrm{Tal}]_{\mathrm{NLI}}$. [1] And the stand we take in this book is that the concentration reflects a convergence of spaces (such as S^n for $n \to \infty$) to the one point space, which we conceptualize further in 3.45.

F. Geometric invariants of measures related to concentration

A general scheme of evaluating the geometric size of a measure μ on a metric space X consists in picking up some metric invariant Inv expressing the idea of the size of a and defining $\mathrm{Inv}(\mu, m-\kappa)$ as the infimum of $\mathrm{Inv}(X')$ over all metric spaces $X' \subset X$ with $\mu(X') \geq m-\kappa$. (Here, m alludes to the mass $\mu(X)$ and then the condition on X' can be replaced by $\mu(X \setminus X') \leq \kappa$, provided $m = \mu(X) < \infty$. And if $\mu(X) = m = \infty$, one should distinguish these conditions). We have already done it to Inv = diam and also to the minimal covering number of a space by subsets of diameter $\leq \varepsilon$. Notice that these $\mathrm{Inv}(\mu, m-\kappa)$ are monotone under 1-Lipschitz pushforwards of measures. This suggests making the next step by defining

$$\mathrm{ObsInv}_Y(X, -\kappa) = \mathrm{Inv}(X \xrightarrow{\mathrm{Lip}_1} Y, m-\kappa)$$

for $m = \mu(X)$, as the supremum of $\mathrm{Inv}(f_*(\mu), m-\kappa)$ over all 1-Lipschitz maps $X \to Y$. Notice that this observable $\mathrm{Inv}_Y(X, -\kappa)$ is also monotone for 1-Lipschitz maps $(X, \mu) \to (X', \mu')$ if the original metric invariant Inv is monotone for surjective 1-Lipschitz maps. In addition to diameter and covering number, this includes the radius $\mathrm{Rad}\, X$ and many other first order invariants (see Ch. $3.\mathrm{D}_+$) such as the full (vector-valued) covering invariant accounting for the possibility of covering X by subsets $X_1, \ldots, X_N$ of diameters $\mathrm{diam}\, X_i \leq D_i$. We can also use higher-order invariants, e.g., the Hausdorff measure of a given exponent (dimension) d.

$3\frac{1}{2}$.30. Separation distance. Another kind of invariant of μ is defined with given numbers $\kappa_0, \ldots, \kappa_N$ as the supremal δ such that there exist subsets $X_i \subset (X, \mu)$, $i = 0, \ldots, N$, with $\mu(X_i) \geq \kappa_i$ and $\mathrm{dist}(X_i, X_j) \geq \delta$

[1] A wealth of new results appear in Sergey Bobkov's 1997 thesis published in Syktyvkar. Also see [Bob], [Bob–Houd], and [Bob–Led].

for $i \neq j$, where

$$\operatorname{dist}(X_i, X_j) \stackrel{\text{def}}{=} \inf_{x_i, x_j} \operatorname{dist}(x_i, x_j)$$

for $x_i \in X_i$ and $x_j \in X$. This is called the *separation distance* and is denoted

$$\operatorname{Sep}(X; \kappa_0, \dots, \kappa_N) = \operatorname{Sep}(\mu; \kappa_0, \dots, \kappa_N).$$

Clearly, this Sep is monotone for 1-Lipschitz maps $(X, \mu) \to (X, \mu')$.

$3\frac{1}{2}$.31. Observable central radius. A metric space Y with a measure ν is called *summable* if the distance function $\delta(y) = \operatorname{dist}(y_0, y)$ is summable, i.e., satisfies $\int_Y \delta(y)\, d\nu(y) < \infty$. If Y carries an *affine* structure in addition to a metric, such as in the case of a Banach space, then our (summable) measure ν on Y has a well-defined *center of mass*

$$c(\nu) = (\nu(Y))^{-1} \int_Y y\, d\nu(y).$$

We define $\operatorname{CRad}(\nu; m - \kappa)$ as the radius ρ of the minimal closed ball $B = B(c(\nu), \rho)$ in Y around $c(\nu) \in Y$ having mass at least $m\nu$

$$\nu(B) \geq m - \nu.$$

Then we define the *observable central radius* of a summable mm space X by minimizing $\operatorname{CRad}(f_*(\mu), m - \kappa)$ over all 1-Lipschitz maps $f \colon X \to Y$. It is clear that

$$\operatorname{ObsCRad} \geq \frac{1}{2} \operatorname{ObsDiam}$$

and

$$\operatorname{ObsCRad}_Y(X; -\kappa) \leq \operatorname{ObsDiam}_Y(X; -\kappa) + \frac{\kappa}{m - \kappa} \operatorname{diam}(X),$$

for $m = \mu(X)$ (compare 3.42(2)). In fact, there is a much better bound on ObsCRad if $\operatorname{ObsDiam}_Y(X; -\kappa)$ grows *slowly* as κ approaches zero. Let $\kappa(\rho)$ be the inverse function of $\rho(\kappa) = \operatorname{ObsDiam}_Y(X; -\kappa)$; thus $\kappa(\rho)$ equals the infimal $t \in [0, m]$, where $m = \mu(X)$, for which $\operatorname{ObsDiam}_Y(X; -\kappa) \leq \rho$. This is a monotone decreasing function on $\mathbb{R}_+ \ni \rho$ which equals m at $\rho = 0$ and goes to zero at infinity. The negative of the differential of $\kappa(\rho)$ gives us a measure, say $\nu_* = -d\kappa$ on $\mathbb{R}_+$ of total mass m. The center of mass of ν_* clearly equals

$$c(\nu_*) = m^{-1} \int_0^\infty \kappa(\rho)\, d\rho \in \mathbb{R}_+$$

(and the center of mass of this measure restricted to $[r, \infty)$ is

$$c_r(\nu_*) = r + (\kappa(r))^{-1} \int_r^\infty \kappa(\rho)\, d\rho).$$

Now, we bound $\text{ObsCRad}_Y(X; -\kappa)$ in terms of $\rho = \text{ObsDiam}$ and $c(\nu_*)$ (which is derived from ObsDiam) by arguing as follows. Let $f\colon X \to Y$ be a 1-Lipschitz map and $\nu = f_*(\mu)$. We have a subset $Y_{1/2} \subset Y$ of diameter $D_{1/2} \leq \rho(m/2)$ which contains at least half of the total mass of ν. Every other subset of mass $\geq m/2$ necessarily intersects $Y_{1/2}$, and so the R-balls in Y centered at some point $y_0 \in Y_{1/2}$ satisfy

$$\nu(B(R)) \geq m - \kappa$$

for $R = \rho(\kappa) + D_{1/2}$ and $\kappa \leq m/2$. What remains is to evaluate the distance between y_0 and the center of mass $c(\nu)$. To do this, we look at the measure ν'_* on $\mathbb{R}_+$ which equals the pushforward of ν under the map $y \mapsto \text{dist}(y_0, y)$ and observe that

$$\text{dist}(y_0, c(\nu)) \leq c(\nu'_*).$$

On the other hand, the measures ν_* and ν'_* on $\mathbb{R}_+$ have equal mass m and are related by the inequality

$$\nu'_*[0, \rho + D_{1/2}] \geq \nu_*[0, \rho]$$

for all $\rho \geq \rho_{1/2} = \rho(m/2)$, and so

$$c(\nu'_*) \geq c(\nu_*) + 2\rho_{1/2}.$$

Therefore,

$$\begin{aligned}\text{ObsCRad}_Y(X; -\kappa) &\leq \text{ObsDiam}_Y(X; -\kappa) + 3\rho_{1/2} + c(\nu_*) \\ &= \text{ObsDiam}_Y(X; -\kappa) + 3\,\text{ObsDiam}_Y(X; -m/2) \\ &\quad + m^{-1}\int_0^\infty \kappa(\rho)\, d\rho,\end{aligned}$$

where $m = \mu(X)$.

For example, in the case of $X = S^n$, we have $\nu_*(\rho)$ proportional to $(\cos\rho)^n\, dt$ on $[0, \pi/2]$, which is bounded by $e^{-n\rho^2/2}$, and so $c(\nu_*) \approx 1/\sqrt{n}$. Thus, $\text{ObsCRad}_{\mathbb{R}}(S^n)$ has the same order of magnitude as $\text{ObsDiam}_{\mathbb{R}}(S^n)$ for $\kappa \geq 1/2$ (with the convention $m = \mu(S^n) = 1$), and so

$$\text{ObsCRad}_{\mathbb{R}}(S^n) = O(\sqrt{\log \kappa^{-1}}/\sqrt{n}),$$

(which improves ($\circ$) in $3\frac{1}{2}$.19).

$3\frac{1}{2}$.32. Product inequalities. If (X, μ) equals the cartesian product of (X_1, μ_1) and (X_2, μ_2) with the (pythagorean) product metric $\text{dist}_X = \sqrt{\text{dist}^2_{X_1} + \text{dist}^2_{X_2}}$, then obviously

$$\text{diam}(\mu, (m_1 - \kappa_1)(m_2 - \kappa_2)) \leq \sqrt{(\text{diam}(\mu_1, m - \kappa_1))^2 + (\text{diam}(\mu_2, m - \kappa_2))^2}.$$

It follows that

$$\operatorname{diam}(X \xrightarrow{\mathrm{Lip}_1} \mathbb{R}^k, (m-\kappa)^k) \le \sqrt{k}\operatorname{diam}(X \xrightarrow{\mathrm{Lip}_1} \mathbb{R}, m-\kappa).$$

Furthermore, one has this inequality without $\sqrt{k}$ if one uses the L_∞-metric $\|y-y'\|_{L_\infty}$ in $\mathbb{R}^k$ corresponding to the norm $\|y\|_{\ell_\infty} = \sup_i |y_i|$, $i=1,\dots,k$.

Corollary: *If a sequence of mm spaces X^n concentrates to a point in the Levy sense, i.e., satisfies* $\operatorname{ObsDiam}_{\mathbb{R}}(X^n) \to 0$ *for each* $\kappa > 0$, *then* $\operatorname{ObsDiam}_{\mathbb{R}^k}(X^n) \to 0$ *for each* $\kappa > 0$ *and* $k = 1, 2, \dots$.

(where we abbreviate $\operatorname{ObsDiam}_{\mathbb{R}}(X) = \operatorname{diam}(X \xrightarrow{\mathrm{Lip}_1}, \mathbb{R}, m\kappa)$ as usual).

The observable central radius $\rho(\kappa)$ of the space $X = X_1 \times X_2$ above satisfies, for all κ_1 and κ_2,

$$\rho(m_2\kappa_1 + m\kappa_2) \le \rho_1(\kappa_1) + \rho_2(\kappa_2)$$

for $\rho_i(\kappa) = \operatorname{ObsRad}_Y(X_i; -\kappa)$, $i = 1, 2, \dots$ Here, $m_i = \mu(X_i)$, $i = 1, 2$, and Y is an arbitrary Banach space. This is seen for a given 1-Lipschitz map $f\colon X \to Y$ with an intermediate map $f_2\colon X_2 \to Y$ obtained by averaging f with respect to the variable x_1,

$$f_2(x_2) = m_1^{-1} \int_{X_1} f(x_1, x_2)\, d\mu_1(x_1).$$

This f_2 is $\mu_1(x)$-close to f outside of a subset of measure $m_2\kappa_1$ in X (where we think of f_2 as $f_2(x_1, x_2) = f_2(x_2)$). On the other hand, f_2 is $\rho_2(x)$-close to its average over X_2, which equals the average of f over X (where the μ-average of f over X equals the center of mass of $f_*\mu$ in Y).

Notice that this bound on $\rho(x)$ applies to the *maximal* product metric on X, also called the *ℓ_1-metric*

$$\operatorname{dist}_X = \operatorname{dist}_{X_1} + \operatorname{dist}_{X_2}.$$

Also notice that for $m_2 = 1$, our bound implies that

$$\rho(\kappa) \le \rho_1(\kappa/2) + \rho_2(\kappa/2).$$

If $Y = \mathbb{R}$, then the argument can be carried over with $\operatorname{ObsDiam} X$, where $f_2(x_2)$ is defined as the *Levy mean* (see $3\frac{1}{2}.19$) of $f(x_1, x_2)$ in the x_1 variable. This f_2 is also 1-Lipschitz, and thus one sees that the observable diameter $D(\kappa) = \operatorname{ObsDiam}_{\mathbb{R}}(X; -\kappa)$ satisfies

$$D(m_2\kappa_1 + \kappa_2) \le 2D_1(\kappa_1) + D_2(\kappa_2)$$

for all $\kappa_2 > 0$ and $0 < \kappa_1 < \mu(X_1)/2$.

Levy radius. To make this clearer, one can introduce the *Levy radius*, $\mathrm{LeRad}(X; -\kappa)$ as the infimal ρ, such that every 1-Lipschitz function $f(x)$ satisfies

$$\mu\{x \in X : |f(x) - a_0| \geq \rho\} \leq \kappa,$$

where a_0 stands for the Levy mean of f. Then the inequality for D above follows from the two obvious inequalities

$$\mathrm{LeRad}(X; -\kappa) \leq \mathrm{ObsDiam}_{\mathbb{R}}(X; -\kappa)$$

for $\kappa < \mu(X)/2$ and

$$D(m_2\kappa_1 + \kappa_2) \leq 2\,\mathrm{LeRad}(X_1; -\kappa_1) + D_2(\kappa_2),$$

which holds for all positive κ_1 and κ_2.

Also notice that the first inequality can be trivially reversed:

$$\mathrm{ObsDiam}_{\mathbb{R}}(X; -\kappa) \leq 2\,\mathrm{LeRad}(X; -\kappa)$$

for all $\kappa > 0$.

Exercise: Define and study the Levy radius for maps of X into a *tree*.

$3\frac{1}{2}$.33. Concentration and separation. Every μ has

$$\mathrm{diam}(\mu, m - \kappa) \geq \mathrm{Sep}(\mu, \kappa, \kappa),$$

and moreover,

$$\mathrm{diam}(X \xrightarrow{\mathrm{Lip}_1} \mathbb{R}, m - \kappa) \geq \mathrm{Sep}(X, \kappa, \kappa),$$

as is seen with the truncated distance function $\kappa \mapsto \max(\delta, \mathrm{dist}(\kappa, X_0))$ in X with $X_0, X_1 \subset X$ having measures $\geq \kappa$ and mutual distance δ.

Furthermore, every measure ν on $\mathbb{R}$ has

$$\mathrm{diam}(\nu, m - 2\kappa) \leq \mathrm{Sep}(\nu, \kappa, \kappa)$$

and consequently

$$\mathrm{Sep}(X; \kappa, \kappa) \geq \mathrm{diam}(X \xrightarrow{\mathrm{Lip}_1} \mathbb{R}, m - 2\kappa)$$

for all X and κ. Thus, *Levy's concentration of X^n (to a point) is equivalent to*

$$\mathrm{Sep}(X^n; \kappa, \kappa) \overset{n \to 0}{\longrightarrow} 0$$

for each $\kappa > 0$.

$3\frac{1}{2}$.34. Observable concentration on compact "screens" $Y \prec X$. Let Y be a metric space which can be covered by N subsets of diameters $\leq \delta$. Then every measure ν on Y satisfies

$$\operatorname{diam}(\nu, m - N\kappa) \leq \operatorname{Sep}(\nu, \kappa, \kappa) + 2\delta.$$

It follows that concentration of X^n to a point implies that (and is obviously implied by)

$$\operatorname{diam}(X^n \xrightarrow{Lip_1} Y, m - \kappa) \xrightarrow{n\to\infty} 0$$

for every *compact* space Y and every $\kappa > 0$. Thus, *every (observable) image of* (X^n, μ_n) *on a given compact "screen"* Y *concentrates to a single (luminous) point for large* $n = n(Y)$.

Notice that the rate of concentration of $\mu_* = f_*(\mu_n)$ on Y to a point depends only on the covering number $N = N(Y, \delta)$ of Y, and even better, this rate can be estimated by the covering properties of the measures μ_*. Namely, one has uniform concentration if (Y, μ_*) runs over a given *precompact* subset $\mathcal{Y} \subset \mathcal{X}$. In particular, *if* X^n *concentrate to a point, and if a sequence of spaces* $Y_i \prec X^{n_i}$ *for some sequence* $n_i \to \infty$ *converges in the* $\underline{\square}_\lambda$*-metric for some* $\lambda \geq 0$*, then the limit is necessarily isomorphic to a one point space. The converse is also true and obvious: if the* X^n *(with masses* $\mu_n(X^n) \leq \text{const}$*) do not concentrate, then there exists a sequence of measures* ν_i *on* $[0,1]$ *coming from* ν_i *on* X^{n_i} *via 1-Lipschitz maps such that* ν_i *converge to a measure* ν *on* $[0,1]$ *which is* not *concentrated at a single point.*

$3\frac{1}{2}$.35. Expansion coefficient and expansion distance. Denote by $\operatorname{Exp}(X; \kappa, \rho)$ the infimal $e \geq 1$ such that

$$\mu(X_0) \geq \kappa \Rightarrow \mu(U_\rho(X_0)) \geq e\kappa$$

for all subsets $X_0 \subset X$, where $U_\rho(X_0)$ denotes the ρ neighborhood of X_0. It is clear that Exp is monotone for 1-Lipschitz maps $(X, \mu) \mapsto (X', \mu')$, i.e.,

$$\operatorname{Exp}(X', \kappa, \rho) \geq \operatorname{Exp}(X, \kappa, \rho),$$

and Exp is well-behaved under the scaling of the measure μ of X

$$\operatorname{Exp}(c\mu, c\kappa, \rho) = \operatorname{Exp}(\mu, \kappa, \rho).$$

Another pleasant feature of Exp is the self-reinforcing property: *if* $\operatorname{Exp}(X, \kappa, \rho) \geq e$ *and* $\operatorname{Exp}(X, e\kappa, \rho') \geq e'$*, then* $\operatorname{Exp}(X, \kappa, \rho + \rho') \geq ee'$*,* *since* $U_{\rho+\rho'}(X_0) \supset U_{\rho'}(U_\rho(X_0))$ (in fact, one has the equality for *path* metric spaces X). This can be repeated several times and shows that

$$\operatorname{Exp}(X, \kappa', \rho) \geq e \Rightarrow \operatorname{Exp}(X, \kappa, k\rho) \geq e^k, \tag{$*$}$$

provided that the first inequality is valid for all $\kappa' \in [\kappa, k\kappa]$.

It is sometimes convenient to express δ as a function of e, and we call this the *expansion distance*, $\operatorname{ExDis}(X, \kappa, e)$, which is the infimal (distance) ρ which suffices for the implication

$$\mu(X_0) \geq \kappa \Rightarrow \mu(U_\rho(X_0)) \geq e\kappa$$

to hold. Here one has the obvious bound on the separation distance by ExDis,

$$\operatorname{Sep}(X; \kappa_0, \kappa_1) \leq \operatorname{ExDis}(X; \kappa_0, e_0) + \operatorname{ExDis}(X, \kappa_1, e_1) \tag{$+$}$$

provided that

$$e_0\kappa_0 + e_1\kappa_1 \geq m = \mu(X).$$

This by itself is not very useful for bounding $\operatorname{Sep}(X, \kappa_0, \kappa_1)$ when e_0, e_1 are small, and where both κ_0 and κ_1 are small compared to m, but one can combine it with $(*)$ and conclude that the bound

$$\operatorname{ExDis}(X; \kappa, e_0) \leq \rho_0$$

for some $e_0 > 1, \rho_0 > 0$, and $\kappa \in [\kappa_0, m/2]$ implies that

$$\operatorname{Sep}(X, \kappa_0, \kappa_0) \leq \frac{1}{2}(\log_{e_0}(m/\kappa_0))\rho_0. \tag{$+\kappa$}$$

In particular, if we have $m = \mu(X) = 1$, then

$$\rho(X, e_0) \overset{\text{def}}{=} \sup_{0 \leq \kappa \leq 1/2} \operatorname{ExDis}(X; \kappa, e_0)$$

for a given $e_0 > 1$, say for $e_0 = 4/3$, essentially majorizes the separation distances and hence the observable diameter of X. Namely,

$$\operatorname{Sep}(X, \kappa, \kappa) \leq \frac{1}{2}\rho(X, e_0) \log_{e_0} \kappa^{-1}$$

(where $\log_{e_0} = \log / \log e_0$), and

$$\operatorname{diam}(X \xrightarrow{\operatorname{Lip}_1} \mathbb{R}; 1 - 2\kappa) \leq \frac{1}{2}\rho(X, e_0) \log_{e_0} \kappa^{-1}.$$

It follows that if *spaces* X^n, $n = 1, 2, \ldots$ *with normalized measures satisfy* $\rho(X, e_0) \to 0$ *for some* $e_0 > 1$, *then they concentrate to a point. In fact,* *one only needs* $\mathrm{ExDis}(X^n, \kappa, e_0) \to 0$ *for* $\kappa \in [\kappa_n, 1/2]$, *where* $\kappa_n \to 0$ *as* $n \to \infty$.

Notice that the expansion inequality $\mathrm{Exp}(X, \kappa, \rho) \geq e_0$ for $\kappa \leq m/2$ implies comparable inequalities for all $\kappa \leq m - \varepsilon$. Indeed, just look at the complement $Y_\rho = X \setminus U_\rho(X_0)$ and observe that $U_\rho(X_0) \setminus X_0 \supset U_\rho(Y_\rho) \setminus Y_\rho$. If $\kappa_0 = \mu(X_0) \geq m/2$, then $\kappa_1 = \mu(Y_\rho) \leq m/2$, and so

$$\mu(U_\rho(X_0)) - \kappa_0 \geq (e_0 - 1)\kappa_1,$$

while $\mu(U_\rho(X_0)) = m - \mu(Y_\rho) = m - \kappa_1$. Thus,

$$\mu(U_\rho(X)) \geq \frac{\kappa_0 - (e_0 - 1)m}{e_0}, \tag{$\diamond$}$$

which is a definite improvement over κ_0 insofar as κ_0 does not approach $m = \mu(X)$. In fact, ($\diamond$) can be derived from (+) above combined with the obvious inequality opposite to (+),

$$\mathrm{ExDis}(X, \kappa, e) \leq \mathrm{Sep}(X, \kappa, m - e\kappa). \tag{$-$}$$

Exercise: Bound $\mathrm{Exp}(X; \kappa, \rho)$ from below in terms of $\mathrm{ObsDiam}(X)$.

Notice that many spaces X with infinite measures have the following expansion property: $\mathrm{Exp}(X, \kappa, 1) \geq e_0 > 1$ for all $\kappa > 0$ (or at least for all $\kappa \geq \kappa_0$). This is essentially equivalent to the *linear isoperimetric inequality* (which amounts to *nonamenability* of the isometry group Γ of X in the case where $V = X/\Gamma$ is compact).

Typical examples of such X's are provided by complete, simply connected manifolds with *negative curvature* $K \leq -\mathrm{const} < 0$, or more generally with $K \leq 0$ and $\mathrm{Ricci} \geq -\mathrm{const} < 0$. Also, every infinite simplicial tree built of unit segments, where each vertex has at least 3 adjacent edges obviously has $\mathrm{Exp} \geq \lambda > 1$. What is not obvious is the existence of sequences of *finite* graphs X^n, built of unit edges and having at most k_0 edges at each vertex (say with $k_0 \geq 3$) and $\mathrm{Exp}(X^n, \kappa, 1) \geq e_0 > 1$, while the number of vertices in X^n goes to infinity as $n \to \infty$. Such graphs X^n have $\mathrm{ObsDiam}_{\mathbb{R}} X^n = O(1)$ (and $AvDi(X^n) \approx \log n$). Their construction is indicated in $3\frac{1}{2}.44$.

$3\frac{1}{2}$.36. Spaces $\mathcal{L}ip_s/\,\mathrm{const}$ and their diameters. Given an mm space X, we denote by $\mathcal{L}ip_s = \mathcal{L}ip_s(X)$ the space of s-Lipschitz functions $f\colon X \to \mathbb{R}$, and we observe that the me_λ-metrics on $\mathcal{L}ip_s$ are invariant under the transformations $f \mapsto f + c$ for all constant (functions) $c \in \mathbb{R}$. Then we pass to the quotient $\mathcal{L}ip_s/\,\mathrm{const}$, where the metrics me_λ descend and are still denoted by me_λ. It is well-known (and obvious from the discussion in $3\frac{1}{2}$.A) that the spaces $\mathcal{L}ip_s/\mathrm{const}$ *are compact with respect to* me_λ *for all* $s \geq 0$ *and* $\lambda > 0$. Then we define the me_λ-*diameter of* X as the me_λ-diameter of $\mathcal{L}ip_s/\mathrm{const}$ for $s = 1$.

Notice that $\mathcal{L}ip_1/\mathrm{const}$ has a distinguished point denoted $0 \in \mathcal{L}ip_1/\mathrm{const}$ corresponding to constants, and one may speak of *the* me_λ *radius of* X, i.e., the minimal radius of the me_λ ball at 0 covering all of $\mathcal{L}ip_1/\mathrm{const}$.

Now we present yet another (obvious) characterization of Levy concentration.

X^n concentrates to a point, i.e., we have $\mathrm{ObsDiam}_{\mathbb{R}}(X^n) \to 0$ *if and only if the corresponding spaces* $(\mathcal{L}ip_1/\mathrm{const}, m_\lambda)$ *Hausdorff-converge to a single point for some (and hence every)* $\lambda > 0$, *that is, if* $\mathrm{diam}_{\mathrm{me}_\lambda}(\mathcal{L}ip_1 X^n/\mathrm{const}) \to 0$ *for* $n \to \infty$.

This raises the question of what happens to X^n if the spaces $\mathcal{L}ip_1(X^n)$ Hausdorff-converge to some (compact metric) space $\mathcal{L}$. This happens, for example, to $X^n = S^1 \times S^2 \times \cdots \times S^n$, and we shall address this question in $3\frac{1}{2}$.53.

The Kuratowski embedding of X into $\mathcal{L}ip_1/\mathrm{const}$. This embedding is defined by $x_0 \mapsto d_{x_0}(x) = \mathrm{dist}(x_0, x) : X \to \mathbb{R}$ with the following factorization by constants, which gives us a topological embedding, say $I : X \to \mathcal{L}ip_1/\mathrm{const}$, and the induced metric on X denoted me^*_λ. Since $\mathcal{L}ip_1/\mathrm{const}$ is compact, X is precompact with respect to me^*_λ, and we may compactify X by completing me^*_λ. Thus, we get the me^*_μ-*compactification* $(\overline{X}, \mathrm{me}^*_\lambda)$ of an arbitrary mm space with finite measure. It follows from the above that the Levy concentration of X^n implies the Hausdorff convergence of $\overline{X}^n$ to a single point, i.e.,

$$\mathrm{ObsDiam}_{\mathbb{R}}(X^n) \to 0 \Rightarrow (\mathrm{diam}(\overline{X}^n), \mathrm{me}^*_1) \to 0.$$

Question. What are the minimal (natural) assumptions on X^n that would imply the reverse implication

$$(\mathrm{diam}(\overline{X}^n), \mathrm{me}^*_1) \to 0 \Rightarrow \mathrm{ObsDiam}_{\mathbb{R}}(X^n) \to 0?$$

In other words, when does the concentration of the *distance functions* $x \mapsto \mathrm{dist}(x_0, x)$ on X for all $x_0 \in X$ imply the concentration of *all 1-Lipschitz functions* f on X? For many examples, such as round spheres S^n and other symmetric spaces, the concentration of the distance function is child's play compared to that for all Lipschitz functions f. But, if we look at more general spaces, say homogeneous, nonsymmetric ones, or manifold X^n with

Ricci $X^n \geq n$, then establishing the concentration for the distance functions becomes a respectable enterprise.

G. Concentration, spectrum, and the spectral diameter

$3\frac{1}{2}$.37. The space Dir_1 and the first eigenvalue. For the time being, let X stand for a *Riemannian* manifold and take some (not necessarily Riemannian) measure μ on X. We denote by $\mathrm{Dir}_1 = \mathrm{Dir}_1(X,\mu)$ the space of locally Lipschitz functions $f\colon X \to \mathbb{R}$ satisfying $\int_X \|\operatorname{grad} f\|^2\, d\mu \leq 1$, where $\|\operatorname{grad} f(x)\|$ of a non-C^1-smooth function f refers to the Lipschitz constant or dilatation at x (denoted by $\mathrm{dil}_x f$ in 1.1). We factorize this Dir_1 by the constants (where the Dirichlet functional $\int \|\operatorname{grad} f\|^2$ vanishes anyway), and we descend the Hilbert norm $\|f\|_2 = \left(\int_X f^2\, d\mu\right)^{1/2}$ from $L_2(X,\mu)$ to L_2/const and thus to our $\mathrm{Dir}_1/\mathrm{const}$. Now we define the *spectral diameter* of (X,μ) as the diameter of $\mathrm{Dir}_1(X,\mu)/\mathrm{const}$ with respect to the descended L_2-metric,

$$\mathrm{SpecDiam}(\mu) = \mathrm{SpecDiam}(X,\mu) = \mathrm{diam}(\mathrm{Dir}_1/\mathrm{const}).$$

Notice that we could make this definition with C^1-smooth functions f. This would not change SpecDiam, since every Lipschitz function can be approximated by smooth ones with the same (local) dilatation at all $x \in X$. Also, observe that

$$\mathrm{SpecDiam} = 2/\sqrt{\lambda_1}$$

for the first eigenvalue λ_1 of the *Laplace-Beltrami* operator on X, assuming that X is compact and that μ is the Riemannian measure of X, where λ_1 is defined via the variational principle (see below). In fact, our definition of $\mathrm{SpecDiam}(X)$ makes sense for all mm spaces X with the convention

$$\|\operatorname{grad} f(x)\| = \mathrm{dil}_x f,$$

and so we could *define* $\lambda_1(X)$ as $4/(\mathrm{SpecDiam})^2$ for all mm spaces X (compare (C) below).

Notice that SpecDiam and λ_1 are invariant under the scaling of the measure and that $\mathrm{SpecDiam}(\ell X,\mu) = \ell\,\mathrm{SpecDiam}(X,\mu)$, where ℓX means $\mathrm{dist}_X \rightsquigarrow \ell\,\mathrm{dist}_X$ as usual.

Now we want to compare SpecDiam with the above me_1-diameter of X, i.e., the me_1-diameter of the space $\mathcal{L}ip(X)/\mathrm{const}$. We assume that μ is normalized, and then we obviously have $\int_X \|\operatorname{grad} f\|^2\, d\mu \leq 1$. Thus,

$$\mathrm{Dir}_1/\mathrm{const} \supset \mathcal{L}ip_1/\mathrm{const}\,.$$

On the other hand, the L_2 metric majorizes the metric me_1 by the Chebyshev inequality, and thus, $\mathrm{diam}_{\mathrm{me}_1})(\mathcal{L}ip_1/\,\mathrm{const})$ can be bounded in terms of SpecDiam. Consequently,

$$\mathrm{SpecDiam}\, X^n \to 0 \qquad \Rightarrow \qquad \mathrm{diam}_{\mathrm{me}_1}((\mathcal{L}ip_1 X^n)/\,\mathrm{const}) \to 0,$$

or, equivalently, $\lambda_1(X) \to \infty \Rightarrow X^n$ *collapses to a point, where we assume that* $\mu_n(X^n) = 1$.

$3\frac{1}{2}$.38. Spectral bound on ObsDiam. Let us give a specific bound on Exdis and thus on ObsDiam in terms of SpecDiam. Take a subset $X_0 \subset X$ of measure $\mu(X_0) \geq \kappa_0 \leq 1/2$ and, following $[\mathrm{Che}]_{\mathrm{RBLD}}$ and $[\mathrm{Che}]_{\mathrm{lbse}}$, consider the (test) function f interpolating between two constants c_0 on X_0 and c_1 on $X_1 = X \setminus U_\rho(X_0)$, for the (open) ρ-neighborhood $U_\rho(X_0)$, by the distance function between X_0 and X_1 as follows:

$$f(x) = \begin{cases} c_0 + \eta\, \mathrm{dist}(x, X_0) & \text{for } x \in U_\rho(X_0), \\ \quad c_1 & \text{on } X_1 = X \setminus U_\rho(X_0), \end{cases}$$

where we take $\eta = \rho^{-1}(c_1 - c_0)$; so the function f is continuous and thus η-Lipschitz on X. The Dirichlet functional of this f obviously equals $\eta^2\kappa'_\rho$, where κ'_ρ denotes the measure of the *closure* of the "annulus" $U_\rho(X_0) \setminus X_0$ (which may be strictly greater than the measure of $U_\rho \setminus X_0$). We take c_1 such that $\eta^2\kappa'_\rho = 1$, i.e.,

$$c_1 = c_0 + \rho/\sqrt{\kappa'_\rho}.$$

The squared L_2-norm of this f equals

$$\|f\|^2_{L_2} = \int_\kappa f^2\, d\mu = c_0^2\kappa_0 + c_1^2\kappa_1$$

for κ_1 being the measure of X_1. Then, assuming $\mu(U_\rho(X_0)) \leq e_0\kappa_0$ with $e_0 \leq 3/2$ (and recalling that $\mu(X_0) \leq 1/2 = \mu(X)/2$), we obtain the following lower bound on the L_2-norm of f,

$$\|f\|^2_{L_2} \geq c_0^2\kappa_0 + \frac{1}{2}\left(c_0 + \frac{\sqrt{2}\rho}{\sqrt{\kappa_0}}\right)^2 \kappa_0 \geq \frac{2}{3}\rho^2,$$

independently of c_0, which shows that $\mathrm{SpecDiam}(X) \geq \rho$ since f was chosen in $\mathrm{Dir}_1(X, \mu)$. In other words,

$$\mathrm{Exdis}(X, \kappa, e) \leq \frac{1}{2}\,\mathrm{SpecDiam}(X) \qquad (*)$$

for all $\kappa \in [0,1/2]$ and $e \leq 3/2$. This implies (see $(+\kappa)$ in $3\frac{1}{2}.35$) that

$$\operatorname{Sep}(X,\kappa,\kappa) \leq \frac{1}{2}\log_{3/2}\kappa^{-1}\operatorname{SpecDiam}(X)$$

and

$$\begin{aligned}\operatorname{ObsDiam}(X,-2\kappa) &= \operatorname{diam}(X \xrightarrow{\operatorname{Lip}_1} \mathbb{R}, 1-2\kappa)\\ &\leq \frac{1}{2}\log_{3/2}\kappa^{-1}\operatorname{SpecDiam}(X)\\ &\leq 2\ (\log\kappa^{-1})\operatorname{SpecDiam}(X) \qquad (\star)\end{aligned}$$

(see [Gro–Mil]$_{\text{TAI}}$, [Milm]$_{\text{HPL}}$, and [Mil–Sch] for further information).

$3\frac{1}{2}$.39. Nonlinear and linear spectra of X. The general variational definition of the spectrum is as follows. Consider some homogeneous functional E on the space of Lipschitz functions f on X, such as, for example,

$$\int (\operatorname{dil}_x f(x))^2\, d\mu_x \Big/ \int f^2(x)\, d\mu_x\ ,$$

(where we recall that dil_x denotes the Lipschitz constant of f at x, which amounts to $\|\operatorname{grad} f(x)\|$ for Riemannian spaces X), or more generally,

$$E(f) = \left(\int (\operatorname{dil}_x f(x))^p\, d\mu_x\right)^{\frac{1}{p}} \Big/ \left(\int f^q(x)\, d\mu_x\right)^{\frac{1}{q}}\ .$$

This is thought of as a function E on the projective space P^∞ of functions on X. There are then several ways to extract numerical invariants from X.

(A) Define λ_i^+ as the infimal λ for which there exists an i dimensional projective subspace $P^i \subset P^\infty$, where $E \leq \lambda$.

(B) Let λ_i^* be the infimal λ such that the inclusion homomorphism

$$H_i(E^{-1}[0,\lambda];\mathbb{Z}_2) \to H_i(P^\infty;\mathbb{Z}_2)$$

is surjective.

Notice that since each $P^i \subset P^\infty$ represents a nonzero class in the homology group $H_i(P^\infty;\mathbb{Z}_2)$, one has the bound $\lambda_i \leq \lambda_i^+$ for all $i = 0,1,\ldots$. And if the homogeneous functionals in question are quadratic (as are the L_2-norm and the Dirichlet functional on Riemannian manifolds), then clearly $\lambda_i = \lambda_i^+$. In particular, $\lambda_1 = 4/(\operatorname{SpeDiam})^2$ for (compact) Riemannian manifolds X, i.e.,

$$\lambda_1 = \inf_f \int_X \|\operatorname{grad} f\|^2\, d\mu \Big/ \int_X f^2,$$

where f runs over all Lipschitz functions f orthogonal to constants, i.e., satisfying $\int_x f = 0$. Also observe that the spectra $\{\lambda_i\}$ and $\{\lambda_i^+\}$ are monotone for the Lipschitz order on mm spaces (see [Gro]$_{\mathrm{DN}\ell\mathrm{S}}$ for further information and [Hof–Zehn] for a symplectic rendition of the spectral idea).

There are certain (truly) quadratic alternatives to $\int_X (\mathrm{dil}_x f)^2$ for non-Riemannian X which correspond to the spectra of *linear* operators. For example, for every (positive symmetric) function K on $X \times X$, e.g., for the indicator function of the ε-neighborhood of the diagonal, one can take the L_2-norm of $K(x, x')(f(x) - f(x'))$ on $X \times X$ as a substitute for $\left(\int \|\operatorname{grad} f\|^2\right)^{1/2}$. (This works pretty well if X is the vertex set of a graph with unit edges, in which case the λ_i constructed in this way when $K(x, x')$ is the incidence matrix of the graph, i.e., the indicator function of the 1-neighborhood of the diagonal in $X \times X$, give us the usual combinatorial spectrum; see [Lub] and the references therein.) The λ_i's one gets in this way essentially constitute the spectrum of the integral operator $\mathcal{K}$ on (functions on) X given by the kernel K. More specifically, one can apply it to the operator $\mathcal{K}_\varphi$ given by the integral kernel $K_\varphi(x, x') = \varphi(\mathrm{dist}(x, x'))$ for a function $\varphi(d)$, thus getting the maps $\varphi \mapsto \{\lambda_i = \lambda_i(\varphi)\}$ serving as a rather comprehensive spectral invariant of X. Furthermore, one can take several functions φ_i and look at the spectra of operators appearing as some (noncommutative) polynomial in the variables $\mathcal{K}_{\varphi_i}$. Finally, one can study the operators given by the kernels $K(x, x')(f(x) - f(x'))$ for functions f on X in the spirit of Alain Connes (see [Connes]). But all these possibilities and their relation to the concentration remain largely unexplored, except for the case of the usual λ_i corresponding to $\mathrm{Dir}_1(f)/\|f\|_{L_2}^2$ on Riemannian manifolds (and graphs) X where $\mathrm{Dir}_1(f) = \int_X \langle \operatorname{grad} f, \operatorname{grad} f\rangle = \int_X f\Delta f$, and so the λ_i appear as the ordinary eigenvalues of the Laplace operator Δ on X.

$3\frac{1}{2}$.40. Evaluation of the spectrum for S^n. The spectrum of the Laplace operator on a sufficiently homogeneous space X can be determined by means of harmonic analysis which combines with $(\star)$ above to give nontrivial bounds on $\mathrm{ObsDiam}_{\mathbb{R}} X$. The simplest example is the unit sphere $X = S^n$, where the first eigenvalue consists of the linear functions $S^n \subset \mathbb{R}^{n+1} \to \mathbb{R}$ and is spanned by the coordinate projections, say $\varphi_i : S^n \to \mathbb{R}$, $i = 0, \dots, n$, which clearly satisfy

$$\sum_{i=0}^{n} \varphi_i^2(x) = 1$$

and

$$\sum_{i=0}^{n} \|\operatorname{grad} \varphi_i(x)\|^2 = n$$

for all $x \in S^n$. Thus,

$$\lambda_1 = \int \|\operatorname{grad} \varphi_i\|^2 / \int \varphi^2 = n, \tag{Δ}$$

and $(\star)$ above implies

$$\operatorname{ObsDiam}(S^n, -2\kappa) = \operatorname{diam}(S^n \xrightarrow{\mathrm{Lip}_1} \mathbb{R}, 1 - 2\kappa) \le 2(\log \kappa^{-1})/\sqrt{n} \tag{$\circ$}$$

which, although somewhat weaker than Levy's inequality $(*)$ in $3\frac{1}{2}.19$ is sufficient to conclude that $\operatorname{ObsDiam}(S^n) \approx 1/\sqrt{n}$.

$3\frac{1}{2}$.41. Evaluation of the observable radius $\operatorname{CRad}(S^n \xrightarrow{\mathrm{Lip}} \mathbb{R}^m, 1-\kappa)$. Start with a measure ν on $\mathbb{R}^m$ having its center of mass at the origin, i.e., satisfying $\int_{\mathbb{R}^m} \varphi \, d\nu = 0$ for all linear functions $\varphi \colon \mathbb{R}^m \to \mathbb{R}$ and let us bound $\lambda_1(\nu)$ as follows. Suppose that the complement to the ρ-ball around the origin, say $C(\rho) = \mathbb{R}^m \setminus B(\rho)$, satisfies $\nu(C(\rho)) \ge \kappa$ and observe that the coordinate (projections) $\varphi_i \colon \mathbb{R}^m \to \mathbb{R}$, $i = 1, \ldots, m$ satisfy

$$\sum_{i=1}^{m} \varphi_i^2(y) \ge \rho^2$$

for $y \in C(\rho)$. It follows that

$$\int_{\mathbb{R}^m} \sum_{i=1}^{m} \varphi_i^2(y) \, d\nu(y) \ge \kappa\rho^2.$$

So there exists a (linear) 1-Lipschitz function φ on $\mathbb{R}^m$ with $\int_{\mathbb{R}^m} \varphi^2(y) \, d\nu_x \ge \kappa\rho^2/m$. Namely, one of the φ_i's will do. Consequently, *if there exists a 1-Lipschitz map $(X, \mu) \to \mathbb{R}^m$ sending μ to ν, then X supports a 1-Lipschitz function f such that*

$$\int_X f \, d\mu(x) = 0 \qquad \textit{and} \qquad \int_X f^2 \, d\mu(x) \ge \kappa\rho^2/m.$$

Therefore, if $\mu(X) = 1$, then

$$\lambda_1(X, \mu) \le m/\kappa\rho^2,$$

and if $\lambda_1(X, \mu) \ge n$, then

$$\kappa\rho^2 \le \frac{m}{n}. \tag{$*$}$$

Corollary:

$$\operatorname{ObsCRad}_{\mathbb{R}^m}(S^n; -\kappa) \le 2\frac{1}{\sqrt{\kappa}}\frac{\sqrt{m}}{\sqrt{n}}. \tag{$**$}$$

Thus, every 1-Lipschitz image of S^n in $\mathbb{R}^m$ concentrates to a point if $n/m \to \infty$.

Remarks and open questions: (a) It is probably not hard to prove a similar result for maps to $\mathbb{R}^m$ with the ℓ_p-metric for all $p > 1$.

(b) It would be nice to improve ($**$) by plugging in $-\log \kappa^1$, or even better $\sqrt{-\log \kappa^{-1}}$ instead of $1/\sqrt{\kappa}$. Some improvement may be possible with the isoperimetric inequality for ν as in Levy's normal law (see $3\frac{1}{2}$.23). In fact, one naively expects that the extremal (least concentrated) maps $S^n \to \mathbb{R}^m$ are $O(m)$-symmetric.

(c) There is a formal similarity between ($**$) and ($+$) in 3.42(9). Are the two related?

(d) For which $R = R(m,n)$ does the scaled sphere $RS^n = S^n(R)$ Lipschitz dominate $S^m = S^m(1)$? Also, one asks this for balls, i.e., when $RB^n \succ B^m$, and, of course, the question stands for all "interesting" mm spaces, such as compact symmetric spaces, product spaces, complex hypersurfaces, etc. And often we just want to know whether $RX_\varepsilon \succ Y_\varepsilon$, where X_ε and Y_ε are ε-close in the $\underline{\square}_\lambda$-metric for a given X and Y.

$3\frac{1}{2}$.42. Spectra of product spaces. It is remarkable, although well-known and easy to prove, that $\lambda_1(X_1 \times X_2) = \min(\lambda_1(X_1), \lambda_1(X_2))$. In particular, $\lambda_1(X \times X \times \cdots \times X) = \lambda_1(X) \geq \delta > 0$ (where there are n factors in the product), for all compact connected Riemannian manifolds X. Thus,

$$\operatorname{ObsDiam} X^n = O(1) \tag{$\square$}$$

for $n \to \infty$. This carries the same price tag as ($\circ$) above, since the characteristic size of the cartesian power X^n is $\approx \sqrt{n}$.

Remark. If X is a compact Riemannian manifold with *boundary*, then our λ_i correspond to the *Neumann* boundary problem for Δ.

Non-Riemannian products. To better appreciate ($\square$), look at the cartesian power $\{0,1\}^n$, i.e., *the set of the vertices* of the unit cube in $\mathbb{R}^n$ with the induced *Euclidean* metric. We claim that *the observable diameter here goes to infinity as $n \to \infty$*. In fact,

$$\operatorname{ObsDiam}\{0,1\}^n \sim n^{1/4}.$$

To see this, we look at the *summation map* from $\{0,1\}^n \subset \mathbb{R}^n$ to $\mathbb{R}$ given by $\sigma_n(x_1, x_2, \ldots, x_n) \mapsto x_1 + x_2 + \cdots + x_n$. This map sends $\{0,1\}^n$ to the segment $[0,n]$, and by the central limit theorem, the subsegments $I_n =$

$[0, n/2-\sqrt{n}]$ and $I'_n = [n/2+\sqrt{n}, n]$ receive definite amounts of the (discrete product) measure of $\{0,1\}^n$. That is, the σ_n-pushforwards of this measure satisfy

$$\mu_n(I_n) = \mu_n(I'_n) \geq \varepsilon_0 > 0$$

for $n \geq 4$ (with, say, $\varepsilon_0 = 1/20$). On the other hand, the maps σ_n on $\{0,1\}^n$ (but *not* on the whole cube $[0,1]^n$) satisfy

$$|\sigma_n(x) - \sigma_n(y)| \leq \|x-y\|^2_{\mathbb{R}^n}.$$

Therefore, the composition of σ_n (on the pullback of $I_n \cup I'_n$ in $\{0,1\}^n$) with the map sending $I_n \to 0 \in \mathbb{R}$ and $I'_n \to n^{1/4} \in \mathbb{R}$ is a 1-Lipschitz map from $\sigma_n^{-1}(I_n \cup I'_n) \subset \{0,1\}^n$ to $\{0, n^{1/4}\} \subset \mathbb{R}$. This can be 1-Lipschitz extended to all of $\{0,1\}^n$ (and even to all of $\mathbb{R}^n$ if we care) with a definite amount ($\approx 5\%$) of measure sent $n^{1/4}$-distance apart. This shows that

$$\mathrm{ObsDiam}(\{0,1\}^n, 1-\kappa) \geq \mathrm{const}(\kappa) n^{1/4},$$

and the opposite asymptotic inequality is explained in $3\frac{1}{2}.62(2)$.

$3\frac{1}{2}$.43. Congruence spaces. Take the group $SL_2\mathbb{R}$ with a left-invariant Riemannian metric and let $\Gamma_p \subset SL_2\mathbb{Z} \subset SL_2\mathbb{R}$, for prime p, consist of the matrices congruent to $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ mod p. It is well-known (and rather obvious) that the quotient spaces $X^p = SL_2\mathbb{R}/\Gamma_p$ have finite volumes. And a celebrated theorem of Selberg asserts that these Riemannian manifolds $X^p = SL_2\mathbb{R}/\Gamma_p$ have $\lambda_1(X^p) \geq \delta > 0$ for all $p = 2, 3, 5, 7, 11, \ldots$, and so their observable diameters remain bounded for $p \to \infty$. This should be seen in conjunction with the (easy) estimate of the characteristic size or of the averaged diameter of these spaces

$$AvDi(X^p) \approx \log p.$$

Thus, the scaled spaces $\varepsilon_p X^p$ for $\varepsilon_p \to 0$ and $\varepsilon_p \log p \geq \mathrm{const} > 0$ provide (highly) nontrivial examples of Levy's concentration phenomenon.

The bound $\lambda_1 \geq \delta > 0$ remains valid for all *congruence spaces*, i.e., for $X^p = G/\Gamma_p$, where G is a semisimple Lie group and Γ_p is a *congruence subgroup* of some arithmetic $\Gamma \subset G$ (where the arithmeticity of Γ implies $\mathrm{vol}(G/\Gamma_p) < \infty$). One can take, for example, $\Gamma_p \subset SL_m\mathbb{Z} \subset SL_m\mathbb{R}$ consisting of the matrices $\equiv \mathbf{1} (\mathrm{mod}\ p)$ and get $X^p = SL_m\mathbb{R}/\Gamma_p$ for all $m \geq 2$. Here, if $m \geq 3$, then the bound $\lambda_1(X^p) \geq \delta > 0$ for $p \to 0$ follows from a pure representation theory for $SL_m\mathbb{R}$, namely from *Kazhdan's T-property*, which implies, in particular, that $\lambda_1(SL_m(\mathbb{R})/\Gamma) \geq \delta > 0$ for all discrete subgroups Γ with finite covolume (see [Alon–Milm], [Alon], [Lub] and the references therein).

Remarks: (a) If G/Γ is noncompact (as is $SL_n\mathbb{R}/SL_n\mathbb{Z}$, for instance), then one should specify the space of (Lipschitz) functions f used in the variational definition of λ_1 via $\int \|\operatorname{grad} f\|^2/\int f^2$. However, the answer is insensitive to the particular space of functions we use. For example, we can take the (small) space of C^∞ functions f with compact support, or the (larger) space of Lipschitz L_2-functions f on $X = G/\Gamma$, and we get the same variational spectrum, as a trivial argument shows.

(b) It is customary to divide G by the maximal compact subgroup K and to study the spectrum on the locally symmetric (orbispace) $K\backslash G/\Gamma$, which corresponds to the spectrum of the Laplace operator on the *K-invariant* functions on G/Γ. For example, if $G = SL_2\mathbb{R}$, we come to the spectrum on the Riemann surface $SO(2)\setminus SL_2\mathbb{R}/\Gamma$ (which has singular points if we allow torsion in Γ). But dividing by K does not essentially change the spectrum; nor does it significantly change the observable diameter (this is rather obvious), and so the discussion above descends to the spaces $K\setminus G/\Gamma_p$.

$3\frac{1}{2}$.44. Expanders. This refers to (sequences of finite) graphs X with the following two properties.

(a) each vertex in X has at most k_0 adjacent edges, (where one can stick to $k_0 = 3$).

(b) $\exp(X, \kappa, 1) \geq e_0 > 1$, for all $\kappa \leq \frac{1}{2} \operatorname{card} \operatorname{ver}(X)$ $(*)$

where we give the path metric to X^n having all edges of unit length, and we assign unit mass to each vertex of X, while the edges carry no measure of their own. Thus, the measure κ above x corresponds to the cardinality of some subset V_0 in the vertex set $\operatorname{ver}(X)$. It is convenient to define the *boundary* ∂V_0 as the set of vertices in $\operatorname{ver}(X)\setminus V_0$ adjacent to V_0, i.e.,

$$\partial V_0 = \{x \in \operatorname{ver}(X) : \operatorname{dist}(x, V_0) = 1\}.$$

Then $(*)$ amounts to *the isoperimetric inequality*

$$\operatorname{card}(V_0) \leq (e_0 - 1)^{-1} \operatorname{card} \partial V_0, \qquad (*')$$

for all $V_0 \subset \operatorname{ver}(X)$ with $\operatorname{card}(V_0) \leq \frac{1}{2} \operatorname{card} \operatorname{ver}(X)$.

It is not hard to exhibit infinite expanders, such as regular trees, but it is a great deal harder to produce a sequence X^n of *finite* graphs with $\operatorname{card}(\operatorname{ver}(X^n)) \to \infty$ satisfying the above $(*)$ and/or $(*')$ with fixed k_0 and e_0, where we repeat that $(*')$ should hold for all subsets V_0 with cardinality less than half of that of ver(X). (We suggest that the reader try to construct such an example on her/his own in order to better appreciate Margulis' construction indicated below. I myself tried something elementary several times, see e.g., 9.2 in [Gro]$_{\text{FRM}}$, unsuccessfully attempting to bypass a Selberg-type theorem in the expander problem.)

Margulis' idea. Start with a finite graph X_1 in the space $X = G/\Gamma$ which generates the fundamental group of X. (Notice that this $\pi_1(X)$ equals Γ if G is a simply connected Lie group.) Then take the pullback $X_1^p \subset X^p$ of X_1 for the congruence covering map $X^p \to X$. Since the first eigenvalue $\lambda_1(X^p)$ remains separated from zero, the expanding coefficient of X^p is $\geq e > 1$. This trivially implies a similar property for X_1^p with a slightly smaller expansion constant $e_0 > 1$. In fact, the vertex set of $X_1^p \subset X^p$ fairly well approximates X^p as far as the expansion is concerned. Namely, the measure of each ball $B = B(R)$ in X^p for $R \geq R_0$ obviously satisfies

$$c_1\mu(B) \leq \operatorname{card}(B \cap \operatorname{ver}(X_1^p)) \leq c_2\mu(B)$$

for some $c_1, c_2 > 0$, provided that $X = G/\Gamma$ is compact. This allows a lower bound on the expansion coefficient of X_1^p in terms of that of X^p in the case where X is compact, and the general (noncompact) case requires a minor adjustment (left to the reader).

Remarks: (a) Instead of bounding $\exp(X_1^p;\kappa,1)$ from below in terms of $\exp(X^p;\kappa,1)$ minorized by $\lambda_1(X^p) \geq \delta > 0$, one could introduce the *combinatorial* first eigenvalue $\lambda_1(X_1^p)$ via $\operatorname{Dir}(f)/\|f\|_{L_2}^2$ for functions f on ver(X_1^p), where

$$\operatorname{Dir}(f) = \sum_{i,j}(f(x_i) - f(x_j))^2,$$

and the sum is taken over all pairs of adjacent vertices x_i, x_j (i.e., those having $\operatorname{dist}(x_i, x_j) \leq 1$ in the combinatorial metric on X_1^p). Then one could bound $\lambda_1(X_1^p)$ from below in terms of $\lambda_1(X^p)$ and finally deduce the lower bound on $\exp(X_1^p)$ by applying from that on $\lambda_1(X_1^p)$. (We suggest that the reader go through these steps in order to slowly get a better feel for λ_1.)

(b) Expanders appear in theoretical computer science. Their existence can be proved by a probabilistic argument as follows. Take a vertex set of the form $V = V' \sqcup V''$, where V' and V'' are sets of equal cardinality n, and let $\Pi_i \colon V' \to V''$, $i = 1, \dots, k$ be bijections. We take X consisting of the edges $(v', \Pi_i(v'))$ for all $v' \in V'$ and $i = 1, \dots, k$, and then evaluate the number of those k-tuples of bijections $(\Pi_1, \dots, \Pi_k)$ for which X fails to be expansive, i.e., contains a subset $V_0 \subset V$ of cardinality $\leq n = \operatorname{card}(V')$, such that the set of the vertices adjacent to V_0 (including those in V_0) has cardinality $\leq e_0 \operatorname{card}(V_0)$ for a suitable e_0, e.g., for $e_0 = 1.1$. Here is how the computation goes. The total number of our graphs equals $(n!)^k$ for $n = \operatorname{card}(V') = \operatorname{card}(V'')$, since these are given by k-tuples of bijections (permutations) $V \leftrightarrow V''$. Then we compute (as in [Lub]) the number of *bad* k-tuples of bijections, such that some subset $A' \subset V'$ of cardinality $r \leq n/2$ goes to a subset $A'' \subset V''$ of cardinality $\leq s = [3r/2]$ under all k bijections, where $[\cdot]$ denotes the integer part of a number. There are $\left(\binom{s}{r}(n-r)!\right)^k$ bad k-tuples of bijections for given A' and A'', and the number of possible A is $\binom{n}{r}\binom{n}{s}$. So, the totality of the bad choices equals $\sum_{r=1}^{[n/2]} B(r)$ for $B(r) = \binom{n}{r}\binom{n}{s}\left(\binom{s}{r}(n-r)!\right)^k$. To bound this sum, we evaluate the ratios $B(r+1)/B(r)$ for $r+1 \leq n/2$, which we write as $R_1R_2(R_3R_4)^k$, where

$$R_1 = \binom{n}{r+1}\Big/\binom{n}{r} = \frac{n-r}{r+1} < n,$$

$$R_2 = \binom{n}{s'}\Big/\binom{n}{s} \leq n^2$$

as $s' = \left[\frac{3(r+1)}{2}\right] \leq \left[\frac{3r}{2}\right] + 2 = s + 2$,

$$R_3 = \binom{s'}{r+1}\Big/\binom{s}{r} \leq 10$$

since $s \leq 3r/2$, and

$$R_4 = (n-r-1)!/(n-r)! = (n-r)^{-1} \leq 2n^{-1}.$$

Thus, $B(r+1)/B(r) \leq 1$ for $k \geq 4$ and $n \geq 100$. Consequently,

$$\sum_{r \leq n/2} B(r) \leq nB(1) \leq n^3(2(n-1)!)^k < (n!)^k$$

for $k \geq 4$ and $n \geq 100$. It follows that not all k-tuples of bijections are bad, provided that $k \geq 4$ and $n \geq 100$.

Now, if a graph X corresponds to a good k-tuple, then every $V_0 = V_0' \cup V_0''$ for $V_0' = V_0 \cap V'$ and $V_0'' = V_0 \cap V''$, clearly has

$$\operatorname{card}(\partial V_0) \geq \operatorname{card}(\partial V_0') - \operatorname{card}(V_0'').$$

If $\operatorname{card} V_0' \geq \operatorname{card} V_0''$, then this makes $\operatorname{card} \partial V_0 \geq \operatorname{card}(V_0)/4 - 1$, which settles the matter. Moreover, if $\operatorname{card} V_0'' > \operatorname{card} V_0'$, then we observe that the computation above shows that the number of bad k-tuples of bad bijections $V' \to V''$ is *much* smaller than $(n!)^k$, and so there is a good bijection for which all k inverse maps $V'' \to V'$ are also good. This makes $\operatorname{card}\partial V_0 \geq \operatorname{card}(V_0)/4 - 1$ in any case.

Isn't it amazing that such a crude computation produces the desired conclusion? And even more amazingly, it does so just barely, leaving us with a pretty big residue $\approx (n!)^k/n^{k-3}$ of bad graphs out of a total of $(n!)^k$. (See [Lub] and [Alon] for information and references.)

H. Observable distance $\underline{H}_\lambda$ on the space $\mathcal{X}$ and concentration $X^n \to X$

$3\frac{1}{2}$.45. The metrics $H_\lambda\mathcal{L}\iota_1$ on d functions and $\underline{H}_\lambda\mathcal{L}\iota_1$ on mm-spaces. We are going to introduce a new metric in the space $\mathbf{X}$ of isomorphism classes of mm spaces, with respect to which the convergence of spaces $X^n \subset \mathbf{X}$ to a one point space would correspond to $\operatorname{ObsDiam}(X^n) \to 0$. We follow the same general scheme as in the definition of $\underline{\square}_\lambda$ and therefore start by defining a new metric on the space D of functions $d\colon X \times X \to \mathbb{R}_+$, where X is a measure space and d are positive symmetric functions satisfying the triangle inequality. Each d on $X \times X$ defines a distinguished subset of functions on X, namely those which are 1-Lipschitz with respect to d, call it $\mathcal{L}ip(d)$, and we set

$$H_\lambda\mathcal{L}\iota_1(d,d') = \operatorname{dist}_H(\mathcal{L}ip_1(d), \mathcal{L}ip_1(d')),$$

where the Hausdorff distance dist_H between subsets of measurable functions on X is taken for the metric me_λ in this space of functions $X \to \mathbb{R}$. It is clear that $H_\lambda\mathcal{L}\iota$ is a metric. Indeed, the map $d \mapsto \mathcal{L}_1(d)$ is injective, since

$$d(x,y) \leq \varepsilon \qquad \Leftrightarrow \qquad \exists f \in \mathcal{L}ip_1(d) \text{ such that } f(x,y) \leq \varepsilon.$$

It is also clear that the map $d \mapsto \mathcal{L}ip_1(d)$ is (uniformly) continuous with respect to the metrics $\square_\lambda$ on D and dist_H (on subsets of functions with the metric me_λ) for every $\lambda > 0$ (where we recall that the metrics $\square_\lambda$ are mutually equivalent for all $\lambda > 0$, and that the metrics me_λ are also mutually equivalent for positive λ). It follows that the identity map

$$(D, \square_\lambda) \longrightarrow (D, H_\lambda\mathcal{L}\iota)$$

is (uniformly) continuous. Thus, the two metrics $\square_\lambda$ and $H_\lambda\mathcal{L}\iota$ are topologically equivalent on each *compact* subset in D.

We define the distance $\underline{H}_\lambda \mathcal{L}\iota$ between two mm spaces X and X' of mass m with the use of a measurable parametrization of X and X' by the segment $[0, m]$, where we minimize the $H_\lambda \mathcal{L}\iota$ distance in $D = D([0, m])$ of the pullbacks d and d' of the dist_X and $\mathrm{dist}_{X'}$ to $[0, m]$ over all such parametrizations. One sees easily that this $\underline{H}_\lambda \mathcal{L}\iota_1$ is indeed a metric, in particular

$$\underline{H}_\lambda \mathcal{L}\iota_1(X, X') = 0 \qquad \Rightarrow \qquad X \text{ is isomorphic to } X'.$$

In fact, $\underline{H}_\lambda \mathcal{L}\iota_1$ is topologically equivalent to $\underline{\square}_\lambda$ on each *compact* subset in $\mathcal{X}_m$, and the implication above follows from the corresponding property of $\underline{\square}_\lambda$. (In fact, one only needs to check this for finite spaces X, since these are $\underline{\square}_\lambda$-dense in $\mathbf{X}$.) Finally, we extend $\underline{H}_\lambda \mathcal{L}\iota_1$ to spaces of different masses m and m' as in Ch. $3\frac{1}{2}$.B$_+$.

Now we recall that the observable diameter of X carries essentially the same message as the me_λ-diameter of the space $\mathcal{L}ip_1(X)$ for some $\lambda > 0$, e.g., $\lambda = 1$, and thus $\mathrm{ObsDiam}_{\mathbb{R}} X^n \xrightarrow[n\to\infty]{} 0$ is equivalent to $\underline{H}_1 \mathcal{L}\iota_1$-convergence of the sequence X^n to a one point space (where we could equally use $\underline{H}_\lambda \mathcal{L}\iota_1$ for a given $\lambda > 0$).

$3\frac{1}{2}$.46. Concentration $X^n \to X$. The word "concentration" refers to $\underline{H}_1 \mathcal{L}\iota_1$-convergence, and we want to get a geometric picture of this, where X consists of more than one point (which is similar to what happens to the Hausdorff collapse in 3.35).

Example: Set $X^n = Y \times Z^n$ with $\mathrm{ObsDiam}\, Z^n \to 0$. Let us show that indeed

$$\mathrm{ObsDiam}\, Z^n \to 0 \qquad \Rightarrow \qquad X^n \xrightarrow{\underline{H}_1 \mathcal{L}\iota_1} Y,$$

where we assume, to avoid notational complications, that $\mu(Z^n) = 1$ and that neither X nor Z^n have atoms. We denote by $p_n : X^n \to Y$ the projection and measurably identify X^n with $[0, m]$ for $m = \mu(X)$. Then our parametrizations of X^n and Y coincide with the identity map $X^n \to X^n$ and with the projection $p_n\colon X^n \to Y$ correspondingly. Let us evaluate the Hausdorff distance between the subspace

$$\mathcal{L}ip_1(X^n) = \{\text{functions on } X^n \text{ with the } \mathrm{me}_1\text{-metric}\}$$

and the pullback $p^*(\mathcal{L}ip_1(X))$. Since p is 1-Lipschitz,

$$p^*(\mathcal{L}ip_1(Y)) \subset \mathcal{L}ip_1(X^n),$$

and all we have to show is that

$$\mathcal{L}ip_1(X^n) \subset U_\varepsilon(p^*(\mathcal{L}ip_1(Y))),$$

where U_ε stands for the ε-neighborhood and where ε should be small for small ObsDiam(Z^n). We observe that $p^*(\mathcal{L}ip(Y))$ consists of exactly those 1-Lipschitz functions on X^n which are constant on the fibers of the map p_n, i.e., on each $x \times Z^n \subset X^n = Y \times Z^n$, and we project $\mathcal{L}ip_1(X^n)$ to $p^*(\mathcal{L}ip_1(Y))$ by taking the Levy mean $\overline{f}(x)$ of each 1 Lipschitz function $f(x) = f(y,z)$ in the z-variable. The me_1-displacement for this projection is controlled by the Levy radius of Z^n as in $3\frac{1}{2}.32$–$3\frac{1}{2}.36$

$$\mathrm{LeRad}(Z^n; -\kappa) \le -\rho \qquad \Rightarrow \qquad \mathrm{me}_1(f, \overline{f}) \le \max(\kappa, \rho).$$

This yields a concentration $X^n \to Y$ under our assumption that $\mathrm{ObsDiam}_{\mathbb{R}}(Z^n) \to 0$, since it is equivalent to $\mathrm{LeRad}(Z^n) \to 0$.

Let us isolate the relevant properties of the projections p_n. We want to comprehend how the general measurable maps $p_n : X^n \to Y$ which, so to speak, *enforce* concentration of X^n to Y, meaning

$$\mathrm{dist}_H(\mathcal{L}ip_1(X^n), p_n^*(\mathcal{L}ip_1(Y)) \xrightarrow{n\to\infty} 0,$$

where clearly a so enforced concentration implies the ordinary one, i.e., the $\underline{H}_\lambda \mathcal{L}\iota_1$-convergence of X^n to Y, provided that the maps $p_n : X^n \to Y$ are measure-preserving. In what follows, we abbreviate $\mathcal{L} = \mathcal{L}ip_1(X)$, $\mathcal{L}_n = \mathcal{L}(X^n)$, and $\mathcal{L}^* = p^*(\mathcal{L}ip_1(Y)) \subset \mathcal{L}$ for a given map $p \colon X \to Y$. And if we deal with a sequence $p_n : X^n \to Y$, we write $\mathcal{L}_n^*$ for $p_n^*(\mathcal{L}ip_1(Y)) \subset \mathcal{L}_n$. Our objective is to bound the Hausdorff distance between $\mathcal{L}$ and $\mathcal{L}^*$ in the space of measurable functions on X with the me_1 metric. Thus, we want to bound ε for which

I. $\mathcal{L}^* \subset U_\varepsilon(\mathcal{L})$

II. $\mathcal{L} \subset U_\varepsilon(\mathcal{L}^*)$,

where U_ε denotes the ε-neighborhood in the me_1 metric.

(A) Giving a name to the inclusion $\mathcal{L}^* \subset U_\varepsilon(\mathcal{L})$. First we notice that

$$\mathcal{L}^* \subset \mathcal{L} \qquad \Leftrightarrow \qquad \text{the map } p \colon X \to Y \text{ is 1-Lipschitz.}$$

We then *define* a map p to be *1-Lipschitz up to (an additive error)* ε if $\mathcal{L}^* \subset U_\varepsilon(\mathcal{L})$. If we spell this out for a sequence of maps $p_n : X^n \to Y$, then we come to the following

$\mathrm{I_A}$ Almost tautology. *One has $\mathcal{L}^* \subset U_{\varepsilon_n}(\mathcal{L})$ with $\varepsilon_n \to 0$ for $n \to \infty$ if and only if the inequality*

$$\mathrm{dist}(p(x), p(x')) \le \mathrm{dist}(x, x') + \delta_n \tag{$*$}$$

holds for all x and x' in subsets $V_n \subset X^n$, where $\delta_n \to 0$ and $\mu(X^n \backslash V_n) \to 0$ for $n \to \infty$.

(This agrees with our earlier definition of "1-Lipschitz up to ε.")

(B) Concentration and parallelism between the fibers. The inclusion $\mathcal{L} \subset U_\varepsilon(\mathcal{L}^*)$ says that every 1-Lipschitz function f on X admits an ε-small perturbation f' (where "ε-small" refers to the metric me_1) such that

II$_B$. The function f' is constant on the fibers of $p\colon X \to Y$, i.e., it equals $g \circ p$ for some function g on Y.

III$_B$. Moreover, one can choose this f' in such a way that g is 1-Lipschitz on Y.

These II$_B$ and III$_B$ require somewhat different properties of p. The first of the two implies, as we shall see presently, that (most of) the fibers $X_y = p^{-1}(y) \subset X$, $y \in Y$, are roughly *ε-concentrated*, i.e., have observable diameters $O(\varepsilon)$. In fact, we shall state and prove this not for the fibers themselves, but for the *δ-fibers* $X_B = p^{-1}(B) \subset X$, with δ-small underlying open subsets $B \subset Y$. The advantage of this is a guaranteed *positive* measure $\mu_X(X_B) > 0$, which facilitates our discussion.

On the other hand, condition III$_B$ makes these (concentrated) fibers X_y and $X_{y'}$ mutually "almost parallel" with distance $\mathrm{dist}(y, y') + O(\varepsilon)$ between them. In fact, due to the concentration of the fiber X_y, the distance $\mathrm{dist}(x, y')$ is close to a constant on most of X_y, and we claim that this constant is close to $\mathrm{dist}(y, y')$. Here again, it is more convenient to state and prove everything for the δ-fibers.

We conclude this discussion by noting that the ε-perturbation $f \rightsquigarrow f'$ which makes f constant on the fibers can be achieved by taking the means (averages) of f along the fibers. Actually, we prefer Levy's mean rather than the average in order to block the possible summability problem. Besides, we shall use the δ fibers with $\delta > 0$ to have enough elbow room with the measure > 0.

(C) Relative concentration of X_n over Y. We say that a sequence of maps $p_n : X^n \to Y$ *effectuates relative concentration of X^n over Y as* $n \to \infty$ if

$$\limsup_{n \to \infty} \mathrm{ObsDiam}(X_B^n, -\kappa) \leq \mathrm{diam}\, B$$

for every bounded open subset $B \subset Y$ and each $\kappa > 0$, where $X_B^n = p_n^{-1}(U)$.

Observation: *If measure-preserving maps p_n enforce concentration of X^n to Y, then they also effectuate relative concentration of X^n over Y.*

Proof. We must show that every 1-Lipschitz function $f = f_n$ on X_B^n has oscillation $\leq \delta + \varepsilon_n$ for $\delta = \operatorname{diam} B$ and $\varepsilon_n \to 0$ outside a subset of measure κ in X_B^n. To show this, we extend f to a 1-Lipschitz function $\tilde{f}$ on all of X^n, where it is ε_n-close to $\tilde{f}' = g_n \circ p_n$ for g_n being 1-Lipschitz on Y and with the ε_n-closeness referring to the me_1-metric. Clearly, the restriction $f_n' = \tilde{f}_n'|_{X_B}$ has oscillation at most $\delta + o(n)$ away from a subset of measure $o(1)$, since the maps f_n are 1-Lipschitz with additive errors $o(1)$ (where $o(1)$ means "going to zero as $n \to \infty$"). On the other hand, f_n' is ε_n-close to f on X_B^n for $\varepsilon_n \to 0$, and the proof follows.

(D) Closeness and parallelisms between the fibers. Let us define the following *κ-distance* between two measurable subsets A_1 and A_2 in an mm space X

$$\operatorname{dist}_+(A_1, A_2; +\kappa) = \sup \operatorname{dist}(A_1', A_2'),$$

where $A_1' \subset A_1$ and $A_2' \subset A_2$ run over the pairs of subsets such that

$$\mu(A_1') \geq \kappa \qquad \text{and} \qquad \mu(A_2') \geq \kappa,$$

and where we recall that

$$\operatorname{dist}(A_1', A_2') = \inf_{\substack{a_1' \in A_1' \\ a_2' \in A_2}} \operatorname{dist}(a_1', a_2').$$

Observation: *Let $p_n : X^n \to Y$ be a measure preserving map as before which enforces concentration of X^n to Y, take $\kappa > 0$ and consider subsets $A_1 = A_1(n)$ and $A_2 = A_2(n)$ in X^n of measures $\geq \kappa$. Denote by B_1 and B_2 their p_n-images in Y and let*

$$\delta = \max(\operatorname{diam} B_1, \operatorname{diam} B_2).$$

Then

$$\operatorname{dist}_+(A_1, A_2; +\kappa) \leq \operatorname{dist}(B_1, B_2) + 2\delta + \varepsilon_n,$$

where $\varepsilon_n \to 0$ as $n \to \infty$. (In fact, $\varepsilon_n = \varepsilon_n(\kappa)$ admits a bound in terms of $\operatorname{dist}_H(\mathcal{L}_n, \mathcal{L}_n^*)$ essentially independently of the particular subsets we choose.

Proof. Suppose we have subsets $A_i' \subset A_i$, $i = 1, 2, \ldots$, of measure $\geq \kappa > 0$ and with

$$\operatorname{dist}(A_1', A_2') \geq \Delta = \operatorname{dist}(B_1, B_2) + 2\delta + \alpha$$

for some $\alpha > 0$ independent of n. Then there is a 1-*Lipschitz* function f on X which equals zero on A_1' and Δ on A_2'. But this f cannot be contained in the ε-neighborhood $U_\varepsilon(\mathcal{L}_n^*)$ if ε is sufficiently small and $\mathcal{L}_n^* \subset U_\varepsilon(\mathcal{L}_n)$ (i.e., p is 1-Lipschitz up to ε), since every function g on Y such that $g|_{B_1''} = 0$ for some $B_1'' \subset B_1$ and $g|_{B_2''} = \Delta$ for some $B_2'' \subset B_2$ has Lipschitz constant at least

$$\Delta/(\text{dist}(B_1, B_2) + 2\delta) \geq 1 + \alpha/\Delta,$$

provided that B_1'' and B_2'' are nonempty.

Concentration remark. If δ is small and $A_i = p_n^1(B_i)$, $i = 1, 2$, then these A_i are almost δ-concentrated for large n, and so the κ distance well-approximates the *actual distance* between the δ-fibers A_i which is given by either of the two functions $x_1 \mapsto \text{dist}(x_1, A_2)$ for $x_1 \in A_1$ and $x_2 \mapsto \text{dist}(x_2, A_2)$ for $x_2 \in A_2$. Namely, if $\kappa \in [\kappa_0, \underline{\kappa}]$ for a fixed $\kappa_0 > 0$ and $\underline{\kappa} \leq \frac{1}{2}\min(\mu(B_1), \mu(B_2))$, then these functions are close (with respect to the me_1-metric) to constants and the observation above implies that the constants are within the range $\text{dist}(B_1, B_2) \pm \delta$. So, we can say that the δ-fibers of the maps $p_n \colon X^n \to Y$ are *essentially parallel* for $n \to \infty$.

Example: If $p \colon X \to Y$ is a Riemannian submersion, e.g., $S^{2n+1} \to \mathbb{C}P^n$, or the projection $X = Y \times Z \to Y$, then the fibers are honestly parallel, i.e., the function $\text{dist}(x, X_{y_0})$ is constant on each fiber $X_y = p^1(y) \in X$.

Characteristic distances between the fibers. If A_1 and A_2 are two sufficiently concentrated fibers of $p \colon X \to Y$ (or δ-fibers with a small δ), then the distance function $\text{dist}(x_1, x_2)$ for $(x_1, x_2) \in A_1 \times A_2$ is close to its Levy mean (or the ordinary mean, i.e., the average, where this $\text{dist}(x_1, x_2)$ must be summable) and this mean, call it $\text{Ledi}(A_1, A_2)$, is a certain function on $Y \times Y$

$$\text{Ledi}(y_1, y_2) = \text{Ledi}(X_{y_1}, X_{y_2}),$$

(where one can replace the points y_i by small δ-balls $B(y_i, \delta)$ if one is bothered by the issue of how to properly define a *nonzero* measure in the fibers). This Ledi is usually significantly greater than the ordinary distance $\text{dist}_Y(y_1, y_2)$. For example, $\text{Ledi}(y, y)$ equals the characteristic size (or the average diameter) of the fiber X_y, which may be far from zero. To get a clear perspective, one should look again at $X = Y \times Z$ with a concentrated fiber, e.g., $Z = S^n$ with large n, where

$$\text{Ledi}(y_1, y_2) \approx \sqrt{\text{dist}^2(y_1, y_2) + \pi^2/4},$$

if we use the ℓ_2-product metric in $Y \times Z$. More generally, one can modify this picture with a *warping factor*, i.e., a function $w = w(y)$ which scales the fiber $X_y = y \times Z$ by $w(y)$, replacing it by $y \times w(y)Z$. In this case, the function $\text{Ledi}(y_1, y_2)$ behaves in a somewhat more complicated way.

Exercise: Evaluate $\mathrm{Ledi}(y_1, y_2)$ for $X = [0,1] \times wS^n$ for a given function $w(y)$, $y \in [0,1]$, e.g., for $w(y) = e^{ay}$ with large a.

$3\frac{1}{2}$.47. A criterion for enforced concentration. We are now about to state and prove an inverse to the tautology and observations above for a given sequence of measure preserving maps $p_n \colon X^n \to Y$.

Proposition: *Let p_n satisfy the following three conditions:*

I. *The p_n's are 1-Lipschitz up to (additive errors) $\varepsilon_n \to 0$ for $n \to \infty$.*

II. *The δ-fibers are highly concentrated for small $\delta > 0$ and large $n = n(\delta)$. Namely, for every $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\mathrm{ObsDiam}(X^n_B, -\kappa) \leq \varepsilon$$

for every δ-ball $B \subset Y$, every $\kappa > 0$, and all sufficiently large $n \geq n(\varepsilon, \delta, \kappa)$.

III. *the κ-distance between the δ-fibers approaches the distance between the underlying balls for suitable values of δ and κ. This means that for every two points y_1 and y_2 in Y and each $\varepsilon > 0$, there exist $\delta > 0$ such that the κ-distance between $A_1 = p_n^{-1}(B(y_1, \delta))$ and $A_2 = p_n^{-1}(B(y_2, \delta))$ satisfies*

$$\mathrm{dist}_+(A_1, A_2; +\underline{\kappa}) \leq \mathrm{dist}(y_1, y_2) + \varepsilon',$$

(where the opposite inequality $\mathrm{dist}_+(A_1, A_2; +\underline{\kappa}) \geq \mathrm{dist}(y_1, y_2) - \varepsilon'$ *follows from I) for $\underline{\kappa} = \frac{1}{2}\min(\mu B(y_1, \delta), \mu B(y_2, \delta))$ (with $B(y, \delta)$ denoting the gd-ball in Y around $y \in Y$) and all $n \geq n(\varepsilon', \delta, \underline{\kappa})$.*

Then the p_n enforce the concentration of X^n to Y.

Idea of the proof. Our task is to slightly perturb a 1-Lipschitz function f on X^n to $f' = g \circ p_n$, where g is 1-Lipschitz on Y. And it is sufficient to have g only 1-Lipschitz up to an additive error $o(1)$ for $n \to \infty$, since this can be further perturbed to the desired honestly 1-Lipschitz function. Now, we make up such g by Levy-averaging f along the δ-fibers of p_n. Namely, we cover most of Y by small δ-balls $B_i = B_{i,n}$, $i = 1, \dots, N$, where $\delta \leq \delta_n \to 0$ for $n \to \infty$ and $N = N(n) \to \infty$, while

$$\mu(X^n \setminus \bigcup_{i=1}^{N} B_i) \leq \kappa_n \xrightarrow{n \to \infty} 0$$

for the measure $\mu = \mu_n$ on X^n. To make everything work right, we must assume that δ_n and κ_n decay very slowly with $n \to \infty$ and $N(n)$ grows very slowly. Actually, we choose our balls for some n and then we allow n to grow for a long time before we switch to a new set of balls B_i. Then we again let the balls stay still for a long time as X^n concentrates more and more to Y as n increases. In other words, $\mathrm{dist}_H(\mathcal{L}_n, \mathcal{L}_n^*)$ must be kept relatively small, i.e., less than some $\varepsilon_n = \varepsilon_n\{B_i\}$, where, in particular, this ε_n is much smaller than δ_n, κ_n and $N^{-1}(n)$. It follows from **II** that the δ-fibers $A_i = p_n^{-1}(B_i)$ are highly concentrated, and so the Levy means g_i of $f|_{A_i}$ are close to $f|_{A_i}$. Then **III** and the König matching theorem (see Ch. $3\frac{1}{2}$.C) imply that

$$|g_i - g_j| \leq \mathrm{dist}(B_i, B_j) + o(1)$$

as $n \to \infty$. Indeed, the concentration of A_j implies high concentration of the distance function $a_j \mapsto \mathrm{dist}(a_j, A_i')$ on A_j for every subset $A_i' \subset A_i$. Then **III** shows that the ρ-neighborhood in X of a subset $A_i' \subset A_i$ with $\mu(A_i') \geq x' > 0$ will contain almost full measure of A_j, provided that

$$\rho \geq \mathrm{dist}(B_i, B_j) + 4\delta + \varepsilon' + \varepsilon = \mathrm{dist}(B_i, B_j) + o(1)$$

for all sufficiently large $n \geq n(\delta, \varepsilon, \varepsilon', x') \to \infty$. The same equality applies to subsets $A_j' \subset A_j$ and leads to a bound on the di_λ-distance between the normalized measures $\mu_i = \mu^{-1}(A_i)\mu|_{A_i}$ and $\mu_j = \mu^{-1}(A_j)\mu|_{A_j}$. Consequently, we can bound the transportation metric via the König theorem. In fact, what we get and what we truly need is that the ρ_{ij}-neighborhood in X of a subset $A_i' \subset A_i$ with $\mu(A_i') \geq \kappa' > 0$ will contain almost full measure of A_j provided that

$$\begin{aligned} \rho_{ij} &\geq \mathrm{dist}(B_i, B_j) + 4\delta + \varepsilon' + \varepsilon \\ &= \mathrm{dist}(B_i, B_j) + o(1) \end{aligned}$$

for all sufficiently large $n \geq n(\delta, \varepsilon, \varepsilon', \kappa') \to \infty$. Then our bound on $|g_i - g_j|$ follows via the following, obvious

(5′) Lemma: *Let μ_1 and μ_2 be two probability measures on a metric space X, such that* $\mathrm{ObsDiam}(\mu_i, -1/4) \leq \Delta$, $i = 1, 2$, *and μ_1 admits a ρ-transportation to μ_2 of deficiency $\kappa \leq 1/4$. Then the Levy means of each 1-Lipschitz function f on X with respect to μ_1 and μ_2, call them g_1 and g_2, satisfy*

$$|g_1 - g_2| \leq \rho + 2\Delta.$$

Warning: The Levy mean of f with respect to a measure μ is not uniformly continuous in μ as is seen by moving a small central atom away from the center.

Here, the transportation of this tiny mass moves the Levy mean of the coordinate t from t_0 to t_1.

Now, the bound $|g_i - g_j| \leq \text{dist}(B_i, B_j) + o(1)$ shows that the function $g(y)$ assigning to each $y \in \bigcup_{i=1}^N B_i$ the value g_i for some $B_i \ni y$ is 1-Lipschitz up to an additive error $o(1)$, and the proof follows. We suggest that the reader fill in the details by setting in order the qualifiers and chasing all the ε's and δ's. (This will go smoothly unless we missed something in our **II** and/or **III**.)

$3\frac{1}{2}$.48. On relativization of concentration invariants. The above discussion suggests a definition of a *δ-invariant* attached to a Borel map $f: X \to Y$ and some geometric invariant Inv as follows.

$$\delta\text{-Inv}(X/Y) = \sup_{X'} \text{Inv}(X')$$

for X' running over all δ-fibers of X.

Example: If $X = Y \times Z$, then the *δ-observable diameter of X over Y* is roughly bounded by $\delta + \text{ObsDiam}(Z)$. In fact, this is literally true for the central δ-radii,

$$\delta\text{-ObsCRad}_B(X/Y; -\kappa) \leq \delta + \text{ObsCRad}_B(Z; -\kappa)$$

for every Banach space B (e.g., $B = \mathbb{R}$), and

$$\delta\text{-LeRad}(X/Y; -\kappa) \leq \delta + \text{LeRad}(Z; -\kappa).$$

This follows from the discussion above.

(6′) Concentration modulo a partition. There is another relativization of our concentration invariants which applies to an arbitrary measurable partition Π of X into subsets X_y for $y \in Y = X/\Pi$, where we stay blind to the geometry of Y. Recall that there exist canonical probability measures in almost all fibers X_y (where Y carries the quotient measure), and so one can average every L_1-function f along the fibers. Then one defines $\text{ObsCRad}_{\mathbb{R}}(X/\Pi; -\kappa)$ as the infimal ρ such that each 1-Lipschitz function $f(x)$ is ρ-close to its fiberwise average $\overline{f}_\Pi(x)$ outside a subset of measure κ,

$$\mu\{x \in X : |f(x) - \overline{f}_\Pi(x)| \leq \rho\} \leq \kappa.$$

One can also use the fiberwise Levy mean $\check{f}_\Pi(x)$ of $f(x)$ and thus define $\text{LeRad}(X/\Pi; -\kappa)$, which has the advantage of not requiring f to be L_1.

One can go further and look for the infimal $\rho = \rho(x)$ such that for each 1-Lipschitz function f, there exists a function f' on X which is constant on each fiber and satisfies

$$\mu\{x \in X : |f(x) - f'(x)| \leq \rho\} \leq \kappa,$$

which relativizes (a slightly modified) ObsDiam. Here, there is an extra modification achieved by restricting to 1-*Lipschitz* functions f' constant on the fibers. We suggest that the reader study the relations among all these relative radii, generalizing what was done in the absolute case in which Y consisted of a single point. We also suggest looking in a similar vein at the space $\mathcal{L}ip_1(X)/\operatorname{const}_\Pi$, where const_Π denotes the space of functions constant in the fibers X_y.

Warning: One could bring the δ-radii closer to the radii of X/Π above by evaluating them for the δ-fibers $X' \subset X$ with *normalized* measures and then sending $\delta \to 0$. Thus, we recapture the (radii of the) fibers themselves.

This works very well for $X = Y \times Z \xrightarrow{p} Y$, but it fails for more general p such as the Hopf fibration $p : S^{2n+1} \to \mathbb{C}\mathrm{P}^n$. Here, the δ-radii with normalized measures are about $\pi/2 = \frac{1}{2} \operatorname{diam} S^1$ for δ approaching zero, while the radii of S^{2n+1}/Π for the partition into the fibers $p^1(y)$, $y \in \mathbb{C}\mathrm{P}^n$, can only be *smaller* than the corresponding absolute radii of S^{2n+1}. The latter, as we know, go to zero as $n \to \infty$, while the δ-radii (with the normalized measure!) remain $\geq \pi/2$. The major reason why this happens is the *non*-precompactness of the family $\{S^{2n+1}\}_{n\to\infty}$ in the $\underline{\Box}_1$-topology, but the geometry of the fibration is also important (compare [Gro–Mil]$_{\mathrm{TAI}}$).

Remark: The above shows that the concentration of X relative to a partition Π may be due to the concentration of all of X rather than to the concentration of the fibers. But, for example, $[0, 1] \times S^{2n+1}$ partitioned into the Hopf circles does not fit either of these two possibilities. In fact, this may very well be the general case: the concentration relative to Π can may always come from a fiberwise concentration of a coarser partition Π' of X.

$3\frac{1}{2}$.49. Concentration $X^n \to Y$ via maps $p_n : X^n \to Y$. The discussion above gives a criterion for concentration (i.e., $\underline{H}_1\,\mathcal{L}\iota_1$-convergence) X^n to Y in terms of suitable maps $p_n : X^n \to Y$. Now we want to show that whenever a sequence of mm spaces X^n concentrates to some Y as $n \to \infty$, there exist maps $p_n : X^n \to Y$ behaving in a manner similar to the projections $Y \times Z^n \to Y$, where $\operatorname{ObsDiam} Z^n \to 0$. Here is our

Proposition: *A sequence X^n concentrates to Y if and only if there exist Borel measurable maps $p_n : X^n \to Y$ with the following properties:*

0.* *The pushforward measures $(p_n)_*(\mu_{X^n})$ converge to μ_Y. (Here we assume that our measures are finite, $\mu(X^n) < \infty$ as well as $\mu(Y) < \infty$, and so the convergence of the measures implies that the masses $m_n = \mu(X^n)$ converge to $\mu(Y)$.)*

I*. *The maps p_n are 1-Lipschitz up to ε_n-errors, where $\varepsilon_n \to 0$ as $n \to \infty$.*

II*. *The fibers of the maps concentrate to points in the following sense*

$$\limsup_{n\to\infty}(\delta\text{-ObsDiam}(X^n/Y; -\kappa)) \leq \delta$$

for all positive δ and κ.

II*. *For every open subset $B \subset Y$ of diameter D, the functions $d(x) = \mathrm{dist}(x, p_n^{-1}(B))$ and $\underline{d}(x) = \mathrm{dist}(p(x), B)$ eventually become D-close in the metric me_1 on functions on X, i.e.*

$$\limsup_{n\to\infty} \mathrm{me}_1(d, \underline{d}) \leq D.$$

Sketch of the proof. The "if" claim follows from the previous discussion, at least for $(p_n)_*(\mu_{X^n}) = \mu_Y$, and the general case requires a minor adjustment. (In fact, if $\mu(X^n) = \mu(Y)$ and the spaces X^n have no atoms, then one can easily perturb the maps p_n, gaining the relation $(p_n)_*(\mu_{X^n}) = \mu_Y$ without disturbing I*, II*, and III*. A more interesting part is "only if," which needs an actual construction of maps p_n given X^n concentrating to Y. To do this, let us look at what can be seen of a given mm space (X, μ, dist) on the Euclidean "screen" $\mathbb{R}^N$ with the ℓ_∞ metric. Our "observables" are 1-Lipschitz maps $F: X \to \mathbb{R}^N$, i.e., N-tuples of 1-Lipschitz functions $(f_1, \dots, f_N)$, and what we see are the pushforward measures $F_*(\mu)$ on $\mathbb{R}^N$. We denote by $\mathcal{M}(X \xrightarrow{\mathrm{Lip}_1} \mathbb{R}^N) \subset \mathcal{M}(\mathbb{R}^N)$ the subset of these measures for all 1 Lipschitz maps $F: X \to (\mathbb{R}^N, \|\cdot\|_{\ell_\infty})$, where $\mathcal{M}(\mathbb{R}^N)$ denotes the space of all Borel measures on $\mathbb{R}^N$ equipped with the di_1-metric.

(A) Observational criterion for concentration. *The following two conditions are equivalent:*

(*i*) *The X^n concentrate to Y as $n \to \infty$.*

(*ii*) *For every $N = 1, 2, \dots$, the subsets $\mathcal{M}(X^n \xrightarrow{\mathrm{Lip}_1} \mathbb{R}^N) \subset \mathcal{M}(\mathbb{R}^N)$ Hausdorff-converge to $\mathcal{M}(Y \xrightarrow{\mathrm{Lip}_1} \mathbb{R}^N) \subset \mathcal{M}(\mathbb{R}^N)$ as $n \to \infty$, where the Hausdorff convergence refers to the di_1-metric in $\mathcal{M}(\mathbb{R}^N)$.*

Proof of $(i) \Rightarrow (ii)$. Recall that $\underline{H}_1\mathcal{L}\iota_1$-closeness between two spaces X and Y means that the spaces of 1-Lipschitz functions $\mathcal{L}ip_1(X)$ and $\mathcal{L}ip_1(Y)$ are Hausdorff close in the ambient space $\mathcal{F}$ of functions with the me_1-metric on the parametrizing space $[0,m]$. Thus, for every N-tuple of 1-Lipschitz functions f_i on X, one can find functions f_i' on Y which are me_1-close in $\mathcal{F}$ to the corresponding f_i, and therefore the pushforward measure $F_*'(\mu_Y)$ on $\mathbb{R}^N$ is di_1-close to the measure $F_*(\mu_X)$. This yields the implication $(i) \Rightarrow (ii)$.

Proof of $(ii) \Rightarrow$ **"only if"** . We use an approximation of Y by its Euclidean images $\underline{Y}_N$ for 1-Lipschitz maps $\Phi_N : Y \to (\mathbb{R}^N, \|\cdot\|_{\ell_\infty})$, where $\underline{Y}_N$ denotes $(\mathbb{R}^N, \underline{\mu}_N, \|\cdot\|_{\ell_\infty})$ with $\underline{\mu}_N = (\Phi_N)_*(\mu_Y)$ (see $3\frac{1}{2}$.17). Such a $\underline{Y}_N$ can be chosen arbitrarily $\underline{\square}_1$-close to X for large N, and thus we have approximate inverse maps $\Psi_N : \underline{Y}_N \to Y$. On the other hand, according to (ii), each $\underline{\mu}_N$ can be arbitrarily closely di_1-approximated by measures $\mu'(n) = \Phi_*'(\mu_{X^n})$ for large $n = n(N)$, where $\Phi' = \Phi_N'(\Phi)$ are some 1-Lipschitz maps $X^n \to \mathbb{R}^N$. Then one can slightly modify the approximate inverse maps, $\Psi_N \rightsquigarrow \Psi' : \mathbb{R}^N \to Y$, such that these $\Psi' = \psi_{N,r}'$ will be 1-Lipschitz up to $\varepsilon = \varepsilon(n,N)$ for the measure $\mu'(n)$ and send $\mu'(n)$ ε close to μ_Y on Y, where $\varepsilon \to 0$ as $N \to \infty$ and $n = n(N) \to \infty$. This can be done, for example, by δ-*regularizing* the map Ψ_N, i.e., by taking $\Psi_{N,\delta}$ which is defined in the δ-neighborhood $U_\delta = U_\delta(\text{Supp}(\underline{\mu}_N)$, and which sends each $u \in U_\delta$ close to the majority of points in the Ψ_N-image of the ball of radius δ around some point $v \in \text{Supp}\,\mu_N$ which is δ-close to u. We leave the quantification of this (which is fairly easy) to the reader.

Finally, we compose Φ' and Ψ', thus getting a maps $p = p_n = \Psi' \circ \Phi'$: $X^n \to Y$ which are 1-Lipschitz up to ε' and which send μ_{X^n} ε'-close to μ_Y, where $\varepsilon' \to 0$ as $n \to \infty$. These maps are good for 0^* and I^* in the proposition above, and now we claim that the automatically satisfy II^* and III^*.

(9) Lemma: *Let X^n be a sequence of mm spaces such that the metric spaces $(\mathcal{L}ip_1(X^n), \text{me}_1)$ Hausdorff-converge to the space $(\mathcal{L}ip_1(Y), \text{me}_1)$ for some mm space Y of finite mass. Let $p_n : X^n \to Y$ be a sequence of maps satisfying 0^* and I^* of our Proposition. Then II^* and III^* are satisfied as well.*

Proof. The properties 0^* and I^* show that the maps p^* almost isometrically, up to error $\varepsilon_n \to 0$, send $\mathcal{L}ip_1(Y)$ to the ε_n-neighborhood of $\mathcal{L}ip_1(X^n)$, while II_A^* and II_B^* say, in effect, that the ε_n'-neighborhood of the image $p^*(\mathcal{L}ip_1(Y))$ contains $\mathcal{L}ip_1(X^n)$ with $\varepsilon_n' \to 0$. One may ε_n-perturb the maps to make them send $\mathcal{L}ip_1(Y)$ to $\mathcal{L}ip_1(X_n)$, invoke the compactness of

$\mathcal{L}ip_1(Y)$, and apply the following trivial

Sublemma: *Let a sequence of metric spaces $\mathcal{L}_n$ Hausdorff-converge to a compact metric space $\mathcal{L}$, and let $q_n : \mathcal{L} \to \mathcal{L}_n$ be Borel maps such that*

$$\operatorname{dist}(q_n(\ell), q_n(\ell')) \geq \operatorname{dist}(\ell, \ell') - \varepsilon_n$$

for $\varepsilon_n \to 0$, $n \to \infty$. Then the images $q_n(\mathcal{L}) \subset \mathcal{L}_n$ are ε'_n-dense for $\varepsilon'_n \to 0$, $n \to \infty$.

In fact, this is seen by looking at maximal δ-separated nets $N_\delta \subset \mathcal{L}$ for $\delta = \delta_n \to 0$ as it is done in showing that every distance increasing self-mapping of a compact metric space $\mathcal{L}$ into itself is a surjective isometry.

Proof of $(ii) \Rightarrow (i)$**.** The Hausdorff convergence $\mathcal{M}(X^n \to \mathbb{R}^N) \to \mathcal{M}(Y \to \mathbb{R}^N)$ trivially implies that the "curvatures" $K_N(\mathcal{L}ip_1(X^n)) \subset M_N = \mathbb{R}^{N(N-1)/2}$ converge to $K_N(\mathcal{L}ip_1(X)) \subset M_N$ in the Hausdorff topology of subsets in M_N. This yields (see Ch. 3.D$_+$) the Hausdorff convergence $\mathcal{L}ip_1(X^n) \to \mathcal{L}ip_1(X)$ and the argument above using the lemma applies.

(10) Concentration $X^n \to Y$ as an ergodic decomposition. When n is "infinitely large" and X^n "infinitely concentrates" to Y, this essentially amounts to X^n being decomposed into "infinitely concentrated" pieces which are the fibers of a suitable map $X^n \to Y$. This is vaguely similar to the decomposition of a measurable transformation into ergodic components (compare 3.42(7)). On the other hand, if $\alpha : M \to M$ is an automorphism of a measure space M and $f : M \to X$ is a measurable map to a metric space X, one can look at the maps $f_n = (f, \alpha \circ f, \ldots, \alpha^{n-1} \circ f) : M^n \to X^n$, give X^n some (say ℓ_p) product metric dist $\times$ dist $\times \cdots \times$ dist and the measure $(f_n)_*(\mu \times \mu \times \cdots \times \mu)$. Then one can relate the ergodic decomposition of α to some concentration of X^n or rather of $\lambda_n X^n$ with $\lambda_n = o(1)$. We suggest that the reader look at specific instances of this picture).

I. The Lipschitz order on $\mathcal{X}$, pyramids, and asymptotic concentration

Recall the order relation $A \prec B$ for mm spaces referring to a 1 Lipschitz map $B \to A$ sending μ_B to $c\mu_A$ with $c \geq 1$ (see $3\frac{1}{2}$.15). Then for a sequence (of mm spaces) $X^n \in \mathcal{X}$, $n = 1, 2, 3, \ldots$, we define its *tail* $\mathcal{T}\{X^n\} \subset \mathcal{X}$ as the set of the limits of the $\underline{\square}_1$-convergent sequences $Y_n \in \mathcal{X}$, where each Y_n is Lipschitz dominated by some X^n, i.e., $Y_n \prec X^n$. Notice that every infinite subsequence X^{n_i} has a bigger tail $\mathcal{T}\{X^{n_i}\} \supset \mathcal{T}\{X^n\}$.

$3\frac{1}{2}$.50. Proposition. *If X^n concentrate to Y, then Y serves as the maximal element in the tail of $\{X^n\}$, i.e., $Y \in \mathcal{T}\{X^n\}$ and $Y \succ Y'$ for all $Y' \in \mathcal{T}\{X^n\}$. Conversely, if $\mathcal{T}\{X^n\}$ admits a (necessarily unique) maximal element Y, then X^n concentrate to Y provided that every infinite subsequence X^{n_i} of X^n has $\mathcal{T}\{X^{n_i}\} = \mathcal{T}\{X^n\}$.*

Proof. If X^n concentrate to Y, then Y equals the $\underline{\square}_1$-limit of some Euclidean 1-Lipschitz images of X^n approximating images of Y in $(\mathbb{R}^N, \|\cdot\|_{\ell_\infty})$ for $N \to \infty$ and $n = n(N) \to \infty$ as follows from (A). And every $Y' \in \mathcal{T}\{X^n\}$ is also a limit of such images of X^n in $\mathbb{R}^{N_n}$, and so it appears as the $\underline{\square}_1$-limit of approximating images of Y in $\mathbb{R}^{N_n}$ as again follows from (A). Then $Y \succ Y'$ by the $\underline{\square}$-continuity of the Lipschitz order. Notice that this equally applies to $Y' \in \mathcal{T}\{X^{n_i}\}$ for every infinite subsequence X^{n_i} and shows, in particular, that $\mathcal{T}\{X^{n_i}\} = \mathcal{T}\{X^n\}$.

Conversely, let $Y \in \mathcal{T}\{X^n\}$ dominate all tails $\mathcal{T}\{X^{n_i}\}$. Then it dominates the limits of images of X^{n_i} in $\mathbb{R}^N$ for $i \to \infty$ and every N. Since Y lies in $\mathcal{T}\{X^n\}$, all of its images in $\mathbb{R}^N$ can be approximated by those of X^n for $n \to \infty$, and so X^n concentrates by (A) again.

$3\frac{1}{2}$.51. Pyramids in $\mathcal{X}$ and their local Hausdorff convergence. A subset $\mathcal{P} \subset \mathcal{X}$ is called a *pyramid* if

(*a*) $\{X \in \mathcal{P}$ and $Y \prec X\} \Rightarrow Y \in \mathcal{P}$,

(*b*) for every pair $X, X' \in \mathcal{P}$, there exists $Y \in \mathcal{P}$ dominating X and X', i.e., such that $Y \succ X$ and $Y \succ X'$.

Example of $\mathcal{P} = \mathcal{P}_X$. This denotes the pyramid with *apex* $X \in \mathcal{X}$, i.e.,

$$\mathcal{P}_X \stackrel{\text{def}}{=} \{X' \in \mathcal{X} : X' \prec X\}.$$

Clearly, this $\mathcal{P}_X$ is a $\underline{\square}_1$-*closed* subset in $\mathcal{X}$ (see 3.41(3)(b)). Furthermore, if X and X' are $\underline{\square}_1$-closed in $\mathcal{X}$, then their pyramids $\mathcal{P}_X$ and $\mathcal{P}_{X'}$ are closed with respect to the Hausdorff metric corresponding to $\underline{\square}_1$ in $\mathcal{X}$. In particular, $\underline{\square}_1$-convergence $X^n \to Y$ implies the Hausdorff convergence $\mathcal{P}_{X^n} \to \mathcal{P}_Y$.

Next, we invoke the following general notion of *local Hausdorff convergence* of subsets $\mathcal{Y}_i$ in a metric space $\mathcal{X}$, e.g., in our $(\mathcal{X}, \mathrm{dist} = \underline{\square}_1)$. We assign to each subset $\mathcal{Y} \subset \mathcal{X}$ the distance function $d_\mathcal{Y}(X) = \mathrm{dist}(X, \mathcal{Y})$ for all $X \in \mathcal{X}$ and observe that the ordinary Hausdorff convergence $\mathcal{Y}_i \to \mathcal{Y}$ amounts to the *uniform* convergence of functions $d_{\mathcal{Y}_i} \to d_\mathcal{Y}$. Then we say that $\mathcal{Y}_i$ *locally (Hausdorff) converge to* $\mathcal{Y}$ *if*

(*a*) $d_{\mathcal{Y}_i}(Y) \to 0$ for each X where $d_\mathcal{Y}(X) = 0$, and

(*b*) $\liminf_{i\to\infty} d_{\mathcal{Y}_i}(X) > 0$ whenever $d_\mathcal{Y}(X) > 0$.

Example: If the functions $d_{\mathcal{Y}_i}$ converge to $d_{\mathcal{Y}}$ at each point $X \in \mathcal{X}$, then $\mathcal{Y}_i$ locally converge to $\mathcal{Y}$. Notice that since the functions $d_{\mathcal{Y}_i}$ are all 1-Lipschitz, pointwise convergence implies uniform convergence on each compact subset in $\mathcal{X}$. This motivates our "local" terminology. Also observe that if $\mathcal{X}$ is a *proper* metric space, i.e., all closed balls $B(R)$ in $\mathcal{X}$ for $R < \infty$ are *compact*, then (a) and (b) imply pointwise convergence of $d_{\mathcal{Y}_i}$ to $d_{\mathcal{Y}}$ (and we suggest that the reader make up an example showing that this is false for nonproper spaces).

Down-to-earth criterion for local convergence. *If $\mathcal{Y}$ is a closed subset in $\mathcal{X}$ and $\mathcal{Y}_i$ locally converge to $\mathcal{Y}$, then $\mathcal{Y}$ equals the set of limits of all convergent subsequences $Y_i \in \mathcal{Y}_i$. Furthermore, every convergent sequence $Y_j \in \mathcal{Y}_{i_j}$ with $i_j \to \infty$ for $j \to \infty$ has its limit contained in Y. Conversely, if $\mathcal{Y}$ consists of the limits of all convergent sequences $Y_i \in \mathcal{Y}_i$ and every convergent sequence $Y_j \in \mathcal{Y}_{i_j}$ has $\lim_{j\to\infty} Y_j \in \mathcal{Y}$, provided that $i_j \to \infty$, then the sequence $\mathcal{Y}_i$ locally converges to $\mathcal{Y}$.*

The proof is straightforward and is left to the reader. Now the proposition above yields the following theorem

"Pyramidal" criterion for concentration. *A sequence $X^n \in \mathcal{X}$, $n = 1, 2, \dots$, concentrates (i.e., $\underline{H}_1\mathcal{L}\iota_1$-converges) to Y if and only if the corresponding pyramids $\mathcal{P}_X \subset \mathcal{X}$ weakly-converge to $\mathcal{P}_Y$ for the metric $\underline{\square}_1$ in $\mathcal{X}$.*

$3\frac{1}{2}$.52. This suggests looking at *the space $\Pi \supset \mathcal{X}$ of all $\underline{\square}_1$-closed pyramids $\mathcal{P}$ in $\mathcal{X}$*, where $\mathcal{X}$ is embedded into Π via $X \mapsto \mathcal{P}_X$. Clearly, every closed pyramid $\mathcal{P}$ appears as the $\underline{\square}_1$-closure of an increasing union of $\mathcal{P}_{X_i}$ for $i = 1, 2, \dots$, where $X_1 \prec X_2 \prec \cdots \prec X_i \prec \cdots$, and so $\mathcal{P}_{X_1} \subset \mathcal{P}_{X_2} \subset \cdots \mathcal{P}_{X_i} \subset \cdots$. Thus, closed pyramids $\mathcal{P} \in \Pi \supset \mathcal{X}$ can be viewed as "ideal limits" of increasing sequences of mm spaces $X_1 \prec X_2 \prec \cdots$. In fact, it is convenient to consider more general

Asymptotic sequences. We call a sequence $X_i \in \mathcal{X}$ *asymptotic* if the corresponding pyramids $\mathcal{P}_{X_i}$ weakly converge to some (pyramid) $\mathcal{P} \in \Pi$. Intuitively, a sequence X_i is asymptotic if it displays definite asymptotic behavior as $i \to \infty$. Clearly, every sequence X_i admits an asymptotic subsequence. In fact, if $\mathcal{X}$ is an arbitrary Polish space, then every sequence $\mathcal{Y}_i \subset X$ admits a weakly Hausdorff-convergent subsequence. Moreover, there (obviously) exists a subsequence $\mathcal{Y}_{i_j}$ such that the distance functions $d_j = d_j \colon X \mapsto \mathrm{dist}(X, \mathcal{Y}_{i_j})$ converge at each point $X \in \mathcal{X}$ for $j \to \infty$.

Exercises: Show that the sequence $S^n(\sqrt{n})$, $n = 1, 2, \dots$, is asymptotic.

Compare the limit pyramids for the spheres $S^n(\sqrt{n})$ and the Gaussian spaces $(\mathbb{R}^n, e^{-\|x\|^2}\,dx)$. Study the limit pyramid for the cartesian powers $X^n = X \times \cdots \times X$ of a given mm space X, e.g., $X = [0,1]$, with the ℓ_p-product metric.

Local convergence via $\mathcal{X}(N,R)$. Let $\mathcal{X}(N,R) \subset \mathcal{X}$ denote the subset of those mm spaces X which correspond to measures in the R-ball in $(\mathbb{R}^N, \|\cdot\|_{\ell_\infty})$, i.e., of X isomorphic to $(\mathbb{R}^N, \|\cdot\|_{\ell_\infty})$ with a measure supported in the R-ball. Clearly $\mathcal{X}(N,R)$ is a compact subset in $\mathcal{X}$, and, by the approximation argument above, $\mathcal{X}$ equals the closure of the union $\bigcup_{N=1}^{\infty} \mathcal{X}(N,N)$. Furthermore, every pyramid $\mathcal{P}$ in $\mathcal{X}$ equals the $\underline{\square}_1$-closure of the union $\bigcup_{N=1}^{\infty}(\mathcal{P} \cap \mathcal{X}(N,N))$. It follows that

$\mathcal{P}_i$ weakly converge to $\mathcal{P}$ as $i \to \infty$ if and only if

$$\mathcal{P}_i \cap \mathcal{X}(N,N) \xrightarrow{\text{Hau}} \mathcal{P} \cap \mathcal{X}(N,N)$$

for each $N = 1, 2, \ldots$, where the Hausdorff convergence refers to that for subsets in the space $\mathcal{X}(N,N)$ with the $\underline{\square}_1$-metric.

$3\frac{1}{2}$.53. Asymptotic concentration and the $\underline{H}_1\,\mathcal{L}\iota_1$ completion $\overline{\mathcal{X}}$ of $\mathcal{X}$. We say that a sequence $X_i \in \mathcal{X}$ *asymptotically concentrates* if it Cauchy-converges with respect to the metric $\underline{H}_1\,\mathcal{L}\iota_1$ on $\mathcal{X}$. The proof of $(i) \Rightarrow (ii)$ in $3\frac{1}{2}$.50(A) shows that

if a sequence X_i asymptotically concentrates, then it is asymptotic.

Thus, every point Y in the $\underline{H}_1\,\mathcal{L}\iota_1$*-completion* of $\mathcal{X}$, denoted $\overline{\mathcal{X}}^{H\,\mathcal{L}\iota}$ is represented by a (unique) pyramid $\mathcal{P} = \mathcal{P}_Y \subset \mathcal{X}$ which is the weak limit of $\mathcal{P}_{X_i}$. Notice that $\mathcal{P}_Y$ is $\underline{\square}_1$-closed in $\mathcal{X}$. Therefore, it is $\underline{H}_1\,\mathcal{L}\iota_1$-closed. In fact,

every $\underline{\square}_1$-closed pyramid $\mathcal{P}$ in $\mathcal{X}$ is obviously weakly closed,

where "weakly closed" means that if pyramids $\mathcal{P}_i \subset \mathcal{P}$ weakly converge to some $\mathcal{P}'$, then $\mathcal{P}' \subset \mathcal{P}$. And, *weakly closed* $\Rightarrow$ $\underline{H}_1\,\mathcal{L}\iota_1$-closed by the previous discussion. It is also clear that

$$\mathcal{P}_Y = \mathcal{P}_{Y'} \qquad \Rightarrow \qquad Y = Y'$$

for arbitrary Y and Y' in $\overline{\mathcal{X}}^{H\,\mathcal{L}\iota}$. In other words, the map $\overline{\mathcal{X}}^{H\,\mathcal{L}\iota} \to \Pi$ given by $Y \mapsto \mathcal{P}_Y$ is one to-one. This is seen with the following criterion for asymptotic concentration $X_i \xrightarrow{\underline{H}_1\,\mathcal{L}\iota_1} Y$.

Denote by $\mathcal{M}_i = \mathcal{M}_i(N, R)$ the set of pushforward measures in the R-ball $B(R) \subset \mathbb{R}^N$ under 1-Lipschitz maps $X_i \to B(R)$, where $B(R)$ comes along with the ℓ_∞-metric. Then, we introduce a set of measures on $B(R) \subset \mathbb{R}^N$ associated to $\mathcal{P} = \mathcal{P}_Y$ and denoted by $\mathcal{M} = \mathcal{M}(Y; N, R)$, which equals the set of those μ for which $(B(R), \mu)$ is isomorphic to some $X \in \mathcal{P} \cap \mathcal{X}(R, N)$.

Observational criterion for asymptotic concentration. *A sequence $X_i \in \mathcal{X}$ asymptotically concentrates to $Y \in \overline{\mathcal{X}}^{H\,\mathcal{L}\iota}$ iff $\mathcal{M}_i$ Hausdorff-converge to $\mathcal{M}$ for every $N \in \mathbb{N}$ and $R \geq 0$.*

The proof is identical to that above (using $3\frac{1}{2}$.50(A)) and is left to the reader.

Finally, we close the circle by calling a pyramid $\mathcal{P}$ *concentrated* if the metric spaces $\mathcal{L} = (\mathcal{L}ip_1(X), \mathrm{me}_1)$ for all $X \in \mathcal{P}$ form a precompact family in the Hausdorff sense.

Proposition: *A pyramid $\mathcal{P}$ is concentrated iff it is of the form $\mathcal{P} = \mathcal{P}_Y$ for some $Y \in \overline{\mathcal{X}}^{H\,\mathcal{L}\iota}$. This Y is unique, and we say that "$\mathcal{P}$ concentrates to Y."*

Proof. Given two $\mathcal{L}$ and $\mathcal{L}'$ as above, i.e., of the form $\mathcal{L}ip_1(X)$ and $\mathcal{L}ip_1(X')$ for $X, X' \in \mathcal{P}$, we can make yet another space $\mathcal{L}''$ corresponding to X'' Lipschitz dominating X and X' and thus isometrically containing both $\mathcal{L}$ and $\mathcal{L}'$. It follows that whenever a sequence $X_i \subset \mathcal{P}$ represents $\mathcal{P}$, in the sense that the $\mathcal{P}_{X_i}$ weakly converge to $\mathcal{P}$, then $\mathcal{L}_i = \mathcal{L}ip_1(X_i)$ Hausdorff-converge to some space $\mathcal{L}$, and by the logic of the proof of $(ii) \Rightarrow (i)$ (see $3\frac{1}{2}$.50), the spaces X_i asymptotically concentrate to some Y such that $\mathcal{P}_Y = \mathcal{P}$. (We leave the details to the reader.)

$3\frac{1}{2}$.54. About the map $\overline{\mathcal{X}} \to \mathcal{H}(= \mathcal{X}_c)$. Let $\mathcal{H}$ denote the Hausdorff moduli space of compact metric spaces (which was called $\mathcal{X}_c$ in $3.11\frac{1}{2}{}_+$). Then the space $\overline{\mathcal{X}} = \overline{\mathcal{X}}^{H\,\mathcal{L}\iota}$ naturally goes to $\mathcal{H}$ for $X \mapsto (\mathcal{L}ip_1(X), \mathrm{me}_1)$. This map is first defined on $\mathcal{X} \subset \overline{\mathcal{X}}$, and then it extends to all of $\overline{\mathcal{X}}$ by continuity with respect to the $\underline{H}_1\,\mathcal{L}\iota_1$ metric, since $\mathcal{X}$ is dense in $\overline{\mathcal{X}}$.

> *The (obviously continuous) map $\overline{\mathcal{X}} \to \mathcal{H}$ is proper, i.e., preimages of compact sets are compact.*

In fact, let $X_i \in \mathcal{X}$ satisfy $\mathcal{L}ip_1(X_i) \xrightarrow{\mathrm{Hau}} \mathcal{L}$. Then we may assume without loss of generality that the sequence $\mathcal{X}_i$ is asymptotic (since we can always pass to a subsequence), and so the $\mathcal{P}_{X_i}$ weakly converge to some pyramid $\mathcal{P}$. This clearly is concentrated, hence the X_i concentrate to some $Y \in \overline{\mathcal{X}}$, and our claim follows. (Again the details are left to the reader.)

Question: Is our map $\overline{\mathcal{X}} \to \mathcal{H}$ one-to-one?

$3\frac{1}{2}$.55. About the embedding $\overline{X} \to \Pi$. The space Π of all pyramids $\mathcal{P} \subset \mathcal{X}$ is clearly compact, and the map $\overline{\mathcal{X}} \to \Pi$ defined by $Y \mapsto \mathcal{P}_Y$ is continuous for the $\underline{H}_1 \mathcal{L}\iota_1$ metric in $\overline{\mathcal{X}}$ and the topology of weak convergence in Π. This is immediate with all we have seen at this stage. Furthermore, this map is injective, as we have mentioned earlier. An additional remark we now wish to make is that this injective map is a *topological embedding*, i.e.,

the weak convergence $\mathcal{P}_{Y_i} \to \mathcal{P}_Y$ implies $\underline{H}_1 \mathcal{L}\iota_1$-convergence $Y_i \to Y$.

This is a mild generalization of the proposition above whose proof is left to the reader.

Now, all of the concentration discussion is expressed in the "pyramidal" language with the exception of the Hausdorff compactness criterion distinguishing $\overline{\mathcal{X}} \subset \Pi$, where we need the map $\mathcal{X} \to \mathcal{H}$ given by $X \mapsto (\mathcal{L}ip_1(X), \mathrm{me}_1)$.

$3\frac{1}{2}$.56. Basic example of asymptotic concentration. Consider partial cartesian products $X_i = F_1 \times F_2 \times \cdots \times F_i$ for some infinite sequence of mm spaces F_i of finite mass. If the observable diameters of the "tails" $\Phi_{ij} = F_i \times F_{i+1} \times \cdots \times F_j$ satisfy

$$\mathrm{ObsDiam}(\Phi_{ij}, -\kappa) \to 0$$

for each $\kappa > 0$ and $i, j \to \infty$, then the X_i clearly asymptotically concentrate, and the limit $Y \in \overline{\mathcal{X}}$ is represented by the infinite product $F_1 \times F_2 \times \cdots \times F_i \times \cdots$, which is a kind of "ideal mm space." An especially attractive instance of this is where the F_i are Riemannian manifolds with $\lambda_1(F_i) \to \infty$, such as the unit Euclidean spheres S^i, where we can bound $\mathrm{ObsDiam}_{\mathbb{R}}(\Phi_{ij})$ in terms of $\lambda_1(\Phi_{ij}) = \sup_{j \geq i} \lambda_1(F_j)$ (compare $3\frac{1}{2}$.62).

$3\frac{1}{2}$.57. Spectral criterion for asymptotic concentration. Let us invoke the space $\mathrm{Dir}_1 = \mathrm{Dir}_1(X)$ of Lipschitz functions f on X satisfying $\int_X \|\mathrm{grad}\, f\|^2 \, d\mu \leq 1$ and look at the behavior of the metric spaces $(\mathrm{Dir}_1(X), \|\cdot\|_{L_2})$ as X runs over some pyramid $\mathcal{P} \subset \mathcal{X}$.

Spectral concentration. Call $\mathcal{P}$ *spectrally concentrated* if for each $X \in \mathcal{P}$, the intersection of $\mathrm{Dir}_1(X)$ with the L_2-ball

$$L_2 B_1(X) = \left\{ f : \int_X f^2(x)\, d\mu_x \leq 1 \right\},$$

is compact with respect to the L_2-metric, and the totality of these intersections

$$\{\mathrm{Dir}_1(X) \cap L_2 B_1(X), \|\cdot\|_{L_2}\}$$

for $X \in \mathcal{P}$ is Hausdorff precompact.

Observe that the L_2-compactness of $\mathrm{Dir}_1 \cap L_2 B_1$ is equivalent to the discreteness of the spectrum of the Laplace operator on X, which can be understood in the usual way if X is a Riemannian manifold and the Dirichlet functional

$$f \mapsto \mathrm{Dir}(f) = \int_X \|\operatorname{grad}(f)\|^2 \, d\mu$$

is quadratic. (This can be extended to the non-Riemannian case with the formalism described in $3\frac{1}{2}.39$). Similarly, Hausdorff precompactness of a family $\{\mathrm{Dir}_1(X) \cap L_2 B_1(X), \|\cdot\|_{L_2}\}$ means, in the Riemannian case, a uniform bound on the number of eigenvalues in every fixed interval $[0, \lambda]$.

Also notice a certain discrepancy between the asymptotic concentration defined via $(\mathcal{L}ip_1(X), \mathrm{me}_1)$ and the spectral one using $(\mathrm{Dir}_1(X), \|\cdot\|_{L_2}$. Namely, in the former case, we did *not* intersect $\mathcal{L}ip_1$ with the unit L_∞-ball of functions. Consequently, if, for example, X_i consists of two atoms of fixed mass with mutual distance i, then the sequence X_i, $i \to \infty$, does *not* asymptotically concentrate with our convention, but it does concentrate spectrally. More generally, the asymptotic concentration excludes the possibility $\mathrm{ObsDiam}_{\mathbb{R}}(X; -\kappa) \to \infty$ for some $\kappa > 0$, while spectral concentration allows *finitely* many pieces of X of definite mass drifting far away from each other. Yet, we have the following trivial

Proposition: *Let $\{X\}$ be a family of mm spaces with the following three properties:*

(a) The masses of all our X's are bounded by a fixed constant $< \infty$.

(b) $\operatorname{diam}(X; m - \kappa) \leq \mathrm{const}_\kappa < \infty$ *for all* $\kappa > 0$. *(This means that there exist subsets $X_\kappa \subset X$ with* $\operatorname{diam} X_\kappa \leq \mathrm{const}_\kappa$ *and* $\mu(X \setminus X_\kappa) \leq \kappa$.*)*

(c) The family $\{\mathrm{Dir}_1(X) \cap L_2 B_1(X), \|\cdot\|_{L_2})$ *is Hausdorff precompact.*

Then the family $\{\mathcal{L}ip_1(X), \mathrm{me}_1\}$ *is also Hausdorff precompact.*

Proof. To clarify the ideas, assume that $\operatorname{diam} X \leq D$ for all our X and let $\overline{m} = \sup_{\{X\}} \mu(X)$. Then every Lipschitz function f on X vanishing at some point $x_0 \in X$ clearly has $\|f\|_{L_2} \leq \overline{m} D^2$, and so it is contained in the slightly enlarged set $\mathrm{Dir}_1 \cap L_2 B_1$, namely in $\overline{m}\,\mathrm{Dir}_1(X) \cap \overline{n} L_2 B_1(X)$ for $\overline{n} = \overline{m} D^2$. On the other hand, the L_2-metrics dominate me_1 by the Chebyshev inequality, and our proposition follows, since we can make every f vanish at x_0 by adding a constant which does not affect relevant Hausdorff distances. Finally, we observe that the very nature of the metric me_1 allows us to throw away subsets of small measure κ, and so we can use (b) instead of the stronger assumption $\operatorname{diam}(X) \leq D$. (The details are left to the reader.)

By specializing the proposition to pyramids and to (monotone) sequences of spaces, we immediately obtain the following

Corollaries: *(i) If a pyramid $\mathcal{P}$ is spectrally concentrated, then it is concentrated provided that (a) and (b) hold for all $X \in \mathcal{P}$.*

(ii) Let X_i be a sequence of Riemannian manifolds satisfying (a) and (b) and such that the number of eigenvalues of the Laplace operator on X_i in each interval $[0, \lambda]$ is bounded by a constant $c_\lambda < \infty$ independent of i. Then X_i admits an infinite subsequence which asymptotically concentrates. Furthermore, if the X_i are Lipschitz monotone increasing, $X_1 \prec X_2 \prec \cdots$, then they asymptotically concentrate themselves.

Remark: Spectral concentration is a by far more restrictive condition than asymptotic concentration. Thus, the "ideal apex" of a spectrally concentrated pyramid may carry far more interesting (spectral) geometry than a general "limit space" $Y \in \overline{\mathcal{X}}$. The first example to look at is the infinite product $F_1 \times F_2 \times \cdots$ considered in the previous section.

Exercises: (A) Sharpen (a) or (b) in the proposition by making them *necessary* for the asymptotic concentration for as well as sufficient in conjunction with (c).

(B) Let X_i be a sequence of Riemannian manifolds of unit mass satisfying $\lambda_k(X_i) \to \infty$ as $i \to \infty$ for some k. In other words, for each $\lambda > 0$, the spectrum of Δ_{X_i} contains at most k points in $[0, \lambda]$ for $i \geq i(\lambda)$ (where the points of the spectrum are counted with multiplicity). Show that the separation distance satisfies $\operatorname{Sep}(X_i; \kappa_0, \ldots, \kappa_k) \to 0$ as $i \to \infty$ and arbitrary $\kappa_0, \ldots, \kappa_k \geq 0$. Then (following Alon and Milman), look at 1-Lipschitz functions $f_i : X_i \to \mathbb{R}$ and observe that the pushforward measures μ_i on $\mathbb{R}$ concentrate to at most k points. Namely, there exist subsets $I_i \subset \mathbb{R}$, each consisting of the union of k intervals of length ε_i, such that $\varepsilon_i \to 0$ as $i \to \infty$ and $\mu_i(I) \to 1$.

(C) **Question:** Let X_i be Riemannian manifolds whose spectra converge to that of a fixed space X. Under what assumptions do the spaces X_i concentrate to X? (Problem (B) above shows that it works for finite spaces X.) In any case, what can be said about asymptotic (i.e., $\underline{H}_1 \mathcal{L}\iota_1$) limit $Y \in \overline{\mathcal{X}}$ of these X_i? (One may strengthen the assumptions by bringing in the nonlinear spectra mentioned in $3\frac{1}{2}.39$).

(D) Take a sequence of probability mm spaces $F_1, F_2, \dots$ and let $X_n = F_1 \times F_2 \times \dots \times F_n$, where the metric on X_n is ℓ_p, i.e.,

$$\left(\sum_{i=1}^{n} \mathrm{dist}_i^p\right)^{1/p},$$

for some $1 \leq p \leq \infty$. Give a criteria for the following behavior:

(*i*) $\underline{\square}_1$-convergence of X_n (compare 3.38(6)(f) and (f')).

(*ii*) Asymptotic concentration.

(*iii*) Spectral concentration.

Consider specifically the case of $F_i = (F, \lambda_i \,\mathrm{dist}, \mu)$ as in 3.38(6)(f), and show that the X_n spectrally (and thus asymptotically) concentrate for $p = 2$ if F is a compact Riemannian manifold with a normalized Riemannian measure μ whenever $\lambda_i \to 0$. Show that this is also true for the Gaussian space $F = (\mathbb{R}, |x - y|, e^{-\pi x^2}\, dx)$.

(E) Consider (full) ellipsoids X_n with principal semiaxes $a_i = a_i(n)$, $i = 1, \dots, n$, i.e.

$$X_n = \{x \in \mathbb{R}^n : \sum_{i=1}^{n} a_i^{-2} x_i^2 \leq 1\}$$

with normalized Euclidean measures and figure out the mode of concentration of X_n in terms of $a_i(n)$ for $n \to \infty$. In particular, prove the Levy concentration of X_n (i.e., concentration to the one point space) for $a_i = \varepsilon_i \sqrt{i}$, $i = 1, \dots, n$ and $n \to \infty$, where $\varepsilon_i \to 0$ for $i \to \infty$. Show that there is no asymptotic concentration for $a_i = \sqrt{i}$, $i = 1, \dots, \sqrt{n}$, and $n \to \infty$, not even for any subsequence of X_n. But if, for example, $a_i(n) = \varepsilon_i \sqrt{n}$, where $\varepsilon_i \to 0$ as $i \to \infty$, then the X_n asymptotically (and even spectrally) concentrate. Moreover, if $\sum_{i=1}^{\infty} \varepsilon_i^2 < \infty$, then the X_n concentrate to an actual mm space, namely to the Hilbert space with a Gauss measure on it. (Yet these X_n may easily $\underline{\square}_1$-diverge.)

(F) Observe similar behavior of more general (convex) subsets in $\mathbb{R}^n$ as well as in other Banach spaces X^n of dimension $n \to \infty$. In particular, look at

$$X_n = \{x \in \mathbb{R}^n : \sum_{i=1}^{n} a_i x_i^a \leq 1\}$$

where $\mathbb{R}^n$ can be given some ℓ_p-metric. Also study the boundaries of such convex sets with induced (Minkowski) measures and metrics.

J. Concentration versus dissipation

A dissipative behavior (of a sequence) of mm spaces is the opposite to that of concentration. Here are specific

$3\frac{1}{2}$.58. Definitions. A sequence X_i *δ-dissipates mass* m if for large $i \to \infty$ there are (arbitrarily) small pieces of X drifting δ-far apart and adding up to mass m. That is,

$$\liminf_{i\to\infty} \mathrm{Sep}(X_i; \kappa_0, \dots, \kappa_N) \geq \delta$$

for every fixed $(N+1)$-tuple of positive numbers κ_i satisfying $\kappa_0+\kappa_1+\cdots+\kappa_N \geq m$ (see $3\frac{1}{2}$.30 for the definition of Sep). If all these X_i themselves have $\mu(X_i) = m$, we say that they are *δ-dissipating.* More generally, a sequence X_i is called δ dissipating if, for each $m > 0$, every subsequence of spaces $Y_j \subset X_{i_j}$ of masses $m_j \geq m$ δ-dissipates mass m.

For example, if X_i is an asymptotic sequence with $\mu(X_i) \overset{i\to\infty}{\longrightarrow} m < \infty$, then it δ dissipates if and only if the (weak) limit pyramid $\mathcal{P} = \lim_{i\to\infty} \mathcal{P}_{X_i}$ contains (the isomorphism class of) every mm space of mass $\leq m$ and diameter $\leq \delta$. (The proof of this is straightforward.)

Finally, $\{X_i\}$ is called *infinitely dissipating* if it δ dissipates for all $\delta > 0$. This is equivalent for the X_i above to having

$$\mathcal{P} = \mathcal{X}.$$

Notice that the full negation of dissipation is weaker than concentration. Namely, a sequence X_i is called *nondissipating* if for every $m, \delta > 0$, there exists $\kappa > 0$ such that the inequalities $\mathrm{Sep}(X_i; \kappa_0, \dots, \kappa_N) \geq \delta$ and $\kappa_0 + \kappa_1 + \cdots + \kappa_N \geq m$ imply that $\sup_{0\leq k\leq N} \kappa_k \geq \kappa$, independently of i. In other words, no subsequence δ-dissipates a mass $m > 0$ unless $\delta = 0$.

Example: The unit Euclidean cubes $[0,1]^n \subset \mathbb{R}^n$, $n = 1, 2, \dots,$ comprise a nondissipating family, but they do not asymptotically concentrate.

Proof. The nonconcentration is obvious. In fact, even if we use the smaller ℓ_∞-metric on I^n, the cubes still do not concentrate, since the n (1-Lipschitz) coordinate projections $I^n \to \mathbb{R}$ make a *non precompact* family for the Hausdorff metric associated to me_1.

But on the other hand, the nondissipation follows from the bound $\lambda_1([0,1]^n) = \lambda_1([0,1]) = \pi^2 > 0$ via the following lemmas.

(A) *If spaces X_i of masses $\leq m \leq \infty$ have large expansion of subsets of small measure, namely*

$$\exp(X_i; \kappa, \rho) \geq e > 1 \tag{$*$}$$

for every i, each given $\rho > 0$, and $0 < \kappa \leq \kappa(\rho)$, then the sequence $\{X_i\}$ does not dissipate.

Indeed $(*)$ shows that the δ-neighborhood of a subset $Y \subset X_i$ of small mass κ has much greater mass than Y and our claim follows.

(B) *If $\lambda_1(X) \geq \varepsilon > 0$, then the X_i satisfy the expansion property above.*

This is a special case of $3\frac{1}{2}$.38.

$3\frac{1}{2}$.59. The subquotient order $X \lhd Y$. This is defined by

$$X \lhd Y \qquad \Leftrightarrow \qquad X \prec Y' \subset Y,$$

i.e., X is Lipschitz dominated by some *closed* subset $Y' \subset Y$ with the metric and measure of Y restricted to Y'. (Little would change if we allowed arbitrary Borel subsets $Y' \subset Y$.) It is indeed a partial ordering, since (obviously)

$$X \lhd Y \lhd Z \qquad \Leftrightarrow \qquad X \lhd Z,$$

and (by an easy argument),

$$X \lhd Y,\ X \rhd Y \qquad \Leftrightarrow \qquad X \simeq Y$$

in $\mathcal{X}$. A useful invariant of an mm space X is the totality of all spaces $Y \lhd X$ consisting of at most k points, $k = 1, 2, \ldots, \infty$, denoted $F_k^{\lhd}X$. These carry the same information as (the totality of) the measures $\underline{\mu}_r^X$ (see 3.39) but are more appropriate for expressing the ideas of concentration and dissipation.

Exercises: (*a*) Reformulate the notions of dissipation above in terms of $F_k^{\lhd}$.

(*b*) Determine which of the invariants from $3\frac{1}{2}$.F are monotone for the subquotient order.

(*c*) Figure out when two mm spaces X and Y are *comparable*, i.e., satisfy

$$X \lhd \lambda Y \qquad \text{and} \qquad Y \lhd \lambda X \qquad\qquad (\lhd)_\lambda$$

for some $\lambda \geq 1$. Study the corresponding metric $\mathrm{dist}^{\lhd}(X, Y)$ defined as $|\log \lambda|$ for the infimal λ in $(\lhd)_\lambda$.

(*d*) What can you say about X and Y if some ε-approximations to X and Y satisfy

$$X_\varepsilon \lhd Y_\varepsilon$$

for $\underline{\Box}_1(X, X_\varepsilon) \leq \varepsilon$ and $\underline{\Box}_1(Y, Y_\varepsilon) \leq \varepsilon$?

(*e*) Define $X \lhd_\sigma Y$ if there is a subset $X' \subset X$ with $\mu(X \setminus X') \leq \sigma\mu(X)$ such that $X' \lhd Y$. Show that the Euclidean ball in $\mathbb{R}^n$ of radius $\sqrt{n}$ and the unit cube satisfy

$$[0,1]^n \lhd B^n(\sqrt{n}) \lhd_\sigma C(\sigma)\sqrt{\log n}[0,1]^n$$

for all $\sigma < 1$, where the ball and the two cubes (the unit and the scaled ones) are given their respective (nonnormalized) Lebesgue measures. (*Hint:* Use the lemma in 5.7 of [Mil–Sch].)

Question. Can one improve upon this $\sqrt{\log n}$?

(*f*) Let $A \subset \mathbb{R}^n$ be an arbitrary (centrally symmetric?) convex body. Does there exist an affine image A' of A such that

$$\lambda_\sigma^{-1} A' \lhd_\sigma B^n \lhd_\sigma \lambda_\sigma A',$$

where B^n is the Euclidean ball and λ_σ is of the order $\log n$?

Remark: A more traditional (also open) problem (communicated to me by V. Milman) is to find A' with

$$\mu((\lambda_\sigma^{-1} A') \cap B^n) \geq \lambda_\sigma^{-1}\sigma\mu(A')$$

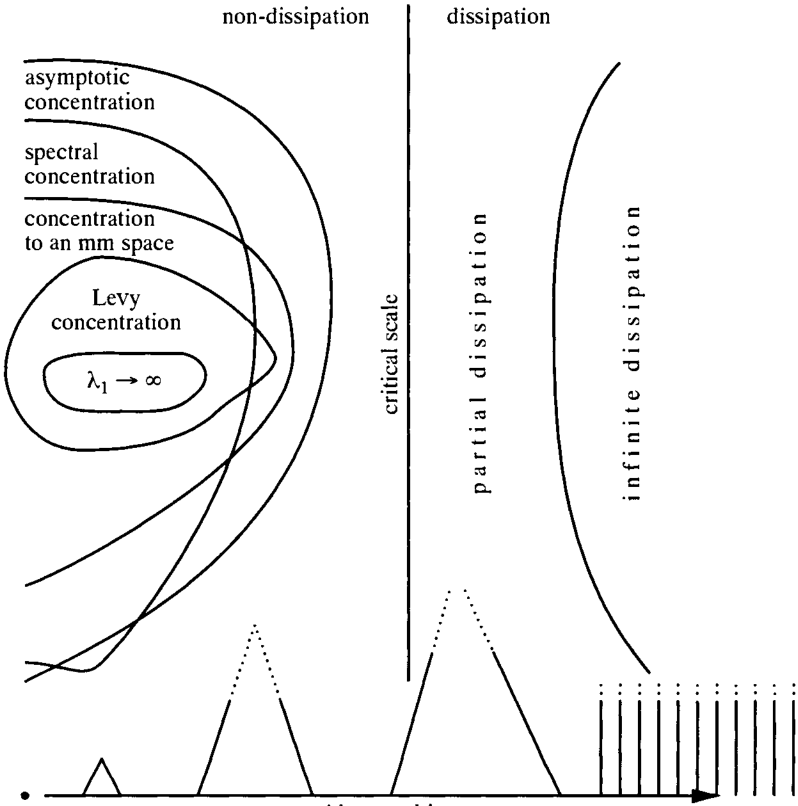
and

$$\mu(B^n \cap \lambda_0 A') \geq \sigma\mu(B^n)$$

with small λ_σ (say of order $\log n$).

$3\frac{1}{2}$.60. A schematic picture of concentration and dissipation. One can think of concentrated (seqences of) mm spaces as those having small (weak limit) pyramids (where one can also use the pyramdids for the subquotient order), while dissipation corresponds to large pyramids. An overall picture is on the next page

$3\frac{1}{2}$.61. Concentration and dissipation for non-Lipschitz functions. One can generalize the above setting by considering, instead of a metric, another structure on a measure space X. What is essential is a distinguished class of observables, i.e., $\mathbb{R}$-valued (or more generally $\mathbb{R}^n$-valued) functions on X replacing our 1-Lipschitz functions. Notice that one can often think of such a class as a subsets of the set of all 1-Lipschitz functions on X for a suitable metric on X, e.g., the *maximal metric* for which all functions in our class are 1-Lipschitz. But it may happen that the distinguished functions are more concentrated than more general 1-Lipschitz functions.



Examples: (*a*) The classical case is that of $\mathbb{R}^n$ with the product measure $\mu_n = \mu \times \mu \times \cdot \times \mu$ for some probability measure μ on $\mathbb{R}$, where one is interested in the rate of concentration of the sum

$$\sigma_n \colon x = (x_1, \dots, x_n) \mapsto x_1 + x_2 + \cdots + x_n \in \mathbb{R}$$

for $n \to \infty$. The law of large numbers asserts the concentration of $\sigma_n(x)/n$ to its mean value, provided that $\int_{\mathbb{R}} x\, d\mu_x < \infty$, and the central limit theorem implies the concentration for $\lambda_n \sigma_n(x)$ whenever $\lambda_n/\sqrt{n} \to 0$, and, say, μ has compact support. Yet, we know that more general Lipschitz functions do not always concentrate in this way, e.g., on $\{0, 1\}^n$.

(*b*) The above generalizes to a measure space X with a given measurable automorphism (or endomorphism) $A \colon X \to X$ and a function $f \colon X \to \mathbb{R}$. Here, one considers the pushforward measures μ_n on $\mathbb{R}^n$ given by n-tuples of functions

$$(f, f_1, \dots, f_{n-1}) \colon X^n \to \mathbb{R},$$

where $f_0 = f$ and $f_i = f_{i-1} \circ A$. The ergodic theorem asserts the concentration of $\sigma_n(x)/n$ as $n \to \infty$, provided that A is ergodic and f is L_1, while more general Lipschitz functions on $(\mathbb{R}^n, \mu_n)$ may be rather dissipating. (This happens, for example, if A is an irrational rotation on the circle, and f is a smooth function, e.g., an eigenfunction for the rotation, i.e., $f(z) = z^n$.)

Problems: Study the (metric) geometry of $(\mathbb{R}^n, \mu_n)$ for μ_n above for "interesting" pairs (A, f).

(c) Let μ_n be some measure on $\mathbb{R}^n$, and let the admissible observables be normal projections to linear subspaces in $\mathbb{R}^n$. A specific example is provided by taking μ_n equal to the normalized Lebesgue measure on some domain $\Omega = \Omega_n \subset \mathbb{R}^n$. Here, the classical Brunn–Minkowski theorem yields a nontrivial concentration for *convex* domains Ω. Namely, if the principal curvatures $0 \leq \kappa_1 \leq \kappa_2 \leq \cdots \leq \kappa_{n-1}$ at every point in the boundary $\partial\Omega$ of Ω satisfy $\kappa_1 \geq \varepsilon_n$, where $\varepsilon_n\sqrt{n} \to \infty$ as $n \to \infty$, then (the pushforwards of μ_n under) the projections $\mathbb{R}^n \to \mathbb{R}^m$ concentrate to point for every fixed p and $n \to \infty$. In fact, this remains true for arbitrary 1-Lipschitz maps $\mathbb{R}^n \to \mathbb{R}^m$ as follows from Levy's theorem. But, the Brunn–Minkowski result also applies to nonstrictly convex domains Ω and partially log-convex measures in these, which gives us a concentration of projections $\Omega \to \mathbb{R}^m$ to k-dimensional subsets in $\mathbb{R}^m$ with k independent of m.

Questions: Let $f_n : \Omega_n \to \mathbb{R}^m$ be arbitrary 1-Lipschitz functions. Do the pushforward measures $\underline{\mu}_n$ concentrate near k dimensional subsets in $\mathbb{R}^m$ if we assume that $\kappa_{k+1} \geq \kappa > 0$? Here we expect that every asymptotic subsequence Ω_{n_i} concentrates to an mm space of dimension $\leq q$. More generally, we ask if this also happens to Riemannian manifolds X with Ricci ≥ 0, where Ricci $\geq \kappa > 0$ normally to subspaces of dimension $\leq q$ in all tangent spaces $T_x(X)$ (compare $3\frac{1}{2}$.27(k)).

$3\frac{1}{2}$.62. Concentration and dissipation of products and towers of fibrations. We are concerned here with products $F_1 \times \cdots \times F_i \times \cdots \times F_n$ for large $n \to \infty$, and so it matters how we define the product metric. In Riemannian geometry, one uses the pythagorean (or Euclidean) ℓ_2 *product*, $\mathrm{dist}_{\ell_2} = \sqrt{\sum_i \mathrm{dist}_i^2}$. More generally, one can use the ℓ_p-product, $\mathrm{dist}_{\ell_p} = (\sum_i \mathrm{dist}_i^p)^{1/p}$ for all $p \in [0, \infty]$, where the ℓ_∞-metric is just the sup metric $\mathrm{dist}_{\ell_\infty} = \sup_i \mathrm{dist}_i$. This metric is the *minimal possible* on the product, such that the projections to the fibers $F_1 \times \cdots \times F_i \times \cdots \times F_n \to F_i$ are 1-Lipschitz. The maximal natural metric is dist_{ℓ_1}. It is maximal among the metrics for which all fiber embeddings $F_i \to f_1 \times \cdots \times F_i \times \cdots f_n \subset F_1 \times \cdots \times F_i \times \cdots \times F_n$ are 1-Lipschitz.

As far as the measure is concerned, we assume here all μ_i on F_i to be normalized, i.e., $\mu_i(F_i) = 1$, and then our products are also normalized.

The problem which begs for resolution is that of the evaluation of basic (concentration) invariants of $X^n = F_1 \times \cdots \times F_n$ in terms of suitable invariants of F_i. For example, we want to know when the X^n (and more generally the $\lambda_n X^n$) concentrate or dissipate. We already know that the observable diameter of X^n is not determined by that of the factors F_i. For example, the observable diameter of the ℓ_2-cartesian power $[0,1]^n$ is bounded as $n \to \infty$, while the vertex set of the cube $[0,1]^n$, i.e., the cartesian power of $\{0,1\}$, has ObsDiam $\sim n^{1/4}$ (see $3\frac{1}{2}.42$). Furthermore, it is clear that the powers $\{0,1\}^n$ dissipate even for the (smallest) ℓ_∞-metric, while the $[0,1]^n$ do not dissipate. In fact, if F is an arbitrary *disconnected* mm space, then the $F^n = F \times F \times \cdots \times F$ ε-dissipate for some $\varepsilon > 0$. This is obvious. What is more interesting is the following

(1) Non-dissipation theorem: *Let F be a compact, connected, and locally connected mm space (where we assume as usual that* $\operatorname{Supp}\mu = F$*). Then the ℓ_∞-powers $F^n = (F \times \cdots \times F, \operatorname{dist}_{\ell_\infty})$ do not dissipate as $n \to \infty$.*

Proof. We start with the following obvious

Majorization lemma: *Let $F_0 \to F$ be a uniformly continuous map between mm spaces pushing forward the measure μ_0 of F_0 to μ on F. If the ℓ_∞-powers F_0^n do not dissipate as $n \to \infty$, then the F^n do not dissipate either.*

Next, we observe that our (compact, connected, locally connected) F can be majorized by the segment $[0,1]$, i.e., it admits (as every child knows) a continuous (and hence uniformly continuous) map $[0,1] \to F$ pushing forward the Lebesgue measure to our μ on F.

Finally, we recall (and this is the only nontrivial point) that the cubes $[0,1]^n$ do not dissipate as $n \to \infty$, even when we use the ℓ_2-metric, which is much greater than ℓ_∞ (see $3\frac{1}{2}.42$), and the theorem follows.

Remarks: (a) The proof above looks like a cheat. Well, sometimes crime pays... (and an honest approach, harder but having the advantage of illuminating noncompact and nonlocally connected spaces, will not be pursued in this book).

(b) The majorization lemma indicates the usefulness of the order between mm spaces coming from uniformly continuous, measure preserving maps $f: X \to Y$. Since we are interested in sequences of spaces, we must fix some modulus of continuity $\delta(\varepsilon)$ and then stick to $\delta(\varepsilon)$-continuous maps. Of course, if $X \to Y$ and $Y \to Z$ are $\delta(\varepsilon)$-continuous, then their composition $X \to Z$ need not be $\delta(\varepsilon)$-continuous, and so we do not get a true order relation for general $\delta(\varepsilon)$ (it is necessary to have $\delta(\varepsilon) \le \varepsilon$). But, it still makes sense to speak of certain (quasi-)monotonicity of geometric invariants under the relation $X \xrightarrow{\delta(\varepsilon)} Y$. We suggest that the reader do that for the invariants we have met so far.

(2) Concentration and random walks. The isoperimetric profile of an mm space X is pretty well reflected in the behavior of the *random walk* on X. We study here only the case where X is a Riemannian manifold, so that the random walk amounts to the *heat flow*, i.e., the family of operators $H(t) = exp(t\Delta)$, where Δ is the positive Laplacian on X, i.e., $-\sum_i(\partial/\partial x_i^2)$. This $H(t)$ carries complete information about the spectrum $\{\lambda_0, \lambda_1, \dots\}$ of Δ for

$$\mathrm{Tr} H(t) = \sum_{i=0}^{\infty} \exp^{-\lambda_i t}.$$

Since $\lambda_0 = 0$, this trace ≥ 1 and if for some $t_0 > 0$ we have $\mathrm{Tr}H(t_0) \le 1 + c_{t_0}$ with $c_{t_0} < 1$, we obtain a nontrivial lower bound on λ_1, namely

$$\lambda_1 \ge -(\log c_{t_0})/t_0. \tag{$*$}$$

In fact,

$$\lambda_1 = \lim_{t\to\infty} -(\log c_t)/t. \tag{$*\infty$}$$

On the other hand, this trace can be evaluated by the *(heat) kernel* of the operator $H(t)$, denoted by $H(x, x'; t)$, as

$$\mathrm{Tr} H(t) = \int_X H(x, x; t)\, dx.$$

This allows a bound on $\mathrm{Tr}H(t)$ and hence on λ_1 in terms of the L_∞-bound on the heat kernel, i.e.,

$$\|H(t)\|_{L_\infty} \overset{\text{def}}{=} \sup_{x,x'} H(x, x'; t).$$

Notice at this point that $H(x, x'; t)$ is a positive symmetric function on $X \times X$ for each $t > 0$ and $\int_X H(x, x; t)\, dx = 1$ for all y and t. Also observe that the heat kernel of the scaled manifold $sX = (X, s\,\mathrm{dist})$ equals

$$H_s(x, x'; t) = H(x, x'; s^{-2}t).$$

What is most important for us is the multiplicative property of the heat kernel for cartesian products $Z = X \times Y$

$$H_Z((x,y),(x',y');t) = H_X(x,x';t)H_Y(y,y';t).$$

This (trivially) implies, for example, the basic inequality

$$\lambda_1(X \times Y) = \min(\lambda_1(X), \lambda_1(Y))$$

used in $3\frac{1}{2}.42$. And, we are about to prove a similar relation for *Riemannian fibrations* $p\colon Z \to X$. This means that Z is a Riemannian manifold with totally geodesic, mutually isometric fibers $Y = Y_x = p^{-1}(x)$, and (the differential of) p is isometric on the vectors normal to the fibers. Every such Z can be built out of a given X and Y by choosing a principal bundle over X with the group $G = \operatorname{Isom} Y$ and taking a connection in this bundle. Then one takes the associated bundle $Z \to X$ with the Y-fibers and constructs in the obvious way a Riemannian metric on Z which turns Z into a Riemannian fibration such that the *horizontal* subbundle $T^{hor}(Z) \subset T(Z)$ consisting of vectors normal to the fibers defines our connection.

The heat flow on such a Z is expressed by the famous *Kac- Feynmann formula*, which states that the heat flow on Z is obtained, roughly speaking, from that on $X \times Y$ by an additional averaging along the fibers corresponding to the holonomy of the connection. This immediately leads to the following

Kato inequality.

$$\|H_Z(t)\|_{L_\infty} \leq \|H_X(t)\|_{L_\infty} \|H_Y(t)\|_{L_\infty} \tag{$\times$}$$

(as was explained to me by Jürg Frölich around 1980).

Corollary:

$$\lambda_1(Z) \geq \min(\lambda_1(X), \lambda_1(Y)) \tag{$**$}$$

Proof. To show this, we need the following modifications of $(*\infty)$ expressing the first eigenvalue λ_1 of a Riemannian manifold X in terms of the heat decay $d_t = \operatorname{vol}^{-1}(X) - \|H_X(t)\|_{L_\infty}$,

$$\lambda_1 = \lim(-\log d_t)/t. \tag{$+\infty$}$$

Clearly, $(\times)$ and $(+\infty)$ imply the desired $(**)$, and so it only remains to prove $(+\infty)$. To do this, we expand the heat kernel into the Fourier series

$$H(x,x';t) = \sum_{i=0}^{\infty} e^{-\lambda_i t}\varphi_i(x)\varphi_i(x') \tag{$\star$}$$

for the orthonormal eigenfunctions φ_i of Δ on X, where we recall that $\lambda_0 = 0$, and φ_0 is constant $= (\mathrm{vol}(X))^{1/2}$ (which makes $\int_X \varphi_0^2\, dx = 1$). The numbers λ_i grow, approximately, like $i^{2/\dim(X)}$, and the φ_i are smooth functions satisfying

$$\|\varphi_i\|_{L_\infty} = O(\lambda_i^{\dim(X)})$$

for large $i \to \infty$, as standard estimates show. Therefore, $(\star)$ implies $(+\infty)$.

Application: *Let* $\cdots \to X^n \to X^{n-1} \to \cdots \to X^0$ *be a tower of Riemannian fibrations with fibers* F_n, $n = 1, 2, \ldots$. *If* $\lambda_1(F_n) \to \infty$, *then the sequence* X^n *asymptotically concentrates, and if* $\lambda_1(F_n) \geq \varepsilon > 0$, *then the* X^n *do not dissipate, provided that each* X^n *is assigned the normalized Riemannian measure.*

Example. The above can be applied to towers of spherical fibrations, i.e., to the iterated unit tangent bundles of a unit sphere S^p, where

$$X_0 = S^p, \ \ X_1 = T_1(S^p), \ \ X_2 = T_1(X^1) \ \ \cdots,$$

where each $T_1(X)$ comes along with the Levi–Civita connection of X.

Question. What is the actual asymptotic behavior of $\lambda_i(X^n)$ for these X^n as $n \to \infty$?

Notice that the Kac–Feynmann formula suggests a sharpening of $(**)$ of the form

$$\lambda_1(Z) \geq \min(\lambda_1(X), \lambda_1(Y) + \varepsilon),$$

where $\varepsilon \geq 0$ is estimated from below in terms of the holonomy of the fibration $Z \to X$. For example, let Z_j be circle fibrations over a fixed base X with curvature forms $i\omega$ for a given *generic* 2-form λ on X. Then an easy argument shows that $\varepsilon \to \infty$ as $j \to \infty$, and hence $\lambda_1(Z) \to \lambda_1(X)$, provided that $\dim X \geq 2$. In fact, the growth of the curvature forces $\|\varphi\|_{L_2}/\|\operatorname{grad}\varphi\|_{L_2} \to 0$ for all functions φ having zero integrals around the fibers, which are themselves circles in this case. For general Riemannian fibrations $Z \to X$, one defines the space Φ of functions on Z which are constant on the orbits of the isometry group $\operatorname{Isom} Y_x$ on each fiber Y_x, $x \in X$ (which are all mutually isometric), and the above applies to functions φ orthogonal to Φ under suitable assumptions of "generic growth" of the curvature of the fibration.

(2′) Everything we said automatically extends to X's with Riemannian metrics and to arbitrary (possibly non-Riemannian) measures (compare

[Gro]$_{\mathrm{SGSS}}$), where a basic example is $X = \mathbb{R}$ with some measure $\mu = \varphi(x)\,dx$. Such an $X = \mathbb{R}$ has $\lambda_1 > 0$, provided that

$$\min(\mu(-\infty, x], \mu(x, \infty]) \leq \mathrm{const}\,\varphi(x)$$

for all $x \in \mathbb{R}$ and some const ≥ 0, as follows from the (trivial special case of the) celebrated inequality by Jeff Cheeger (see [Gro]$_{\mathrm{SGSS}}$). For instance, this is satisfied by $\varphi(x) = e^{-|x|^p}$ for $p \geq 1$. Consequently,

> *the Euclidean space* $\mathbb{R}^n$ *with the measure* $\exp(-\sum_{i=1}^n |x_i|^p)\,dx$ *has* $\lambda_1 \geq \varepsilon > 0$ *independently of* n *(and* p*).*

This, as was pointed out to me by G. Schechtman, applies to an evaluation of the observable diameters of ℓ_p-balls in $\mathbb{R}^n$ (see (5) below).

(3) ℓ_1-products, martingales, and Ex_β. We shall indicate here an elementary approach to finding a bound on $\mathrm{ObsDiam}\,X^n$ for $X^n = X \times X \times \cdots \times X$ (n factors) equipped with the ℓ_1-product metric. In particular, we shall see that

$$\mathrm{ObsDiam}\,X^n = O(\sqrt{n}),$$

as $n \to \infty$, provided that $\mathrm{diam}\,X < \infty$, which is not bad since the ordinary diameter, as well as the characteristic size of this X^n, are $\approx n$. Also, we shall see that X^n with the ℓ_p-product metric has $\mathrm{ObsDiam} = O(n^{1/2p})$ (compare for $p = 2$), while the diameter and the characteristic size of these X^n are $\approx n^{1/p}$.

We use in our evaluation of $\mathrm{ObsDiam}\,X^n$ the classical idea of the Laplace–Gibbs transform of the measure $f_*\mu$ on $\mathbb{R}$ and set

$$\mathrm{Ex}_\beta \stackrel{\mathrm{def}}{=} \sup_f \int_X e^{\beta f(x)}\,d\mu_x,$$

where the f runs over all 1-Lipschitz functions with $\int_X f(x)\,d\mu_x = 0$. This Ex_β carries essentially the same information as the observable diameter or rather the central observable radius, but it is more suitable for the study of product spaces. Namely, *the ℓ_1-product of two mm spaces with normalized measures satisfy*

$$\mathrm{Ex}_\beta(X \times Y) \leq \mathrm{Ex}_\beta(X)\,\mathrm{Ex}_\beta(Y) \qquad (\times\times)$$

Proof. Take a function $f = f(x, y)$ on $X \times Y$ and let $\overline{f} = \overline{f}(x)$ be the function obtained by averaging f over the fibers $Y = x \times Y \subset X \times Y$, i.e.,

$$\overline{f} = \int_Y f(x, y)\,d\mu y,$$

where we note that

$$\int_X \overline{f}(x)\,d\mu_x = \int_X \int_Y f(x,y)\,d\mu_x\,d\mu_y.$$

Write $f(x,y) = \overline{f}(x) + g(x,y)$ for $g \stackrel{\text{def}}{=} f - \overline{f}$ and observe that g has zero integral over all fibers $x \times Y$. Thus,

$$\int_X \int_Y e^{\beta f} = \int_X d\mu_x \int_Y e^{\mu \overline{f}(x)} e^{\beta g(x,y)}\,d\mu_y \leq \int_X e^{\beta \overline{f}(x)}\,\mathrm{Ex}_\beta\,Y,$$

provided that f, and hence g, are 1-Lipschitz in the y-variable. Next, assuming that f is 1-Lipschitz in x, we observe that $\overline{f}$ is Lipschitz as well, and consequently,

$$\left(\int_X e^{\beta \overline{f}(x)}\right) \mathrm{Ex}_\beta(Y) \leq (\mathrm{Ex}_\beta\,X)(\mathrm{Ex}_\beta\,Y),$$

provided that $\int\int f(x,y) = \int \overline{f}(x) = 0$.

Corollary: *The ℓ_1-cartesian power X^n of X satisfies*

$$\mathrm{Ex}_\beta\,X^n \leq (\mathrm{Ex}_\beta\,X)^n \tag{$\times^n$}$$

for all $\beta \in \mathbb{R}$.

Now, in order to evaluate $\mathrm{ObsDiam}\,X^n$, we must choose a suitable β, and we try $\beta = \pm 1/\sqrt{n}$. If $\mathrm{diam}(X) = d$, then every 1-Lipschitz function f on X with zero mean is bounded by d, and

$$\int_X e^{\beta f} = \int_X 1 + \beta f + \frac{1}{2}\beta^2 f^2 + O(\beta^3 f^3)$$

for small β, and since we assume $\int_X f = 0$, we get

$$\int_X e^{\pm f/\sqrt{n}}) = 1 + \frac{d^2}{2n} + O(n^{-3/2}).$$

It follows that

$$\mathrm{Ex}_{\pm 1/\sqrt{n}}(X^n) \approx \left(1 + \frac{d^2}{2n}\right)^n = \delta \approx e^{d^2/2}.$$

This trivially implies that

$$\mathrm{ObsCRad}(X^n, -\kappa) = O(\sqrt{n}(\log\delta - \log\kappa)),$$

since

$$\int e^{\beta f} \geq \kappa_+ e^{\beta R}$$

for $\beta \geq 0$ and $\kappa_+ = \mu\{x \in X : f(x) \geq R\}$, as well as

$$\int e^{\beta f} \geq \kappa_- \operatorname{Ex} \beta R$$

for $\beta \leq 0$ and $\kappa_- = \mu\{x \in X : f(x) \leq -R\}$.

Remark: The above generalizes in an obvious way to towers of fibrations $\to X^n \to X^{n-1} \to \cdots \to X_0$, where the sequence of functions $\overline{f}_i$ on X for a given $f = \overline{f}_n$ on X_n is defined with the partition of X^{i+1} into the fibers of the map $X^{i+1} \to X^i$ as follows. We assume that these maps are measure-preserving for all $i = 0, 1, \ldots, n-1$, and we recall the canonical probability (i.e., normalized) measures in the fibers. Then the function $\overline{f}_{n-1}$ is defined by averaging $f = \overline{f}_n$ along the fibers of $X^n \to X^{n-1}$, where we doublethink of this $\overline{f} = \overline{f}_{n-1}$ as a function on X^{n-1} as well as on the original space X^n where it comes from X^{n-1} by composing with the projection $X^n \to X^{n-1}$. Then we do the same to $\overline{f}^{n-1}$ on X^{n-1} and get $\overline{f}^{n-2}$ on X^{n-2}, which is also thought of as the function on X^n constant on the fibers of the composed projection $X^n \to X^{n-2}$. Eventually, we obtain the sequence $\overline{f}^n, \overline{f}^{n-1}, \ldots, \overline{f}^0$, where $\overline{f}^i$ on X^n is constant on the fibers of $X^n \to X^i$, and so can be thought of as a function on X^i. This sequence is called a *martingale* associated to the partitions of X^n into smaller and smaller fibers of the projection $X^n \to X^i$ as i runs through $n, n-1, \ldots$.

All we need to make the argument work is the preservation of the 1-Lipschitz property under the averaging of functions f on X^{i+1} along the fibers of $X^{i+1} \to X^i$. This can be expressed in terms of our Lid-metric. We suggest that the reader work out the details and consult §7 in [Mil- Sch] for basic information and applications of martingales to the concentration problem.

Now we derive from the bound on $\operatorname{ObsDiam} X^n$ for the ℓ_1 cartesian power the corresponding inequality for the ℓ_p-product. To do this, we observe that our ℓ_1-bound

$$\operatorname{ObsDiam} X^n = O(\sqrt{n})$$

is uniform for all spaces X with $\operatorname{diam} X \leq 1$. In particular, it holds true with the same implied constants for the space X_k consisting of k atoms of weight $1/k$ with *unit* mutual distances. Then the ℓ_p-metric on the cartesian power X_k^n obviously satisfies

$$\operatorname{dist}_{\ell_p} = (\operatorname{dist}_{\ell_1})^{1/p},$$

and every 1-Lipschitz function f on $(X_k^n, \mathrm{dist}_{\ell_1})$ is $n^{11/p}$-Lipschitz for the ℓ_p-distance on X_k^n. Conversely, let f be 1-Lipschitz on (X_k^n, ℓ_p). We restrict f to a maximal δ-separated net $Y_\delta \subset X_k^n$, $\delta = n^{1/2p}$ and observe that f is λ-Lipschitz on $(Y_\delta|_{\mathrm{dist}_{\ell_1}})$ for $\lambda = \delta/\delta^p = n^{1/2p-1/2}$. This f extends to a λ Lipschitz function f' on all (X_k^n, ℓ_1), where clearly $\|f' - f\|_{L_\infty} \le \lambda\delta^p + \delta = 2\delta$. The function f' pushes the essential part of the measure of X_k^n to a segment in $\mathbb{R}$ of length about $\lambda\sqrt{n} \approx n^{1/2p}$, since $\mathrm{ObsDiam}(X_k^n, \mathrm{dist}_{\ell_1}) = O(\sqrt{n})$. Hence

$$\mathrm{ObsDiam}(X_k^n, \mathrm{dist}_{\ell_p}) = O(n^{1/2p}). \tag{$*$}$$

Finally, we recall that the spaces X_k majorize, in the limit as $k \to \infty$, all spaces X with $\operatorname{diam} X \le 1$. This means that every such X appears as a $\underline{\square}_1$-limit of some spaces $Y_k \prec X_k$ as $k \to \infty$. This allows our extension of $(*)$ to all spaces X with $\operatorname{diam} X \le 1$, since the bound $(*)$ is uniform in k. Thus,

$$\mathrm{ObsDiam}(X^k) = O(n^{1/2p}) \tag{$\star$}$$

for all $p \ge 1$ and all spaces X with $\operatorname{diam} X < \infty$. (We suggest that the reader quantify the argument and observe that everything is uniform in $p \to \infty$.)

(4) Concentration of alternating functions on cartesian powers. Whenever we have an mm space acted upon by a group G, we can ask for the concentration behavior of functions f on this space transformed by G according to a given representation of G. A particular example of interest is the power space $X^n = X \times X \times \cdots \times X$ acted upon by the permutation group S_n. Then one distinguishes two spaces of functions f on X^n: *symmetric* and *alternating*, where the former means invariance under S_n, i.e.,

$$f(x_1, \dots, x_i, x_{i+1}, \dots, x_n) = f(x_1, \dots, x_{i+1}, x_i, \dots, x_n),$$

and the latter signifies the change of sign under odd permutations

$$f(x_1, \dots, x_i, x_{i+1}, \dots, x_n) = -f(x_1, \dots, x_{i+1}, x_i, \dots, x_n).$$

The symmetric functions can be viewed as functions on the quotient space X/S_n, and the concentration properties of this space do not seem to be much better (at least to a casual eye) than those of X^n (but a significant difference may emerge under closer inspection, compare (5) below).

The alternating functions on X^n do not come from any (commutative, compare [Connes]) quotient of X^n, but they can be studied directly on X^n by looking, for example, at the *alternating spectrum* of X consisting of the eigenvalues of Δ acting on the space of alternating functions. Every n-tuple of pairwise nonequal eigenfunctions on X, say $\varphi_{k_1}, \varphi_{k_2}, \ldots, \varphi_{k_n}$ defines an alternating eigenfunction

$$\Phi = \Phi(x_1, \ldots, x_n) = \det(\varphi_{k_i}(x_j))$$

on X^n, called an *orbital* in quantum mechanics, with the eigenvalue $\lambda_{k_1} + \lambda_{k_2} + \cdots + \lambda_{k_n}$. The lowest orbital, made with the first eigenfunctions $\varphi_0, \varphi_1, \ldots, \varphi_{n-1}$ on X, gives us the lowest alternating eigenvalue of X^n by an elementary linear algebraic argument. Thus,

$$\lambda_1^{alt}(X^n) = \sum_{i=0}^{n} \lambda_i(X) \tag{$*$}$$

for all compact Riemannian manifolds X. It follows that $\lambda_1^{alt}(X^n)$ *is asymptotic to* $n^{1+2/d}$ *for* $d = \dim X$, *and the (obviously defined) alternating observable diameter of* X^n *decays like* $n^{-1-2/d}$. (This fast decay is closely related to the stability of matter, see [Lieb]).

Problem: Evaluate the alternating visual diameter for the ℓ_p-power X^n of a given mm space X. Study in general the asymptotic geometry of X^n with the action of S_n taken into account. In particular, decide for which sequences of scaling numbers ρ_i the sequence $\rho_i X^n$ concentrate or dissipate and what happens on the *critical scale* where ρ_i is asymptotic to $(\mathrm{ObsDiam}(X^n))^{-1}$ (with a given type of symmetry).

(5) Euclidean geometry of simplices and ℓ_p-balls. Besides products, such as the cube $[0,1]^n$, there are many other interesting spaces, e.g., polyhedra with S_n actions. The first among these is the Euclidean simplex

$$\Delta^n = \{x_0, \ldots, x_n \geq 0 : \sum_{i=0}^{n} x_i = 1\},$$

where we start by observing that the distance function from a vertex gives us the measure $\frac{x^n}{n!}\,dx$ on $\mathbb{R}_+$ whose essential part spreads over the segment of length $1/n$ around zero. Thus, the observable diameter of Δ^n must be *at least* const $/n$. Next, we recall that the moment map $\mathbb{C}\mathrm{P}^n \to \delta^n$ (see $3\frac{1}{2}.1$) pushes the $U(n+1)$-invariant measure of $\mathbb{C}\mathrm{P}^n$ forward to the Lebesgue measure on Δ^n, and since this map is 1-Lipschitz, we have

$$\mathrm{ObsDiam}\,\Delta^n \leq \mathrm{ObsDiam}\,\mathbb{C}\mathrm{P}^n = O\left(\frac{1}{\sqrt{n}}\right).$$

In fact,

$$\mathrm{ObsDiam}\,\Delta^n = O\left(\frac{1}{n}\right), \tag{$\star$}_\Delta$$

as follows from Schechtman's theorem (see below), but even the $o(1/\sqrt{n})$-concentration of the distance function on Δ^n from the center is not totally obvious. This concentration can be expressed in terms of the volumes of the intersections of Δ^n with the Euclidean R-balls in $\mathbb{R}^{n+1} \supset \Delta^n$ with variable R, and then the relevant information about Δ^n follows from the corresponding volume properties of intersections of the Euclidean and ℓ_1-balls in $\mathbb{R}^{n+1}$. In fact, Schechtman and Schuckenschläger (see [Sch–Sch]) evaluated the intersections between ℓ_p- and ℓ_q-balls in $\mathbb{R}^n$ for all p and q as follows.

Let $B_p \subset \mathbb{R}^n$ denote the ball with respect to the ℓ_p-metric of radius R_p, such that the Euclidean volume of $B_p = 1$. For example, if $p = 2$, this is the ordinary Euclidean ball of radius

$$(\Gamma(1+n/2))^{1/n}/2\Gamma(3/2) \approx \sqrt{n}$$

and if $p = 1$, this is the polyhedron

$$B_1 = \{x_1, \dots, x_n : \sum_{i=1}^{n} |x_i| \le \frac{(n!)^{1/n}}{2} \approx n\},$$

where the faces are the copies of Δ^{n-1} scaled by $\lambda_n \approx n$. *Then there is a critical constant* $c_{pq} > 1$ *(explicitly computed in [Sch- Sch]) such that as* $n \to \infty$, $\mathrm{vol}(B_p^n \cap tB_q^n) \to 0$ *for* $t < C_{pq}$ *and* $\mathrm{vol}(B_p^n \cap tB_q^n) \to 1$ *for* $t > C_{pq}$. We use this for $p = 1$ and $q = 2$ and conclude that the distance function $x \mapsto \|x\|_{\ell_2}^2$ on the unit ℓ_1-ball $\{x_1, \dots, x_n : \sum_{i=1}^{n} |x_i| \le 1\}$ has concentration $o(1/\sqrt{n})$ for $n \to \infty$. This implies the similiar concentration on the boundary $\partial B_1^n(1)$ and hence on the simplex Δ^{n-1}.

Spectral evaluation of $\mathrm{ObsDiam}\, B_p^n(1)$**.** We recall that the (Lebesgue measure on) the ℓ_p ball

$$B_p^n(n^{1/p}) = \{x_1, \dots, x_n : \sum |x_i|^p \le n\}$$

is well-approximated by the (canonical) product measure $\exp\left(\sum |x_i|^p\right) dx$ on $\mathbb{R}^n$ (see $3\frac{1}{2}.25$). Since this measure has $\lambda_1 \ge \varepsilon > 0$, it has ObsDiam $=$ $O(1)$ and thus we have a similar bound for the above balls. In fact, by rescaling these ℓ_p-balls to unit size and then taking $3\frac{1}{2}.25$ into account, we arrive at the following result (communicated to me by G. Schechtman)

$$\mathrm{ObsDiam}\, B_p^n(1) = O(n^{-1/p}). \qquad (\star)_p$$

Notice that $(\star)_\Delta$ easily follows from $(\star)_1$.

Remarks and open problems: (a) Probably the above argument can be adjusted for the evaluation of the spectral radii of the ℓ_p-balls and that of Δ^n, as well as for the study of the alternating observable diameters of these spaces.

(*b*) There are many other interesting polyhedra with S_n-actions besides $[0,1]^n$, $B_1^n(1)$, and Δ^n, where the problem of evaluating the observable diameter (with and/or without symmetries) begs for solution. An attractive case is that of the space of *bistochastic matrices*,

$$\{x_{ij} \geq 0 : \sum_{i=1}^{n} x_{ij} = 1 \text{ for } j = 1, \dots, n, \text{and } \sum_{i=1}^{n} x_{ij} = 1 \text{ for } i = 1, \dots, n\}$$

and such generalizations as

$$\{x_{ijk} \geq 0 : \sum_{i} x_{ijk} = 1, \sum_{j} x_{ijk} = 1, \sum_{k} x_{ijk} = 1\}.$$

Furthermore, one can start with a given (convex) subset $B \subset \mathbb{R}^m$ and make, for example,

$$B_n(R) = \{x_{ij} \in B : \sum_{i=1}^{n} x_{ij} = R = \sum_{j=1}^{n} x_{ij}\},$$

leading to the same spectrums of open problems (compare [Barv]).

(*c*) The canonical (and microcanonical) measures on $\mathbb{R}^n$ arising in statistical mechanics provide us with an inexhaustible source of examples where we wish to evaluate our invariants for $n \to \infty$.

$3\frac{1}{2}$.63. Fine geometry in $\mathcal{X}$. Eventually, one wants to study the asymptotic mm geometry of particular (sequences of) spaces with much greater precision than provided by the metric and other structures on $\mathcal{X}$. The major examples are supplied by various *configuration spaces* coming from geometry, probability theory, statistical mechanics, and quantum field theory. A famous case is the *self-avoiding random walk* on the standard lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$. Here, our X^n is the space of *injective* maps x from $\{1, 2, \dots, n\}$ to $\mathbb{Z}^n$ with $\operatorname{dist}(x(i), x(i+1)) = 1$ for $i = 1, \dots, n-1$. One gives equal measures to all atoms x constituting X^n (maybe one should introduce smart weights?) and wants to understand the asymptotic behavior of functions $f(x)$ on X^n recording the geometry of a walk $x\colon \{1, 2, \dots, n\} \to \mathbb{Z}^2$, e.g., the diameter of its image. Amazingly, no *nontrivial rigorous* result seems to be available for these X^n.

What can the space $\mathcal{X}$ offer us as a tool for approaching real, hard problems? As it stands, precious little. Actually, even rather crude invariants, such as the first eigenvalue $\lambda_1(X)$ is *not* continuous on $(\mathcal{X}, \underline{\square}_1)$, which shows how coarse the geometry of $\mathcal{X}$ is. There are several avenues for improvement. For example, one can make λ_1 more continuous on $\mathcal{X}$ by using the following standard ε-regularization trick of setting

$$\lambda_1^{\varepsilon}(X) = \sup_{X'} \lambda_1(X'),$$

where X' runs over all mm spaces with $\underline{\square}_1(X, X') \leq \varepsilon$. (We suggest that the reader think of what this trick does to the whole spectrum.) However, it seems unlikely that a truly deep asymptotic theory can be developed for all spaces in $\mathcal{X}$. Apparently, one should distinguish a certain class (or classes) or *regular* spaces X (as well as of regular observables f on these X) and then introduce a fine "smooth" structure on the space $\mathcal{X}_{\mathrm{reg}} \subset \mathcal{X}$ of the regular X's. Here are some properties which make our mm space look regular.

1. All λ_i are defined and are bounded away from zero. If we deal with a sequence X^n, then the $\lambda_i(X^n)$ must have "regular" asymptotics.

2. The ratios of measures of small balls of given radii in an individual X are bounded, and such balls, say $B(R)$ and $B(r)$ in X^n, have "regular asymptotics" for $\mu(B_n(R))/\mu(B_n(r))$, e.g., as $(R/r)^{\alpha n}$ for $n \to \infty$.

3. The isoperimetric properties of X^n behave regularly for $n \to \infty$.

4. Regular observables $f : X^n \to \mathbb{R}^k$ should not only concentrate, but the pushforward measures $\underline{\mu}_n = f_*(\mu)$ on $\mathbb{R}^k$ must have reasonable density functions $s_n(y)$ (with respect to the Lebesgue measure on $\mathbb{R}^k$), such that $(1/n) \log s_n(y)$ would converge as $n \to \infty$ to a nice (concave, piecewise analytic, etc.) function $\sigma(y)$ on $\mathbb{R}^k$.

Additionally, the "fine structure" on $\mathcal{X}_{\mathrm{reg}}$ can be expressed, for example, by a metric which is (much) stronger than $\underline{\square}_1$, but more likely it should be a more sophisticated infinitesimal structure reflecting particular types of asymptotics of X^n as those established (or at least expected) in statistical mechanics. We humbly hope that the general ambiance of $\mathcal{X}$ can provide a friendly environment for treating asymptotics of many interesting spaces of configurations and maps.

Chapter 4

Loewner Rediscovered

A. First, some history (in dimension 2)

In 1949, Loewner proved the following (unpublished, see [Pu] or the proofs in [Berger]$_{\text{Cours}}$, [Berger]$_{\text{ombre}}$, [Berger]$_{\text{LGRG}}$):

4.1. Theorem: *Let* $(\mathbb{T}^2, g)$ *be a 2-torus equipped with a Riemannian metric* g *and let* $l(g)$ *be the infimum of the lengths of all closed curves in* $\mathbb{T}^2$ *not homotopic to* 0. *Then* $\text{Area}(\mathbb{T}^2, g) \geq (\sqrt{3}/2)l^2(g)$. *Moreover, if* $\text{Area}(\mathbb{T}^2, g) = (\sqrt{3}/2)l^2(g)$, *then* $(\mathbb{T}^2, g)$ *is necessarily the equilateral flat torus, i.e., defined by* $\mathbb{R}^2/\Lambda$, *where* Λ *is the lattice in* $\mathbb{R}^2$ *generated by the vectors* $(0,1)$ *and* $(1/2, \sqrt{3}/2)$.

The proof of Loewner's theorem begins with an application of the fundamental theorem of conformal representation in order to write $g = f \cdot g_0$, where f is a function on $\mathbb{T}^2$ and g_0 is flat. Since g_0 is flat, it has a transitive group G of isometries; an appropriate averaging of f by the action of G has the effect of decreasing area and increasing l. Since G is transitive, this averaging process yields a constant function, and one is reduced to the case of a flat torus, for which the result is simple (compare [Keen]).

In [Pu], P.M. Pu remarks that the Loewner method can be directly applied to the projective plane $(\mathbb{R}\mathrm{P}^2, g)$ to give (see [Berger]$_{\text{Cours}}$ or [Berger]$_{\text{LGRG}}$, p. 303)

4.2. Proposition: *Given the projective plane* $\mathbb{R}\mathrm{P}^2$ *equipped with a Riemannian metric* g, *let* $l(g)$ *denote the infimum of the lengths of closed curves not homotopic to zero. Then* $\text{Area}(\mathbb{R}\mathrm{P}^2, g) \geq (2/\pi)l^2(g)$, *and if* $\text{Area}(\mathbb{R}\mathrm{P}^2, g) = (2/\pi)l^2(g)$, *then* g *is necessarily the canonical metric on* $\mathbb{R}\mathrm{P}^2$.

4.3. The Loewner technique does not apply to surfaces V of genus > 1 (except asymptotically, see [Katok]), but by the method of harmonic forms and, independently, Accola and Blatter (see [Acc], [Blat]) we have

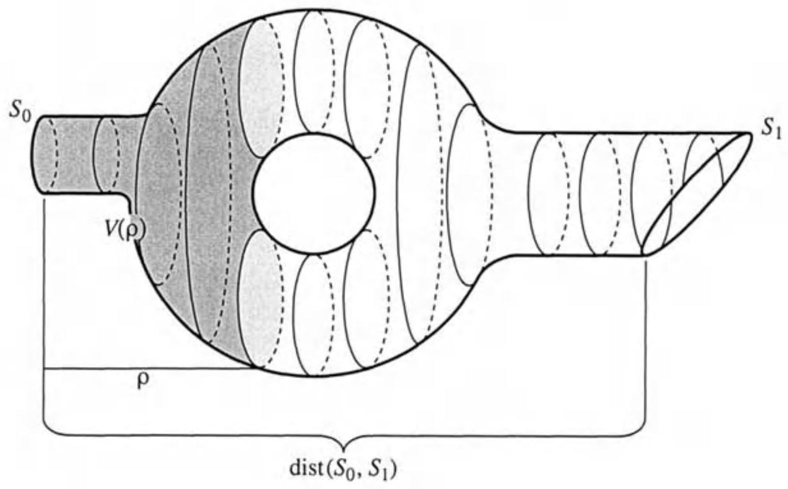
4.4. Theorem: *For each $\gamma > 1$, there is a constant $C(\gamma) > 0$ such that* $\mathrm{Area}(V, g) > C(\gamma) l^2(g)$ *for any surface (V, g) of genus γ.*

4.5_+ . The constant $C(\gamma)$ guaranteed by the method of Accola and Blatter is far from optimal. In fact, these authors rely on the embedding of a Riemann surface into its Jacobian (torus) and then apply the Minkowski theorem (see 4.31). This gives $C(\gamma) \sim \gamma^{-1}$, which tends to zero as $\gamma \to \infty$. On the other hand, one can get a universal constant independent of γ with the following

$4.5\frac{1}{2}_+$ Besicovitch lemma (compare [Besi] and 4.28bis). *Let V be a compact surface with two boundary components $\partial V = S_0 \sqcup S_1$, and let ℓ denote the length of the shortest closed curve in V homologous to S_0 (and hence to S_1) with respect to a given Riemannian metric g on V. Then the area of (V, g) is bounded from below by*

$$\mathrm{Area}(V, g) \geq \ell \, \mathrm{dist}(S_0, S_1). \qquad (*)$$

Sketch of the proof. Look at the levels $S(r)$ of the distance function $r(x) = \mathrm{dist}(x, S_0)$ on V.



These start with $S(0) = S_0$ and remain closed (possibly nonsmooth and disconnected) curves homologous to S_0 for $r \leq \mathrm{dist}(S_0, S_1)$. The classical (and rather obvious) *coarea formula* says that the area $a(\rho)$ of the region $V(\rho) = \{x \in V : r(x) \leq \rho\}$ satisfies

$$a(\rho) = \int_0^\rho \operatorname{length} S(r)\, dr,$$

and so

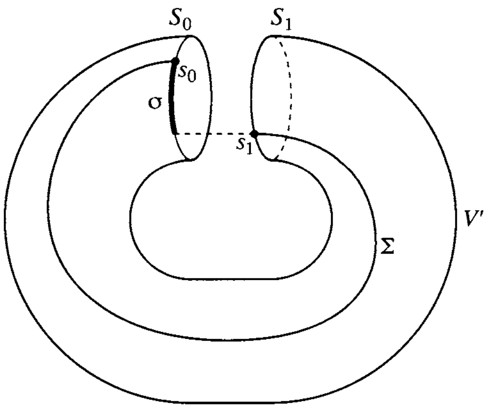
$$\operatorname{Area}(V) \geq \operatorname{Area} V(\rho) = a(\rho) \geq \ell\rho$$

for all $\rho \leq \operatorname{dist}(S_0, S_1)$, which yields the lemma for $\rho = \operatorname{dist}(S_0, S_1)$.

$4.5\frac{3}{4}{+}$ Corollary. *Every closed, orientable surface V of genus ≥ 1 admits a closed curve of length ℓ which is not null-homologous and satisfies*

$$\frac{1}{2}\ell^2 \leq \operatorname{Area}(A). \qquad (**)$$

Proof. Cut V along the *shortest* non-dividing (and hence not null-homologous) curve S and apply Besicovitch's lemma to the resulting surface V' with the two boundary components S_0 and S_1 corresponding to S.

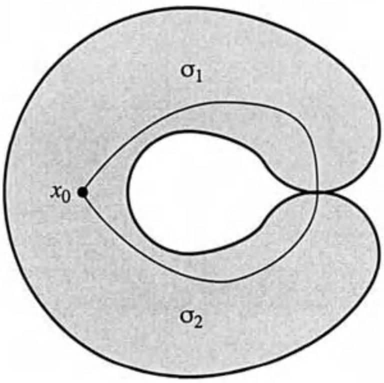


Then, according to $(*)$, there is a curve Σ in V' of length $\delta \leq (\operatorname{Area}(V'))/\ell$ joining two points $s_0 \in S_0$ and $s_1 \in S_1$. This Σ, together with a segment $\sigma \in S$ of length $\leq \frac{1}{2} \operatorname{length} S = \frac{1}{2}\ell$ make a closed curve in V which is not homologous to zero and has length $\leq \delta + \ell/2$, which cannot be shorter than ℓ. Thus, $\delta + \ell/2 \geq \ell$, or $\ell \leq 2\delta$, and so

$$\operatorname{Area}(V) = \operatorname{Area}(V') \geq \ell\delta \geq \frac{1}{2}\ell^2.$$

This argument using the *coarea method* is more elementary than Loewner's *length-area* proof (as it is called by complex analysts), but it does not give us the sharp constant for the torus. Still, it seems to serve better than the Jacobian embedding of Accola–Blatter, at least for large genus γ. And yet, the Jacobian is vindicated by other geometric applications. For example, the very discrepancy between the Accola–Blatter $C(\gamma) \sim \gamma^{-1}$ and the (asymptotically) optimal $C_0(\gamma) \sim \gamma/(\log(\gamma))^2$ (see below) led Buser and Sarnak to a new approach to the *Schottky problem* on the characterization of Jacobi varieties of algebraic curves (see [Gro]$_{\mathrm{SII}}$ and the references therein).

4.6$_+$ Should we be content with having improved $C(\gamma) \sim \gamma^{-1}$ to $C(\gamma) \geq 1/2$? To get a perspective, try surfaces V of genus $\gamma \geq 2$ with a metric g of constant curvature -1. An obvious way (going back to Gauss and Minkowski) of finding a *noncontractible* loop in V based at a given point $x_0 \in V$ is to consider the maximal simply connected open(or minimal non-simply connected closed) disk $B(x_0, R)$ which meets itself at the boundary.



Then the loop S based at x_0 and comprised of two geodesic segments σ_1 and σ_2 joining x_0 with a meeting point at the boundary $\partial B(x_0, R)$ has length $\ell_0 = 2R$. On the other hand, Area $V \geq$ Area $B(x_0, R) \simeq e^R$ for large R, say for $R \geq 1$. On other other hand, the area of this $V = (V, g_0)$ is proportional to the Euler characteristic $\chi(V) = 2(\gamma - 1)$, i.e., Area$(V) = 4\pi(\gamma - 1)$. Thus, $\ell_0 \leq C \log(\gamma)$ and Area$(V, g_0) \approx \gamma$, which implies that

$$\mathrm{Area}(V, g_0) \geq C_0(\gamma)\ell_0^2$$

for

$$C_0(\ell) \geq \mathrm{const}\, \gamma/(\log \gamma)^2$$

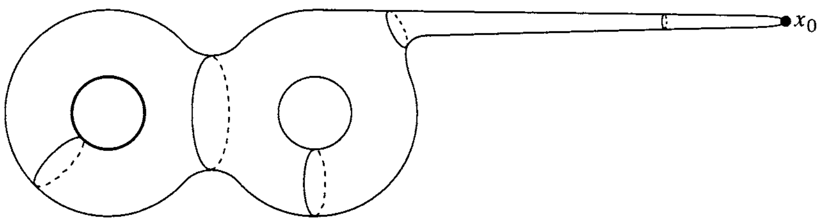
and some universal constant const > 0. This suggests that $C(\gamma) \sim \gamma/(\log\gamma)^2$ for all metrics g. Indeed we have the following

Theorem (see [Gro]$_{\text{SII}}$). *An arbitrary closed (possibly nonorientable) surface (V, g) admits a closed geodesic nonhomologous to zero mod 2 of length ℓ, such that*

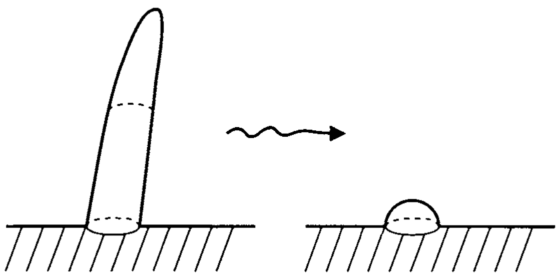
$$\operatorname{Area}(V, g) \geq C(\gamma)\ell^2 \qquad (*)$$

for $C(\gamma) = \text{const}\,\gamma/(\log\gamma)^2$, *where* const > 0 *is a universal constant.*

About the proof. If we try to find a short loop S based at a *given* point $x_0 \in V$, then we run into trouble if x_0 lies at the tip of a long narrow finger.



So, we must avoid such x_0 and look for S somewhere else. In fact, we can cut away such long fingers and replace them with hemispherical caps.



This does not shorten the length of the minimal curves nonhomologous (and/or nonhomotopic) to zero and diminishes the area of V at the same time. So, we may reduce the general case to a "fingerless" one where the balls $B(R) \subset V$ harboring no loops which do not contract in M, satisfy

$$\operatorname{Area} B(R) \geq R^2. \qquad (+)$$

This, combined with the above argument employed for the metric g_0 of constant curvature, yields a noncontractible curve of length $\ell \leq 2\sqrt{\operatorname{Area}(V, g)}$

which is actually slightly worse than the above (**). However, if we fully exploit the inequality (+) for *all* balls in V, and not just for a single $B(R) = B(x_0, R)$, then we can do much better by relating the area of V (satisfying (+)) to the *simplicial volume* of V (see Ch. 5.F_+) which happens to be equal to $-c_0\chi(V)$ for *surfaces* V. (See [Gro]$_{\mathrm{FRM}}$ and [Gro]$_{\mathrm{SII}}$ for the actual proof).

4.7$_+$ It is not at all obvious what the extremal metrics g on surfaces of genus γ with the *maximal* ratio $\mathrm{Area}(V, g)/\ell^2(g)$ should be. They are definitely *not* of constant curvature for $\gamma \geq 2$. See [Cal] for some information about these g_{extreme}.

B. Next, some questions in dimensions ≥ 3

4.8. As a reference for the discussion in this section, the reader may consult [Berger]$_{\mathrm{GR}}$ and [Berger]$_{\mathrm{Pu}}$. We now consider the case of manifolds of dimension ≥ 3. In [Pu], Pu notes that the Loewner technique applies to conformal Riemannian structure to homogeneous Riemannian structures; unfortunately, the space of conformal classes of Riemannian structures in dimension 3 is infinite-dimensional and thus any inequality similar to that of Loewner's theorem for a single conformal class is of limited interest.

4.9. Note that if (V, g) is a Riemannian n-manifold and if we continue to denote by $\nu(g)$ its volume and $l(g)$ the infimum of the lengths of closed curves not homotopic to zero, then it is futile to hope for the existence of a constant $C(V) > 0$ such that $\nu(g) \geq C(V)l^n(g)$ for any nonsimply connected V. For example, if W is simply connected, then the ratio $\nu(g)/l^n(g)$ can be made arbitrarily small for the manifold $S^1 \times W$ equipped with the product metric.

In order for $C(V)$ to be positive, it appears to be necessary, as in the case of $\mathbb{T}^2$ or $\mathbb{RP}^2$, that the degree-1 homology generate all of the topology of V, or at very least the fundamental class of V. The likely candidates for such a theorem might therefore pertain to $\mathbb{T}^n$ and $\mathbb{RP}^n$. Thus,

4.10. Question: *Does there exist a constant $C_{\mathbb{T}}(n) > 0$ such that $\nu(g) \geq C_{\mathbb{T}}(n)l^n(g)$ for every Riemannian metric g on $\mathbb{T}^n$? If so, does equality characterize the flat tori for which the ratio is minimal? (For the corresponding, often nonunique lattices, called* critical lattices, *see [Cass], p. 141, and [Klin], p. 133). The same question (cf. 4.7) can be posed with $\nu(g) \geq C(n)l_1(g)\cdots l_n(g)$.*

4.11. Question: *The same question as in 4.10 can also be posed with* $C_P(n)$ *for* $\mathbb{R}P^n$, *with equality implying that* $(\mathbb{R}P^n, g)$ *is isometric to* $\mathbb{R}P^n$ *with its canonical metric.*

4.12. Next, we can pose the following generalization of Question 4.11: Let $K\mathrm{P}^n$ denote the projective space of k-dimension n over the field K equal to $\mathbb{C}$ if $k = 2$, $\mathbb{H}$ if $k = 4$, and **Ca** if $k = 8$ (so that necessarily $n = 2$). The real dimension is then kn. Given a Riemannian structure g on $K\mathrm{P}^n$, we denote by $\nu(g)$ its volume and by $m(g)$ the infimum of the volumes of k-dimensional submanifolds of $K\mathrm{P}^n$ that are homologous to the (canonically embedded) projective line $K\mathrm{P}^1$.

4.13. Question: *Is there a constant* $C(n, k) > 0$ *such that for every metric* g *on* $K\mathrm{P}^n$, *we have* $\nu(g) \geq C(n, k)m^n(g)$? *What does equality imply?*

To conclude, consider the product $V = V_1 \times V_2$ of two compact manifolds. Given a Riemannian structure on V, let $\nu_i(g)$ denote the infimum of the volumes of submanifolds of dimension $n_i = \dim(V_i)$ and homologous to V_i.

4.14. Question: *Under what conditions on the* V_i *does there exist a constant* $C(V_1, V_2)$ *such that*

$$\nu(g) \geq C(V_1, V_2)\nu_1(g)\nu_2(g)$$

for every g *on* $V_1 \times V_2$? *What does equality imply?*

We note that if $V_1 = V_2 = S^1$ (cf. 4.7), or, more generally, if each V_i is a torus, then a more subtle definition of the $\nu_i(g)$ is needed in order for such a constant to exist.

In the following section, we will provide answers to most of these questions.

C. Norms on homology and Jacobi varieties

4.15. Norms on the homology groups of a Riemannian manifold. Recall that a *singular integral Lipschitz chain* (resp. *real chain*) of dimension p in a Riemannian manifold V is a formal sum of the form $a = \sum_i \lambda_i c_i$, where the λ_i are integers (resp. real numbers) and the c_i are Lipschitz mappings of the standard p-simplex into V. The *mass* of a Lipschitz 1-chain a is defined as the number $M(a) = \sum_i |\lambda_i| \operatorname{length}(c_i)$. Evidently, $M(a) = \sup |\int_a \omega|$, where ω ranges over all differential 1-forms for which $\sup_{v \in V} |\omega(v)| \leq 1$ where $|\cdot|$ is the Euclidean norm on $T_v^* V$.

In order to equip the space of p-chains with a norm that has the same additivity property and that is dual to a sup-like norm on the space of p-forms, we proceed as follows (cf. [Laws], p. 51, and [Fed]$_{\mathrm{GMT}}$).

Given a Euclidean space E, we denote its norm and the induced norms on the exterior powers $\Lambda^p E$ by $|\cdot|$. The *mass* norm on $\Lambda^p E$ is then defined by

$$\|e\| = \inf\left\{\sum_i |e_i| : e = \sum_i e_i, \text{ each } e_i \text{ nondecomposable}\right\},$$

and, on $\Lambda^p E^*$, we define the *comass* as

$$\|\varphi\|^* = \sup\{|\varphi(e)| : e \in \Lambda^p E \text{ and } \|e\| \leq 1\}.$$

The *comass* of a differential form ω is by definition

4.16.

$$M^*(\omega) = \sup |\omega(\tau_1, \ldots, \tau_p)|,$$

where the sup ranges over the orthonormal frames in $T(V)$. The mass of a chain is the dual norm,

4.17.

$$M(a) = \sup\left\{\left|\int_a \omega\right| : \omega \text{ is a differential form with } M^*(\omega) \leq 1\right\}.$$

These definitions imply that $M(\sum_i \lambda_i c_i) = \sum_i |\lambda_i| \operatorname{vol}(c_i)$. For example, in $\mathbb{C}P^2$ equipped with its standard metric, as well as in every Kähler manifold, the comass of the Kähler form ω is 1, as follows from Wirtinger's inequality (see 6.35).

We now equip $H_p(V;\mathbb{Z})$ and $H_p(V;\mathbb{R})$ with quotient norms by defining $\|\alpha\| = \inf\{M(a) : a \in \alpha\}$ for any integer or real class α. By 5.1.6 of [Fed]$_{\mathrm{GMT}}$ (or [Fed–Flem], 9.6, completed by [Fed]$_{\mathrm{VP}}$, §3 for the real case), these are indeed norms. Let α denote an integer class and $\alpha_{\mathbb{R}}$ the real class that it determines. Since $\alpha \in \alpha_{\mathbb{R}}$, we have $\|\alpha\| \geq \|\alpha_{\mathbb{R}}\|$, and the following theorem shows that the norm $\alpha \mapsto \|\alpha_{\mathbb{R}}\|$ on $H_p(V;\mathbb{Z})$ plays the same role as the limit norm on the homotopy groups introduced in Proposition 2.22.

4.18. Theorem (cf. [Fed]$_{\mathrm{VP}}$ §5): *For each class $\alpha \in H_p(V;\mathbb{Z})$,*

$$\lim_{m\to\infty} \frac{\|m\alpha\|}{m} = \|\alpha_{\mathbb{R}}\|.$$

Remark: Examples are known for which $\|\alpha\| > \|\alpha_{\mathbb{R}}\|$ and even for which $(1/m)\|m\alpha\| > \|\alpha_{\mathbb{R}}\|$ for all m. This phenomenon has been described by

F. Almgren (cf. [Fed]$_{VP}$, p. 397) in the case of a degree-1 class in an oriented 3-manifold. In [Laws], there are also striking examples of flat tori and a systematic study of the norms $\|\alpha\|$ and $\|\alpha_{\mathbb{R}}\|$ (see [Fed]$_{VP}$, p. 394, and section 4.33 below).

4.19. Definition: The norm $\alpha \mapsto \|\alpha_{\mathbb{R}}\|$ on $H_p(V;\mathbb{Z})$ is called the *stable norm.* An integral class α is called *stable* if there exists an integer m such that $(1/m)\|m\alpha\| = \|\alpha_{\mathbb{R}}\|$.

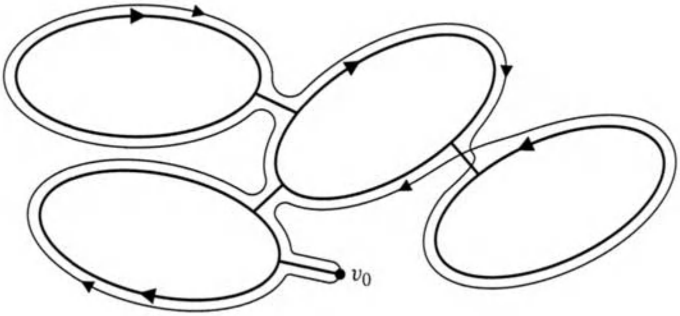
Remark: To the extent that we do not always obtain the best inequalities, the introduction of the mass and comass is not really necessary, since they are equivalent to the norms defined in an analogous way starting with the Euclidean norm on T_v^*V. Nevertheless, the mass and comass help to see that for $p > 2$, the quotient norm on $H_p(V;\mathbb{R})$ is not Euclidean in general, even if V is $S^2 \times S^2$ or a flat torus.

4.20. The Hurewicz homomorphism. In one dimension, the Hurewicz homomorphism (see [Hur–Wall], p. 148) is easy to visualize: Let $\ell\colon S^1 \to V$ be a loop based at a point $v \in V$. The various ways to triangulate S^1 enable us to associate with ℓ an integral 1-chain whose mass equals the length of ℓ. Additionally, the homology class of this chain depends only on the homotopy class of ℓ, so that the *Hurewicz homomorphism* $h\colon \pi_1(V,v) \to H_1(V;\mathbb{Z})$ is well defined.

$4.20\frac{1}{2}{}_+$ Comparison between the length and the mass. Assume V is connected, fix a point $v_0 \in V$ and recall the (length) norm V on the fundamental group $\pi_1(V,v_0)$ which measures each class of loops in $\pi_1(V,v_0)$ by the length of the shortest representative. Now, we pass this norm to $H_1(V)$ via the Hurewicz homomorhism by defining length(α) for $\alpha \in H_1(V)$ as the infimum of lengths of the loops in V based at v_0 which represent (i.e., are homologous to) α. Clearly, this length majorizes the mass $\|\alpha\| = \|\alpha\|_{H_1}$ as the definition of the latter allows a representation of α by a possibly *disconnected* collection of closed curves in V. It is equally obvious that length$(\alpha) \le C\|\alpha\|_{H_1}$ for some constant $C = C_V$, since every collection of closed curves can be replaced by a single one in the same homotopy class without drastically increasing the total length. Moreover, we have the following sharpening of this.

$4.20\frac{1}{2}$bis$_+$ Lemma: *If V is compact, then there exists a function $C(\ell) = C_V)(\ell)$ satisfying $C(\ell) \to 0$ for $\ell \to \infty$ such that*

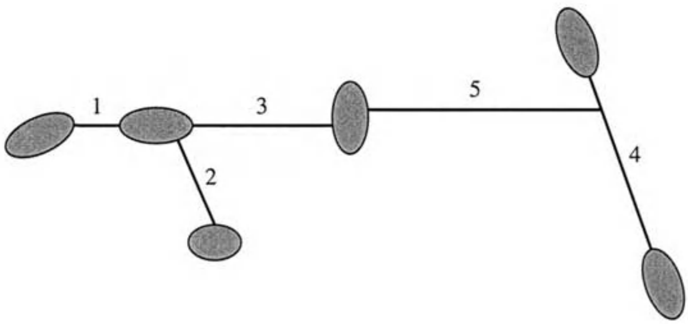
$$1 \le \text{length}(\alpha)/\|\alpha\|_{H_1} \le 1 + C(\text{length}(\alpha)). \qquad (\circ)$$



Proof. What we need is a system of bridges connecting a collection of closed curves in V, so that the total length of these bridges is small compared to the sum of the lengths of the curves. Since all curves may be assumed to be noncontractible, their total length is at most a constant times (number of components), and so our claim follows from:

$4.20\frac{2}{3}_+$ Sublemma: *Every collection of N connected subsets in V can be connected by geodesic segments of total length $\leq \mathrm{const}_V\, N^{1-1/n}$, $m = \dim V$.*

Proof. Every nonempty subset in V has the ε-neighborhood of volume $\geq \varepsilon^n$, and so there is an edge of length $\leq N^{-1/n}$ between some pair of our subsets. Then there is an edge of length $\leq (N-1)^{1/n}$ between the resulting $N-1$ connected subsets, and so on.



Their total length is $\leq \sum_{i=1}^{N} i^{-1/n} \simeq N^{1-1/n}$ for $n \geq 1$ and the proof follows with a special (trivial) consideration of the case $\dim V = 1$.

Remarks $_+$: (*a*) Since the curves in question have diameters $\geq \delta > 0$, their ε-neighborhoods have volumes $\geq \delta\varepsilon^{n-1}$, and so one can bound $C(\ell)$ by $C(\ell) \leq \mathrm{const}_V\, \ell^{-1/(n-1)}$ for $n \geq 2$ (where the case $n = 2$ needs a little extra care).

(*b*) The original version of the lemma (4.20bis in the French edition) claimed an even stronger bound, $C(\ell) \leq \mathrm{const}_V\, \ell^{-1}$, which amounts to

$$\mathrm{length}(\alpha) - \|\alpha\| \leq \mathrm{const}_V, \tag{$\Box$}$$

but the accompanying proof was incorrect. Nevertheless, the inequality ($\Box$) is valid and follows trivially from a theorem due to D. Burago cited in 3.16 which implies that

$$\mathrm{length}(\alpha) \leq \|\alpha\|_{H_1}^{\lim} + \mathrm{const}_V, \tag{$\triangle$}$$

where the limit norm equals, as we know, $\|\alpha_{\mathbb{R}}\|$. We suggest that the reader prove ($\Box$) independently of ($\triangle$), but I must admit that I have no recollection of whether such an argument is easily given.

4.21. The Jacobi variety. Here we will assume that the group $H_1(V;\mathbb{Z})$ has no torsion, so that it can be considered as a lattice within $H_1(V;\mathbb{R})$ equipped with its stable norm. We will denote by $\langle\ ,\ \rangle$ the duality pairing between $H_1(V;\mathbb{R})$ and $H^1(V;\mathbb{R})$.

Now let E be a vector space of closed 1-forms representing the cohomology $H^1(V;\mathbb{R})$ (in [Lichn], E is the space of harmonic forms, but we want to maintain a purely differential point of view). We can then identify E^* with $H_1(V;\mathbb{R})$. Given $v \in V$, we define a map $\tilde{f}$ from the set C_v of paths in V starting at v into E by

$$\langle \tilde{f}(c), \alpha \rangle = \int_c \alpha$$

for $c \in C_v$ and $\alpha \in E$.

If c, c' have the same endpoint x, then

$$\int_c \alpha - \int_{c'} \alpha = \int_{c-c'} \alpha = \int_\gamma \alpha,$$

where γ is the homology class of $c - c'$. The set of linear forms on E given by $\alpha \mapsto \int_\gamma \alpha$ identifies via biduality with $H_1(V;\mathbb{Z})$.

Thus, if we identify paths in C_v having the same endpoint, the mapping $\tilde{f}$ induces a mapping f of V into the torus $H_1(V;\mathbb{R})/H_1(V;\mathbb{Z})$. *A priori*, the map f depends on the choice of v and E, but a change of the space E

representing $H^1(V;\mathbb{R})$ transforms f into a mapping that is homotopic to f.

The torus $J_1 = H_1(V;\mathbb{R})/H_1(V;\mathbb{Z})$ is called the *Jacobi variety* of V, and the map f (well defined up to homotopy) is the *Jacobi mapping*. If V is equipped with a Riemannian metric, then we equip $\mathbb{T}$ with the left-invariant Finsler metric induced by $\| \cdot_{\mathbb{R}} \|$. If $H_1(V;\mathbb{Z})$ has torsion, then we take the lattice $H(V;\mathbb{Z})/\text{torsion} \subset H_1(V;\mathbb{R})$ instead of $H_1(V;\mathbb{Z})$ and define the torus I_1 as earlier.

The object of the next proposition is to geometrically approximate the Jacobi variety of V, which was presented so far in a rather formal manner.

4.22$_+$. Consider the self-coverings of the Jacobian torus $J_1 = J_1(V)(= \mathbb{T}^d$ for $d = \operatorname{rank} H_1(V))$ given by $t \mapsto mt$, $m = 1, 2, 3, \ldots$, and let $\tilde{V}_m \to V$ be the induced covering of a compact manifold V with $\operatorname{rank}_{\mathbb{R}} H_1(V) = d$, where we think of V as being mapped to $J_1 = H_1(V;\mathbb{R})/H_1(V;\mathbb{Z})$ via the Jacobi mapping. Notice that the fundamental group of this covering $\pi_1(\tilde{V}_m) \subset \pi_1(V)$ is generated by

1. the commutator $[\pi_1(V), \pi_1(V)]$,
2. all m-th powers $\gamma^m \in \pi_1(V)$,
3. all γ which are torsion modulo commutator (i.e., the torsion in $H_1(V)$).

Now we take the induced path metric in the covering $\tilde{V}_m \to V$, scale it by m^{-1}, and let $m \to \infty$.

Proposition: *The metric spaces $m^{-1}\tilde{V}_m$ Hausdorff-converge to the Jacobian (torus) J_1 of V with the metric coming from the $\mathbb{R}$-mass $\| \cdot_{\mathbb{R}} \|$ on its universal covering $\tilde{J} = H_1(V;\mathbb{R})$.*

Proof. The above lemma trivially implies that the space (group) $H_1(V)$ with the metric length$(\alpha - \beta)$ converges, after scaling, to $H_1(V;\mathbb{R})$ in the pointed Hausdorff topology,

$$m^{-1}H_1(V) \xrightarrow{\text{Hau}} H_1(V;\mathbb{R})$$

for $m \to \infty$, where one should keep in mind that our $H_1(V;\mathbb{R})$ with the metric $\|(\alpha - \beta)_{\mathbb{R}}\|$ is scale-invariant. This $H_1(V)$ can be seen geometrically in the abelian covering $\tilde{V}_{\text{Ab}} \to V$ induced by the universal covering $H_1(V;\mathbb{R}) = \tilde{J}_1 \to J_1$. Notice that $\tilde{V}_{\text{Ab}} \to V$ is the maximal Galois covering with a free abelian Galois (i.e, deck transformation) group, which is $\mathbb{Z}^d = H_1(V;\mathbb{Z})/\text{torsion}$. Thus, the $\mathbb{Z}^d$ orbit in $\tilde{V}_{\text{Ab}}$ of the lifted reference point

$v_0 \in V$ identifies with the group $\mathbb{Z}^d = H_1(V;\mathbb{Z})/$ torsion with the length norm (induced from $H_1(V;\mathbb{Z})$ under the factorizing away the torsion).

Next, we observe that $(H_1(V;\mathbb{Z}), \text{length})$ and $(\mathbb{Z}^d, \text{length})$ are both separated by a bounded Hausdorff distance. Since $(\mathbb{Z}^d, \text{length})$ is an orbit of cocompact action in $\tilde{V}_{\text{Ab}}$, it also has finite distance from $\tilde{V}_{\text{Ab}}$. Thus,

$$m^{-1}\tilde{V}_{\text{Ab}} \xrightarrow{\text{Hau}} H_1(V;\mathbb{R}) = \tilde{J}_1,$$

where "Hau" refers to the pointed Hausdorff topology.

Finally, we observe that

$$\tilde{V}_m = \tilde{V}_{\text{Ab}}/m\mathbb{Z}^d,$$

and so $m^{-1}(\tilde{V}_m) = (m^{-1}\tilde{V}_{\text{Ab}})/m\mathbb{Z}^d$ converge to $(m^{-1}\tilde{J}^1)/m\mathbb{Z}^d = \tilde{J}^1/\mathbb{Z}^d$.

4.23$_+$. Notice that the proof provides nontrivial geometric information about every mapping $f\colon V \to J_1$ in the Jacobi homotopy class, or rather about the corresponding covering map $\tilde{f}\colon \tilde{V}_{\text{Ab}} \to \tilde{J}_1 = H_1(V;\mathbb{R})$, where $\tilde{V}_{\text{Ab}}$ is given the path metric $\widetilde{\text{dist}}$ induced from V. Namely, we (obviously) have the following

4.24$_+$ Lemma: *The map $\tilde{f}$ is isometric at infinity,*

$$\widetilde{\text{dist}}(\tilde{v}_1, \tilde{v}_2)/\|(\tilde{f}(v_1) - \tilde{f}(v_2)\| \to 1$$

as $\text{dist}(\tilde{v}_1, \tilde{v}_2) \to \infty$. *Moreover, it admits a (discontinuous) almost inverse map $\varphi : \tilde{J}_1 \to \tilde{V}_{\text{Ab}}$, such that* $\text{dist}(\varphi \circ f - \text{id}) \leq \text{const} < \infty$ *and* $\text{dist}(f \circ \varphi - \text{id}) \leq \text{const} < \infty$. *In other words, f is a quasi-isometry with implied bi-Lipschitz constant equal to one at infinity.*

4.25. Definition: Let W be a manifold equipped with a metric d_i not necessarily induced by a Riemannian structure. We define the *volume* of W as

$$\sup_g \left\{ \text{vol}(W, g) : \begin{array}{l} g \text{ is a Riemannian metric on } W \text{ whose} \\ \text{distance function } \text{dist}_g \text{ is smaller than } d_i \end{array} \right\}.$$

When $W = J_1$, the distance d is translation invariant, and so it suffices to consider the supremum over translation-invariant Riemannian metrics, i.e., those metrics arising from a Euclidean norm on $H_1(V;\mathbb{R})$. The supremum is then attained, since the function $\|\cdot\| \to \text{vol}(J_1, \|\cdot\|)$ is continuous on the space of Euclidean norms on $H_1(V;\mathbb{R})$.

4.26. Remark: The notion of volume defined in 4.25 differs in general from the n-dimensional Hausdorff measure of W, for example, as well as from other reasonable notions of volume (see [Fed]$_{\mathrm{GMT}}$, Ch. 2, sec. 10).

4.27. Theorem: *If V is oriented,* $\operatorname{rank} H_1(V) = n = \dim(V)$, *and the Jacobi mapping $f\colon V \to J_1$ has nonzero degree, then*

$$\operatorname{vol}(V) \geq \operatorname{vol}(J_1). \tag{$*$}$$

Remarks: (a) The hypothesis that $\deg(f) \neq 0$ indicates that the fundamental class of V is the product of n classes of degree 1, a condition that only pertains to the topology of V and not to its metric.

(b_+) In $4.29\frac{1}{2}$, we will prove the optimal inequality

$$\operatorname{vol} V \geq \operatorname{vol}_+ J_1 \tag{$+$}$$

for

$$\operatorname{vol}_+ J_1 = \inf_g \{\operatorname{vol}(J_1, g)\},$$

where the infimum ranges of Riemannian metrics g on J_1 whose distance function is greater than the metric on J_1 corresponding to the $\mathbb{R}$-mass $\|\cdot_{\mathbb{R}}\|$ on $H_1(V;\mathbb{R}) = \tilde{J}_1$.

The proof of Theorem 4.27 is modeled on that of the following lemma (which won't be used but whose interest is worth the pain of bringing it up):

4.28. Generalized Besicovitch Lemma ([Der]). *Let K be a Riemannian cube and denote by (F_i, G_i), $1 \leq i \leq n$, its pairs of opposite faces. If d_i is the distance between F_i and G_i, then*

$$\operatorname{vol}(K) \geq \prod_{i=1}^{n} d_i.$$

Proof. For $x \in K$ we set $f_i(x) = d(x, F_i)$ and $f = (f_1, \ldots, f_n)\colon K \to \mathbb{R}^n$. We will show that $K' \subset f(K)$, where $K' = \prod_{i=1}^n [0, d_i] \subset \mathbb{R}^n$.

By hypothesis, K is a cube, i.e., there exists a homeomorphism $h\colon K \to K'$ such that $h|_{\partial K}\colon \partial K \to K'$ maps each face onto a face. In other words, if y_i is the i-th coordinate in $\mathbb{R}^n$, then $y_i \circ h(F_i) = 0$ and $y_i \circ h(G_i) = d_i$. However, $y_i \circ f(F_i) = 0$ and $y_i \circ f(G_i) = f(G_i) > d_i$, and so by setting $f_t(x) = (1-t)f(x) + th(x)$, we obtain a homotopy from $f|_{\partial K}$ to $h|_{\partial K}$ with values in $\mathbb{R}^n \setminus K'$. If there exists $y \in K' \setminus f(K)$, then we have just shown that $f|_{\partial K}$ is homotopic to $h|_{\partial K}$ in $\mathbb{R}^n \setminus y$, and thus is a generator of $[\partial K, \mathbb{R}^n \setminus y]$. But $f|_{\partial K}$ extends to a map $f\colon K \to \mathbb{R}^n \setminus y$, i.e., $f|_{\partial K}$ is homotopic to 0 in $\mathbb{R}^n \setminus y$, a contradiction.

To complete the proof, we approximate f by smooth maps. It is clear that each function $f_i\colon K \to (\mathbb{R}, \mathrm{can})$ is short; for each $\varepsilon > 0$ there exists a C^∞ mapping $f_{i,\varepsilon}$ such that $d(f_i, f_{i,\varepsilon}) \le \varepsilon/n$ and $\mathrm{dil}(f_{i,\varepsilon}) \le 1+\varepsilon$. Arguing as above, we find that the image of $f_\varepsilon = (f_{1,\varepsilon}, \dots, f_{n,\varepsilon})$ contains the set $K'_\varepsilon = \prod_{i=1}^n (\varepsilon, d_i + \varepsilon)$. If ω_0 is the volume form on $(\mathbb{R}^n, \mathrm{can})$, then $\omega_0 = dy_1 \wedge \cdots \wedge dy_n$, and so $f_\varepsilon^*\omega_0 = \bigwedge_{i=1}^n f_\varepsilon^* dy_i = \bigwedge_{i=1}^n df_{i,\varepsilon}$. From this, we can deduce that $\|f_\varepsilon^*\omega_0\| \le (1+\varepsilon)^n$, and so

$$\begin{aligned}\mathrm{vol}(K) &\ge (1+\varepsilon)^{-n} \int_K \|f^*\omega_0\|\, \nu_g \ge (1+\varepsilon)^{-n} \left| \int_K f^*\omega_0 \right| \\ &\ge (1+\varepsilon)^{-n} \left| \int_{f(K)} \omega_0 \right| \ge (1+\varepsilon)^{-n} \prod_{i=1}^n (d_i - 2\varepsilon)\end{aligned}$$

for each $\varepsilon > 0$.

4.28bis. Remark: There is another version of Besicovitch's lemma that applies to connected manifolds (M, g) whose boundary has two components $\partial M_1, \partial M_2$. If d is the distance $d(\partial M_1, \partial M_2)$ and a is the infimum of the masses (cf. 4.17) of cycles homologous to ∂M_1, then $\mathrm{vol}(g) \ge ad$.

4.29. Proof of Theorem 4.27. Fix a Euclidean norm $\|\cdot\|_e$ on $H_1(V;\mathbb{R})$ that is strictly smaller than $\|\cdot_{\mathbb{R}}\|$, so that $f^0 : V^0 \to (H_1(V;\mathbb{R}), \|\cdot\|_e)$ is "short at infinity" by Lemma 4.24. Choose a parallelepiped K_0 in the Euclidean space $(H_1(V;\mathbb{R}), \|\cdot\|_e)$ and denote by (F_i^0, G_i^0) its pairs of opposite faces, $d_i = d(F_i^0, G_i^0)$ with respect to the Euclidean distance $\|\cdot\|_e$, and K_1 the homothetic kK_0, where k is an integer. Set $K = (f^0)^{-1}(K_1)$, $F_i = (f^0)^{-1}(kF_i^0)$, and $G_i = (f^0)^{-1}(kG_i^0)$. If $\varepsilon > 0$, then for sufficiently large k we have $d(F_i, G_i) \ge (1+\varepsilon)^{-1} k d_i$. For $x \in K$, set $g_i(x) = d(x, F_i)$, $g = (g_1, \dots, g_n)\colon K \to (\mathbb{R}^n, \mathrm{can})$, and

$$K' = \prod_{i=1}^n (0, (1+\varepsilon)^{-1} k d_i) \subset \mathbb{R}^n.$$

Modulo an isometry $E \to \mathbb{R}^n$ sending K_0 onto $(1+\varepsilon)^n K'$, we may assume that f^0 takes values in $\mathbb{R}^n$ and $f^0(\partial K) \cap K' = \emptyset$. As in the preceding lemma, it turns out that that $f^0|_{\partial K}$ is homotopic to $g|_{\partial K}$ in $\mathbb{R}^n \setminus K'$. However, f has nonzero degree, and consequently, f^0 assumes almost every value in K' so that $f^0|_{\partial K}$ is not homotopic to 0 in $\mathbb{R}^n \setminus y$ for $y \in K'$. The same is true for g, so that $K' \subset g(K)$ and, since $g\colon K \to \mathbb{R}^n$ is short,

$$\mathrm{vol}(K) \geq \mathrm{vol}(g(K)) \geq (1+\varepsilon)^{-n} \prod_{i=1}^{n} kd_i = (1+\varepsilon)^{-n}\,\mathrm{vol}(K_1)$$

in E.

Choose a basis of $H_1(V;\mathbb{Z})$. The parallelepiped P (not necessarily a rectangle for $\|\cdot\|_e$) thus determined is a fundamental domain for the action of $H_1(V;\mathbb{Z})$, and so $\mathrm{vol}((\mathbb{T}, \|\cdot\|_e)) = \mathrm{vol}(P)$. If k is sufficiently large, then the parallelepiped $K_1 = kK_0$ contains a number N of translates of P and is contained within a union of $N(1+\varepsilon)$ of translates of P. Since f^0 is bijective on fibers, the set $Q = (f^0)^{-1}(P)$ is a fundamental domain for the action of $\mathrm{Aut}(V)$, and so

$$\mathrm{vol}((f^0)^{-1}(K_1)) \leq N(1+\varepsilon)\,\mathrm{vol}(V)$$

$$\mathrm{vol}(K_1) \geq N\,\mathrm{vol}(P) = N\,\mathrm{vol}((\mathbb{T}, \|\cdot\|_e)),$$

and so finally

$$\begin{aligned} \mathrm{vol}(V) &\geq (N(1+\varepsilon))^{-1}\,\mathrm{vol}(K) \\ &\geq \frac{1}{N}(1+\varepsilon)^{-n-1}\,\mathrm{vol}(K_1) \\ &\geq (1+\varepsilon)^{-n-1}\,\mathrm{vol}(P)), \end{aligned}$$

from which we can conclude that $\mathrm{vol}(V) \geq \mathrm{vol}((\mathbb{T}, \|\cdot\|_e))$ for every Euclidean norm on $H_1(V;\mathbb{R})$ less than $\|\cdot_{\mathbb{R}}\|$, hence $\mathrm{vol}(V) \geq \mathrm{vol}((\mathbb{T}, \|\cdot_{\mathbb{R}}\|))$.

$4.29\frac{1}{2}_+$. Remarks, conjectures, exercises. (a) Observe the following basic topological fact underlying the proof of the Besicovitch lemma.

Surjectivity lemma: *Let φ be a self-mapping of a compact polyhedron P which sends each face of P into itself. Then the map φ is onto.*

In fact, this map is homotopic to the identity via face-preserving maps, and so φ maps each face of P into itself with *degree one*, which implies surjectivity.

(b) There are various versions of Besicovitch's lemma bounding the volume of a polyhedron from below in terms of the distances between the faces. For example, *if Δ^n is a simplex with a riemannnian metric, such that the sum of the $n+1$ distances from each point in Δ^n to the $(n-1)$-faces of Δ^n is $\geq \delta$, then*

$$\mathrm{vol}\,\Delta^n \geq \mu_n(\delta) = \mu_n(1)\delta^n,$$

where $\mu_n(\delta)$ denotes the volume of the regular Euclidean n-simplex

$$\Delta^n_\delta = \{x_i \geq 0 : \sum_{i=0}^{n} x_i = \delta\}$$

of height δ.

Proof. Map Δ^n to $\mathbb{R}^{n+1}_+$ by the $n+1$ distance functions to the $(n-1)$-faces of Δ^n and then radially project to Δ^n_δ. This sends Δ^n onto Δ^n_δ with $|\,\text{Jacobian}\,| \leq 1$. In fact, every map $x\colon V \to \mathbb{R}^2$ with coordinate functions $x_i(v)$ has

$$|\,\text{Jacobian}(v)| \leq \left(n^{-n}\left(\sum_{i=1}^{q} \|dx_i(v)\|^2\right)^n\right)^{1/2},$$

i.e., the differential $D = D_v(x)\colon T_v(V^n) \to \mathbb{R}^q$ satisfies

$$[\det(D^*D)]^{1/2} \leq \left[n^{-n}\,(\operatorname{trace} D^*D)^n\right]^{1/2},$$

by the arithmetic-geometric mean inequality. (See §7.3 in [Gro]$_{\text{FRM}}$ for further examples of a similar kind.)

(*c*) Since the map $\tilde{f}\colon \tilde{V}_{\text{Ab}} \to \tilde{J}_1 = H_1(V;\mathbb{R})$ is isometric at infinity, the inequality $\operatorname{vol} V \geq c \operatorname{vol} J_1$ is equivalent, for every $c > 0$, to the (asymptotic) lower bound on the volumes of large R-balls $\tilde{B}(R)$ in $\tilde{V}_{\text{Ab}}$ by those of the balls $B(R)$ in $\tilde{J}_1 = H_1(V;\mathbb{R})$ with the mass norm, $\liminf_{R\to\infty} \operatorname{vol}\tilde{B}(R)/\operatorname{vol} B(R) \geq c$. This is seen by covering large balls by fundamental domains. On the other hand, since $\tilde{J}_1 = H_1(V;\mathbb{R})$ is a normed space, and every norm $\|\cdot\|$ can be minorized by a Euclidean norm $\|\cdot\|_e$ satisfying $\|\cdot\|_e \geq \|\cdot\|/\sqrt{d}$ for $d = \operatorname{rank} H_1(V)$ by the F. John theorem (see [John] and [Mil–Sch]). This makes $\operatorname{vol} B(R) \geq d^{-d/2}\nu_d R^d$, where ν_d denotes the volume of the unit ball in $\mathbb{R}^d$, and consequently, under the assumptions of 4.27

$$\liminf_{R\to\infty} R^{-n} \operatorname{vol}\tilde{B}(R) \geq n^{-n/2}\nu_n.$$

The unfortunate coefficient $n^{-n/2}$ was removed in a recent paper by Burago and Ivanov (see [Bur–Iv]$_{\text{AV}}$ and [Bur–Iv]$_{\text{RT}}$), who used the full power of John's theorem. This says that

> *there is a (unique) Euclidean norm $\|\cdot\|_e^+$ on a given d-dimensional normed space $(H, \|\cdot\|)$ such that $\|\cdot\|_e^+ \geq \|\cdot\|$ and*

$$(\|\cdot\|_e^+)^2 = \sum_{i=1}^{q} \ell_i^2,$$

where ℓ_i are some linear forms on X depending on $\|\cdot\|$ which satisfy $\sum_{i=1}^q \|\ell_i\|^2 \leq d$.

This theorem of F. John yields the following (minor generalization of) the result by Burago-Ivanov:

(*d*) **Theorem:** *Let X be an n-dimensional Riemannian manifold and let $H = (H, \|\cdot\|)$ be a normed space. Suppose that there exists a proper, asymptotically short map $F\colon X \to H$, i.e., such that*

$$\limsup_{\operatorname{dist}(x,y)\to\infty} \frac{\|F(x)-F(y)\|}{\operatorname{dist}(x,y)} \leq 1.$$

If $\dim H = n$ and F has nonzero degree mod 2, then F is asymptotically volume-increasing for the volume (element) vol_+ in H associated to some Euclidean norm $\|\cdot\|_e^+$ on H majorizing $\|\cdot\|$,

$$\liminf_{R\to\infty} \frac{\operatorname{vol} F^{-1}(B(R))}{\operatorname{vol}_+ B(R)} \geq 1.$$

Furthermore, if V is oriented, then

$$\liminf_{R\to\infty} \frac{\operatorname{vol} F^{-1}(B(R))}{\operatorname{vol}_+ B(R)} \geq |\deg(F)|,$$

where $B(R)$ are the $\|\cdot\|$-balls in H,

$$B(R) = \{h \in H : \|h\| \leq R\}.$$

Proof. We "transplant" the linear function ℓ_i from H to X as follows. First, we compose each ℓ_i with F and restrict the resulting functions on X to some ρ-separated ρ-net $X_\rho \subset X$. If ρ is large, then these functions are λ_i-Lipschitz on X_ρ for $\lambda_i \approx \|\ell_i\|$, i.e., they are $(\|\ell_i\| + \varepsilon(\rho))$-Lipschitz with $\varepsilon(\rho) \to 0$ as $\rho \to \infty$. Then the functions $\ell_i \circ F$ extend from X_ρ to λ_i-Lipschitz functions $\tilde{\ell}_{i,\rho}$ on X, which clearly satisfy

$$\sup_{x\in X} |\ell_{i,\rho}(x) - \ell_i \circ F(x)| \leq C\rho$$

for $C = \max\|\ell_i\|$. Thus, the map $\tilde{L}_\rho = (\tilde{\ell}_{1,\rho}, \tilde{\ell}_{2,\rho}, \dots, \tilde{\ell}_{q,\rho})\colon X \to \mathbb{R}^q$ lies within a bounded distance from $L \circ F\colon X \to \mathbb{R}^q$ for $L = (\ell_1, \dots, \ell_q)$ and so the filling volume of $\partial F^{-1}(B(R)) = F^{-1}(\partial B(R))$ mapped to $\mathbb{R}^q$ is roughly the same as that for $\partial B(R)$ mapped to $\mathbb{R}^q$ by L, i.e., the difference between the two is $o(R^n)$ as $R \to \infty$. (This can be seen from the perspective of §7.3

in $[\mathrm{Gro}]_{\mathrm{FRM}}$ or by just normally projecting $\mathbb{R}^q$ onto the image $L(H) \subset \mathbb{R}^q$.) On the other hand, the Jacobian of $\tilde{L}_\rho$ is bounded by

$$\left(n^{-n}\left(\sum_{i=1}^{q}\lambda_i^2\right)^n\right)^{1/2} = \left(n^{-n}\left(\sum_{i=1}^{q}\|\ell_i\|^2\right)^n\right)^2 + o(1) \leq 1 + o(1)$$

for $\rho \to \infty$. It follows that the volume of $F^{-1}(B(R))$ is "essentially" greater than the filling volume of $\partial B(R) \subset L(H) \subset \mathbb{R}^N$, which obviously equals $\mathrm{vol}_+(B(R))$ (where one should not take "filling" too seriously, since the action can be moved to $L(H)$ via the normal projection $\mathbb{R}^q \to L(H)$). Finally, we erase "essentially" by letting $R \to \infty$ and then $\rho \to \infty$.

(d_1) **Corollary (see [Bur–Iv]):** *If the map F is isometric at infinity, e.g., $X = \tilde{V}_{\mathrm{Ab}}$ and F equals the cover $\tilde{f}$ of a Jacobian map f from $V = X/\mathbb{Z}^n$ to a $J_1 = H_1(V;\mathbb{R})/\mathbb{Z}^n$, $\mathbb{Z}^n = H_1(V;\mathbb{R})/\mathrm{tor}$, then the balls $\tilde{B}(R)$ about a fixed point x_0 in X satisfy*

$$\liminf_{R\to\infty} R^{-n}\,\mathrm{vol}\,\tilde{B}(R) \geq \nu_n|\deg(f)| \tag{$*$}$$

(where ν_n denotes the volume of the unit Euclidean n-ball).

(d_2) **Rigidity.** It is clear from the above that equality in $(*)$ implies that the norm $\|\cdot\|$ in H equals the Euclidean (John's) norm $\|\cdot\|_e^+$, i.e., $(H, \|\cdot\|)$ is Euclidean to start with. Furthermore, the composition $\overline{F}_\rho$ of $\tilde{F}_\rho$ with the normal projection $\mathbb{R}^q \to \mathbb{R}^n = L(H)$ is *almost isometric on average.* This means that among all pairs of points $\{x, y \in B(R) : \mathrm{dist}(x,y) \leq 1\}$, those which satisfy

$$|\,\mathrm{dist}(x,y) - \|\overline{F}_\rho(x) - \overline{F}_\rho(y)\| \geq \varepsilon > 0$$

have the *relative* measure $\leq \delta_\rho$ in $B(R)$ for all sufficiently large R, where $\delta_\rho \to 0$ for every fixed $\varepsilon > 0$ and $\rho \to \infty$. It then trivially follows (compare Ch. 2.E$_+$) that there exists a sequence of points $x_i \in X$ such that (X, x_i) Hausdorff-converge to $\mathbb{R}^n$ with the maps $\overline{F}_\rho$ converging to isometries along with $\rho \to \infty$. In particular, *if the isometry group is cocompact on X, e.g., $X = \tilde{V}_{\mathrm{Ab}}$ for a compact V, then $X = \mathbb{R}^n$ (see [Bur–Iv]).* (This obviously remains true if X comes along as a *minimal leaf* of a compact foliation.)

(d_3) **Technical remarks.** (i) When F is isometric at infinity, one can produce the $\tilde{L}$-companion of L more constructively, in the spirit of the argument in 4.29 (see also [Bur–Iv]). Namely, for each $i = 1, \dots, q$, one could take the pullback $y_i = (\ell_i \circ F)^{-1}(0) \subset X$ and set

$$\tilde{\ell}_i = \pm\|\ell_i\|\,\mathrm{dist}(x, y_i)$$

with the $\pm$ sign determined by $\operatorname{sign}(\ell_i \circ F(x))$.

(*ii*) When X is isometrically acted upon by $\mathbb{Z}^n$ (e.g., $X = \tilde{V}_{\mathrm{Ab}}$), then one can average the differentials $d\tilde{\ell}_i$ over this action, thus getting measurable closed 1-forms $\tilde{\delta}_i$ on $X/\mathbb{Z}^n$ cohomologous to the corresponding forms δ_i on $H/\mathbb{Z}^n$ coming from ℓ_i. These forms allow an alternative proof of the rigidity result, since they provide an isometry $X/\mathbb{Z}^n \to H/\mathbb{Z}^n$ sending each δ_i to $\tilde{\delta}_i$ (an exercise for the reader). One may wonder if these $\tilde{\delta}_i$ are good for something else.

(*e*) **About** (+). The inequality

$$\operatorname{vol} V \geq \operatorname{vol}_+ J_1 \tag{+}$$

now follows from the Burago–Ivanov theorem by 4.24_+. Furthermore, if the Jacobian map $f\colon V \to J_1$ has degree > 1, then

$$\operatorname{vol} V \geq c_n |\deg f| \operatorname{vol}_+ J_1$$

for some universal constant $c_n > 1$. This follows by confronting our present arguments with those in Ch. $2.\mathrm{E}_+$. (Notice that $\deg f$ equals the minimal integer > 0 such that the cup product of arbitrary classes $h_i \in H_1(V;\mathbb{Z})$, $i = 1, \ldots, n$, is divisible by d without being zero.)

(*f*) **Exercises.** (*i*) Let X_j, $j = 1, 2, \ldots$ be compact n-dimensional manifolds which Hausdorff-converge to a compact Finsler manifold Y, such that all X_j are uniformly (in j) locally contractible. Show that

$$\liminf_{j\to\infty} \operatorname{vol} X_j \geq \operatorname{vol}_+ Y,$$

where equality implies that Y is Riemannian (compare [Bur-Iv]).

(*ii*) Consider a proper asymptotically Lipschitz map $F : X \to H$ where $n = \dim X > \dim H = d$ and let $\operatorname{systdeg}(f)$ (systolic degree) denote the infimum of $(d-n)$-volumes of cycles in X homologous to the pullback of a regular value of F (assuming F is smooth around some point in H). Prove that

$$\liminf_{R\to\infty} \frac{\operatorname{vol}_n F^{-1}(B(R))}{\operatorname{vol}_+ B(R)} \geq \operatorname{systdeg}(f)$$

and study the case of equality.

(*g*) Let X be a complete Riemannian manifold and define $\operatorname{supVol}(R)$ as the supremum of volumes of the R-balls in X. One dreams of a universal lower bound $\operatorname{supVol}(R) \geq \operatorname{const} R^n$ (especially for const $= \nu_n$) for all R

(especially for $R \to 0$ and $R \to \infty$) under suitable geometric and topological assumptions on X, such as uniform contractibility (see $[\mathrm{Gro}]_{\mathrm{FRM}}$ and 4.44 below for some evidence in favor of such conjectures). Here one knows that uniformly contractible surfaces satisfy $\sup \mathrm{vol}(R) \geq 3R^2$ for all $R \geq 0$ (an exercise for the reader, also see 5.2.A in $[\mathrm{Gro}]_{\mathrm{FRM}}$) and one (naively?) expects $\sup \mathrm{vol}\, R \geq \pi R^2$ in this case. Particular X's where one wants to know $\sup \mathrm{vol}\, B(R)$, besides universal coverings of compact aspherical manifolds, are *contractible minimal* leaves of compact foliations, e.g., stable and unstable Anosov leaves, such as the horospheres of (compact) manifolds with negative curvature. Do the latter have (at least) Euclidean volume growth?

4.30_+ Theorem: *Let V be a closed Riemannian n-manifold which admits n 1 dimensional cohomology classes in $H^1(V;\mathbb{R})$ with nonzero cup-product. If* $\mathrm{rank}\, H_1(V;\mathbb{R}) = n = \dim V$*, then there is a nontorsion class $\alpha \in H_1(V;\mathbb{Z})$ such that*

$$\|\alpha_{\mathbb{R}}\| \leq \nu^{-1/n} 2(\mathrm{vol}(V))^{1/n}, \qquad (++)$$

where $\nu_n \approx n^{-n/2}$ denotes the volume of the unit Euclidean ball.

Proof. First let V be a flat Riemannian torus. Then $(++)$ holds for some nontrivial $\alpha \in H_1(V)$ by the *Minkowski theorem.* In fact, the required short loop comes along with the minimal simply-connected (Euclidean!) ball $B(R) \subset V(\simeq T^n)$ which has

$$\nu_n R^n = \mathrm{vol}\, B(R) \leq \mathrm{vol}\, V$$

and thus delivers a loop of length $2R$. Now, we invoke the inequality $(+)$ above, which says that the volume $\mathrm{vol}\, V$ majorizes that of the Jacobian with the Riemannian metric coming from $\|\cdot\|_e^+$ on $\tilde{J} = H_1(V;\mathbb{R})$, where we recall that $\|\cdot\|_e^+ \geq \mathbb{R}$-mass on $H_1(V;\mathbb{R})$. Thus, the general case follows from the Minkowski theorem above applied to

$$J_1 = (H_1(V;\mathbb{R}), \|\cdot\|_e^+)/H_1(V;\mathbb{Z}).$$

4.31_+ Remarks. (a) We had a weaker inequality with an extra $\sqrt{n}$ factor in the French edition of this book, since Burago-Ivanov's refinement of $(*)$ was only conjectural at that time.

(b) Our argument actually shows that

$$\|\alpha_{\mathbb{R}}\| \leq \sigma_n (\mathrm{vol}\, V)^{1/n},$$

where the number $\sigma_n < 2\nu_n^{-1/n}$ equals the length of the minimal geodesic γ in the "worst" flat torus of unit volume, where length(γ) is the largest possible. In fact, everything about the $\mathbb{R}$-mass on $H_1(V;\mathbb{Z})$ can be bounded (in terms of $\operatorname{vol} V$) by what happens to flat tori.

(c) The discussion above generalizes Loewner's theorem (see 4.1 above) since the stabilized norm $\| \cdot_{\mathbb{R}} \| = \| \cdot \|^{\lim}$ on $H_1(V;\mathbb{Z})$ equals the original length according to the following.

4.32. Lemma: *If a curve α in an orientable surface S is length minimizing in its homotopy class, then its iterates α^p are length minimizing in their respective homotopy classes.*

Proof. Indeed, the lift $\tilde{\alpha}$ of α to the universal cover $\tilde{S}$ of the surface is an infinite geodesic that divides $\tilde{S}$ into two domains. A loop β based on α and having minimal length in some multiple of the class of α would have a lift β, which, unable to intersect $\tilde{\alpha}$ without being nonminimizing, would lie on one side of $\tilde{\alpha}$. In this case, however, $\tilde{\beta}$ would have at least one self-intersection, which would permit us to shorten it within its homotopy class. Thus, $\alpha = \beta$.

4.33_+ Additional remarks and exercises. (a) Our reasoning can be adjusted to the more general situation where $\operatorname{rank} H_1(V) \geq n = \dim V$, but this brings an extra constant depending on $\operatorname{rank} H_1$ (see §7.4, 7.5 in $[\mathrm{Gro}]_{\mathrm{FRM}}$). On the other hand, we shall later see from a different angle that one can always find a short curve which is nonhomologous to zero in V and for which the implied constant depends only on $n = \dim V$. Probably this could also be achieved with a refinement of our argument.

(b) Suppose that $d = \operatorname{rank} H_1(V) < n = \dim V$, define systdeg($f$) of the Jacobian map $f : V \to J_1 = \mathbb{T}^d$ as above, and prove the following bound on the mass of the minimal $\alpha \in H_1(V;\mathbb{Z})$ generalizing (++),

$$\|\alpha_{\mathbb{R}}\|^{n-d} \operatorname{systdeg}(f) \leq \nu_n^{-1} 2^n \operatorname{vol} V.$$

(c) Extend the logic of the exercise (b) to the Jacobian mapping of the Heisenberg manifold H_{2n+1}/Γ to $\mathbb{T}^{2n}$ with a homotopy version of systdeg (compare 7.4.C′ in $[\mathrm{Gro}]_{\mathrm{FRM}}$).

(d) **Question**. Consider H_{2n+1} with its canonical contact structure and look at the Carnot-Caratheodory metrics on H_{2n+1} compatible with this structure. Are there reasonably general conditions on such metrics which would ensure the lower asymptotic bound on the CC-volumes of the R balls in such H_{2n+1} by $\mathrm{const}_n\, R^{2n+2}$? In particular, we want such bounds for (reasonable?) CC-metrics invariant under a cocompact nilpotent isometry group Γ acting on H_{2n+1}.

D. An application of geometric integration theory

4.34. A dual norm on cohomology. If we represent the cohomology of a manifold V by differential forms ω, then by proceeding as in 4.15, we obtain a norm on $H^p(V;\mathbb{R})$ by setting

$$\|\eta\| = \inf\{M^*(\omega) : \omega \in \eta\}$$

for each $\eta \in H^p(V;\mathbb{R})$. Indeed, $d\Omega^{p-1}$ is a closed subspace of Ω^p, as follows from the fact that exact forms are those having zero integral over any cycle.

Even though the duality of the norms M and M^* takes place on the level of chains and forms, and not on the level of cycles and closed forms, the quotient norms on $H_p(V;\mathbb{R}), H^p(V;\mathbb{R})$ are still dual to one another. In other words,

4.35. Proposition: *For each class $\alpha \in H^p(V;\mathbb{R})$, we have*

$$\|\alpha\| = \sup\{|\omega(\alpha)| : \|\omega\| \le 1\}.$$

Proof. (See also $[\mathrm{Fed}]_{\mathrm{VP}}$, 4.10). The fact that $\sup\{|\omega(\alpha) : \|\omega\| \le 1\}$ follows from the inequality $|\omega(\alpha)| \le \|\omega\|\|\alpha\|$ that is itself a consequence of the same inequality for forms η representing ω and cycles c representing α.

To prove the reverse inequality, we consider a cycle c_0 representing α and a linear form L on $B_p \oplus \mathbb{R}c_0$ (here, B_p denotes the space of p-dimensional boundaries) defined by $L(B_p) = 0$ and $L(c_0) = \|\alpha\|$. Since B_p is M-closed in the space of chains (as follows from the fact that a chain c is a boundary precisely when $\int_c \eta = 0$ for each C^∞ closed form η), L is continuous on $B_p \oplus \mathbb{R}c_0$ and has norm less than 1. By the Hahn–Banach theorem, there exists an extension of L (which we will also denote by L) having norm ≤ 1 to the space of chains.

Thus, L is a M-continuous, co-closed co-chain (since it vanishes on B_p), and so by [Whit], p. 157, 261, it can be represented by a measurable, bounded differential form η, and by 4.16 and 4.17 we have $\|L\| = M^*(\omega)$. If ω is the class of η, defined by smoothing η, we have

$$\omega(\alpha) = \|\alpha\| \qquad \text{and} \qquad \|\eta\| \le 1.$$

From this duality we can deduce "stable" isoperimetric inequalities of the form of those in Section 4.C, providing a partial answer to Question 4.13.

Theorem 4.36 *For every Riemannian metric g on $\mathbb{C}P^n$, a generator α of $H_2(\mathbb{C}P^n;\mathbb{Z})$ satisfies the inequality*

$$\|\alpha_{\mathbb{R}}\|^n \leq n!\ \mathrm{vol}(\mathbb{C}P^n, g).$$

Proof. Let ω be a generator of $H_2(\mathbb{C}P^n;\mathbb{Z})$. Then ω^n is a generator of $H_{2n}(\mathbb{C}P^n;\mathbb{Z})$, and for each 2-form η representing ω, we have

$$1 = \int_{\mathbb{C}P^n} \eta^n \leq n! M^*(\eta)^n \,\mathrm{vol}(\mathbb{C}P^n, g),$$

from which we conclude that $1 \leq n!\|\omega\|(\mathbb{C}P^n, g)$. By 4.34, however (here we use the duality between the 1-dimensional spaces $H_2(\mathbb{C}P^n;\mathbb{Z}), H^2(\mathbb{C}P^n;\mathbb{Z})$), $\|\alpha\| = 1/\|\omega\|$.

4.37. Remarks: We have the same inequalities for $\mathbb{H}P^n$ and $\mathbf{Ca}P^n$ (cf. [Berger]$_{\mathrm{Pu}}$); the preceding proof uses only the fact that one has a manifold V of dimension an with $H^a(V;\mathbb{Z}) = \mathbb{Z}$, such that the n-th power of a generator of $H^a(V;\mathbb{Z})$ is a generator of $H^{an}(V;\mathbb{Z})$.

The case of $\mathbb{C}P^n$ is particularly nice, thanks to the reduction of 2-forms and the resulting Wirtinger inequality (cf. [Fed]$_{\mathrm{GMT}}$, p. 40). The inequality of Theorem 4.36 includes $n!$ since the n-th power of the form is taken and not $(2n)!/2^n$, which is the best constant for the comass of the product of n arbitrary 2-forms (cf. [Fed]$_{\mathrm{GMT}}$, Ch. 1, or [Whit], Ch. 1).

For the same reason, this inequality is optimal: If $\mathbb{C}P^n$ is equipped with a Kähler metric, then the norms $\|\cdot\|$ and $\|\cdot_{\mathbb{R}}\|$ on $H^2(\mathbb{C}P^n;\mathbb{Z})$ coincide and the geometric invariants $m(g)$ and (cf. 4.13) $\nu(g)$ are those of the standard metric.

Of course, on $\mathbb{R}P^n$, where the homology is completely torsion, the real homology is zero and this technique produces nothing. On the other hand, it is particularly well adapted to product spaces.

4.38. Theorem: *There exists a constant $C(m,n) > 0$ such that each metric g on $S^m \times S^n$ ($m \neq n$) satisfies the following. For classes α, β from the factor spaces,*

$$\mathrm{vol}(g) \geq C(m,n)\, \|\alpha_{\mathbb{R}}\|\, \|\beta_{\mathbb{R}}\|.$$

For each k, n, there exists a constant $D(k,n) > 0$ such that each metric g on $V = S^k \times \cdots \times S^k$ (n factors) satisfies $\mathrm{vol}(g) \geq D(k,n)A$, *where we set*

$$A = \inf\{\|\alpha_{\mathbb{R}}\| : \alpha \in H(V;\mathbb{Z}) \setminus 0\}.$$

Remark: For $m = 1$, we recover Theorem 4.30 when $V = \mathbb{T}^n$.

Proof. The first part of the assertion is proved as in the proof of Theorem 4.36. The second proceeds as follows. The real vector space $H_m(V;\mathbb{R}) = E$ has dimension n and contains $H_m(V;\mathbb{Z})$ as a lattice. We equip E with the norm $\|\cdot\|$ introduced in 4.17 and denote by K the convex, symmetric set that defines this norm, i.e., $K = \{x \in E : \|x\| \leq 1\}$. We equip $E^* = H^m(V,\cdot)$ with the norm $\|\cdot\|^*$ introduced in 4.15 and set $K^* = \{f \in E^* : \|f\|^* \leq 1\}$, $\Lambda^* = H^m(V,\mathbb{Z})$. By definition, Λ^* is the dual lattice of Λ, i.e.,

$$\Lambda^* = \{f \in E^* : (f(x)) \in \mathbb{Z} \text{ for all } x \in \Lambda \}.$$

The duality described in Proposition 4.35 for the norms $\|\cdot\|$ and $\|\cdot\|^*$ implies that the convex sets K, K^* are *reciprocal* (or *polar*).

4.39. Lemma: *For $A = \inf\{\|x\| : x \in \Lambda \setminus 0\}$, there always exists a basis $(t_i)_{i=1,\dots,n}$ of Λ^* such that*

$$\prod_{i=1}^{n} \|t_i\|^* A^n \leq (n!)^2 \left(\frac{3}{2}\right)^{\frac{1}{2}(n-1)(n-2)}.$$

Proof. The proof consists of the combination of three classical results of geometric number theory to conclude that

1. $\mathrm{vol}(K)\,\mathrm{vol}(K^*) \geq (n!)^{-2}4^n$,
2. There exists a basis $(t_i)_{i=1,\dots,n}$ of Λ^* such that
$$\prod_{i=1}^{n} \|t_i\|^* \,\mathrm{vol}(K^*) \leq 2^n \left(\frac{3}{2}\right)^{\frac{1}{2}(n-1)(n-2)}.$$
3. $\mathrm{vol}(K)A^n \leq 2^n$.

Since E, E^* are arbitrary, we can set $E = E^* = \mathbb{R}^n$ and $\Lambda = \Lambda^n = \mathbb{Z}^n$. In [Lekerk], the proof of (1) appears on p. 106 in 14.2, (2) on p. 59 of 10.2, and finally (3) on p. 33 of 5.1. Note that the volumes are taken with respect to Lebesgue measure and therefore disappear when we take the product of the three inequalities!

End of the proof of Theorem 4.38. Choose exterior differential n-forms ω_i such that $\|\omega_i\| = \|t_i\|^*$ for $i = 1,\dots,n$, so that, as in Theorem 4.36, we have

$$1 = \int_V \omega_1 \wedge \cdots \wedge \omega_n \leq \prod_{i=1}^{n} \|\omega_i\| \, \nu(g).$$

See [Gro]$_{\mathrm{FRM}}$ and [Heb] for further results in this direction.

Question: What are the best constants and extremal manifolds for all these inequalities?

E_+ Unstable systolic inequalities and filling

4.40. The notion of a *systole* $\mathrm{syst}_k(V)$ suggested by M. Berger refers to the minimal volume of a k-cycle *nonhomologous to zero* in a Riemannian manifold V, and for $k = 1$ one may also speak of *homotopy k-systoles* corresponding to minimal *noncontractible* curves in V. A general bound on the latter is known for all closed *aspherical* manifolds V, i.e., those having *contractible* universal coverings $\tilde{V}$. This includes, for example, manifolds homeomorphic to the torus $\mathbb{T}^n$ or those admitting metrics of nonpositive sectional curvatures. Moreover, such a bound on syst_1 remains valid for all *essential* n-dimensional manifolds and polyhedra V, where a polyhedron V is called *n-essential* if there exists an aspherical (possibly infinite) polyhedron K, and a continuous map $V \to K$ which does not contract to the $(n-1)$-skeleton of K.

4.41. Theorem: *Every compact n-dimensional n-essential polyhedron with a piecewise Riemannian metric admits a noncontractible curve S_1 with*

$$\mathrm{length}\, S_1 \leq C_n (\mathrm{vol})^{1/n}$$

for some universal constant $C_n > 0$.

See Appendix 2 in [Gro]$_{\mathrm{FRM}}$ for a proof and [Gro]$_{\mathrm{SII}}$ for an introductory survey of the subject.

Remarks: (a) A basic case of the theorem concerns a closed manifold V for which the classifying map $V \to K(\Pi; 1)$ for $\Pi = \pi_1(V)$ and $K(\Pi; 1)$ being the (aspherical!) Cartan-Eilenberg space, is nonhomologous to zero. For example, the real projective space $\mathbb{R}\mathrm{P}^n$ is essential since it is nonhomologous to zero in $K(\mathbb{Z}_2; 1) = \mathbb{R}\mathrm{P}^\infty$. Thus, the theorem above generalizes Pu's result, albeit *not* with a sharp constant.

(b) One does not truly need the "piecewise Riemannian" condition. The conclusion of the theorem remains true for an arbitrary metric on V with the n-dimensional Hausdorff measure instead of the volume. (This needs a minor adjustment of the argument in [Gro]$_{\mathrm{FRM}}$.)

4.42. To get a grasp on 4.41, let us start with an aspherical space K which is given a metric with the homotopy systole $\geq \sigma > 0$. This means that every curve in K of length $< \sigma$ is contractible. This also can be seen in the universal covering $\tilde{K}$ of K: for every point $\tilde{k} \in \tilde{K}$ and every *nontrivial* deck transformation $\gamma \in \pi_1(K)$ acting on $\tilde{K}$, one has

$$\operatorname{pathdist}(\tilde{k}, \gamma(\tilde{k})) \geq \sigma.$$

Notice that we do not even assume that our metric on K is a path metric. Nevertheless, "pathdist" makes sense as the infimum of lengths of curves in $\tilde{K}$ (between $\tilde{k}$ and $\gamma(\tilde{k})$) where this length is measured in the metric on K after projecting the curves to K. So, we do not exclude the possibility pathdist $= \infty$.

By appealing to our theorem, we then claim that every n-dimensional cycle V in K nonhomologous to zero, satisfies

$$\operatorname{vol}(V) \geq \varepsilon_n \sigma^n \tag{$\geq$}$$

for some universal constant $\varepsilon_n > 0$ (i.e., $\varepsilon_n = c_n^{-n}$ for the c_n above). Here, V is a *singular cycle*, i.e., an integer combination of singular simplices which are continuous maps $\varphi_i : \Delta^n \to K$. The volume of such $V = \sum_i k_i \varphi_i$ is defined to be $\leq t$ if there exist piecewise Riemannian metrics g_i on Δ^n such that all φ_i are 1-Lipschitz on (Δ_n, g_i) and $\sum_i |k_i| \operatorname{vol}(\Delta^n, g_i) \leq t$. And $\operatorname{vol} V \geq t$ means that such g_i *do not* exist.

The idea for proving the lower bound on $\operatorname{vol} V$ comes from geometric measure theory, where one knows that *minimal* n-dimensional subvarieties tend to have large(!) volumes. For example, if V is a complete, minimal subvariety in $\mathbb{R}^N$, then the R-balls in V around some point $v \in V$ grow *faster* with R than with the balls in the flat $\mathbb{R}^n$. And one knows that this property of V (called "volume monotonicity") depends on the classical *filling* (also called *isoperimetric*) *inequality* in $\mathbb{R}^N$,

every $(n-1)$*-cycle* S *in* $\mathbb{R}^N$ *bounds an* n*-chain* B *with*

$$\operatorname{vol}_n(B) \leq \operatorname{const}_n(\operatorname{vol}_{n-1} S)^{n/(n-1)}. \tag{$*$}$$

And a similar inequality in K along with a suitable existence theorem of *minimal* cycles V in a given homology class would imply ($\geq$) by a rather simple argument.

Actually, the filling idea was already used in 4.6_+, where we were cutting away long fingers from a surface V and then filling in by spherical caps, which played the role of the B above in the sense of ($\geq$). That can be generalized from surfaces to all dimensions with the following

Filling inequality: *Let S be an (oriented) $(n-1)$-dimensional pseudomanifold (i.e., a possibly singular variety where the singularity has codimension ≥ 2 so that S makes an abstract cycle) with a piecewise smooth Riemannian metric. Then S can be filled in by an (oriented) pseudomanifold B of dimension n with boundary, i.e., the boundary ∂B is identified with S, and then B can be given a piecewise Riemannian metric with the following three properties:*

1. *The distance function restricted from B to S equals the distance function of the original metric on S.*
2. *The volume of B is bounded by*
$$\mathrm{vol}_n(B) \leq c_n(\mathrm{vol}_{n-1} S)^{n/(n-1)}$$
for some constant in the interval $0 < c_n < n^n\sqrt{n!}$.
3. *The distance from each point $b \in B$ to $\partial B = S$ satisfies*
$$\mathrm{dist}(b, S) \leq c'_n(\mathrm{vol}_{n-1} S)^{1/(n-1)}$$
for $0 < c'_n < n^{n+1}\sqrt{n!}$.

This is proved in $[\mathrm{Gro}]_{\mathrm{FRM}}$. Here, one should note that once we are granted B satisfying (1) and (2), it can be easily improved (by cutting away long fingers) in order to also satisfy (3). On the other hand, (3) alone leads to a bound on the 1-systole of an aspherical manifold V. Namely, we fill-in V instead of S by some $(n+1)$-dimensional pseudomanifold W (instead of B) satisfying (1) and (2). Then, assuming that all short curves in V are contractible, we could construct by a rather trivial topological argument a retraction of W to $V = \partial W \subset W$. But this clearly is impossible, and so short, noncontractible curves in V must exist. (We suggest that the reader try this by him/herself, or consult $[\mathrm{Gro}]_{\mathrm{FRM}}$ for details.)

Remark: We had to introduce pseudomanifolds to bypass the cobordism problem for manifolds. Alternatively, we could work with a filling W homeomorphic to $V \times [0, \infty)$ with some *complete* metric satisfying (1) and (2), but certainly *not* (3).

4.43. Filling volume and filling radius. The filling inequality concerns two basic metric invariants of a Riemannian manifold V.

I. **Filling Volume.** This is the infimal t such that V can be filled by W (with $\partial W = V$ satisfying (1), i.e., having $\mathrm{dist}_W|_V = \mathrm{dist}_V$ and $\mathrm{vol}_{n+1} W = t$).

II. **Filling Radius.** Defined as the infimal R such that V can be filled by W satisfying (1) and having $\mathrm{dist}(w, V) \leq R$ for all $w \in W$.

Exercise. Show that this definition of FilRad agrees with the one given in 3.35 via the Kuratowski embedding of V into the space of functions on V.

4.44. Filling noncompact manifolds and uniform contractibility. The notions of FilVol and FilRad generalize to complete, noncompact Riemannian manifolds, and the noncompact filling inequalities follow easily from the compact one above. An amusing application concerns *uniformly contractible* Riemannian manifolds V (also called *geometrically contractible* in [Gro]$_{\mathrm{FRM}}$). This means that there exists a *contractibility function* $R(r) = R_V(r)$ such that every r-ball in V can be contracted within a concentric $R(r)$-ball (compare Ch. 3.E$_+$).

Theorem: *Every complete, uniformly contractible manifold V has infinite volume.*

This can be easily reduced to a bound on the (noncompact) filling radius of V issuing from the (compact) bound stated in 4.42. We leave this to the reader, who is referred to [Gro]$_{\mathrm{FRM}}$, [Katz]$_{RFR}$, and the papers by M. Katz cited in [Gro]$_{\mathrm{SII}}$ for a further study of FilRad and FilVol. Here we indicate several unsolved

Problems. (a) What is $\mathrm{FilVol}\, S^n$? One suspects that the minimal filling is achieved with the hemisphere $S^{n+1}_+ \supset \partial S^{n+1}_+ = S^n$, but this is unknown even for $n = 1$. Notice that the difficulty in the case of S^1 is due to the possibly nontrivial topology of a surface W with $\partial W = S^1$. Such a W of positive genus (but not for genus $W = 0$) may have $\mathrm{Area}\, W < \mathrm{Area}\, S^2_+ = 2\pi$, where we insist that $\mathrm{dist}_W |_{S^1} = \mathrm{dist}_{S^1}$.

(b) Does the largest ball of radius R in a *uniformly contractible* manifold V have $\mathrm{vol}\, B(R) \geq \mathrm{const}_n\, R^n$ for $n = \dim V$? (Compare 4.29$\frac{1}{2}$ and [Gro]$_{\mathrm{FRM}}$).

(c) Does every closed manifold V with *positive scalar curvature* $Sc(V) \geq \delta > 0$ have

$$\mathrm{FilRad}(V) \leq \mathrm{const}_n\, \delta^{-1/2}?$$

(d) What are the true constants in the general systolic and filling inequalities, and what are the extremal manifolds? Does, for example, the standard $\mathbb{R}\mathrm{P}^n$ have minimal volume per given systole? (One may ask the same question for flat tori, but it is harder to believe that any of them is systolically extremal for $n \geq 3$.)

Remarks: Most techniques employed for bounding syst_1 and related invariants do not discriminate between Riemannian and Finslerian metrics, since they rely only on the Banach geometry of L_∞ spaces (see [Gro]$_{\mathrm{FRM}}$). But then there is little chance to arrive at extremal objects and best constants in the Riemannian category. One should try instead some L_2-techniques better adapted to the Riemannian needs (compare [Bur–Iv] and [B–C–G] for successful proofs of *sharp* volume inequalities).

4.45. On higher dimensional systoles. There are easy examples of n-manifolds V where $\mathrm{syst}_1\,\mathrm{syst}_{n-1}$ is much bigger than $\mathrm{vol}(V)$. For instance, if a simply-connected manifold V_0 admits a free isometric action of S^1, then $V = (V_0 \times \mathbb{R})/\mathbb{Z}$ for a suitable free cyclic subgroup, i.e., copy of $\mathbb{Z}$ in $S^1 \times \mathbb{R} \subset \mathrm{Isom}(V_0 \times \mathbb{R})$, may have an arbitrarily large ratio $\mathrm{syst}_1\, syst_{n-1} / \mathrm{vol}$. Just rotate V_0 by a small angle $\alpha = 2\pi/i$, translate $\mathbb{R}$ by $\varepsilon \simeq \alpha^2$, and observe that

1. the shortest noncontractible curve in V has

$$\text{length} = \mathrm{syst}_1 \geq \alpha$$

2. the minimal $(n-1)$-cycle has

$$\mathrm{vol} = \mathrm{syst}_{n-1} = \mathrm{vol}\, V_0.$$

Since $\mathrm{vol}\, V$ obviously equals $\varepsilon \,\mathrm{vol}\, V_0$, we get

$$\mathrm{syst}_1\, \mathrm{syst}_{n-1} \,/\, \mathrm{vol} \simeq \alpha/\varepsilon = i \to \infty.$$

One can play further with such examples, dressing the essential geometric core with various topological garments via a suitable surgery. Probably one can produce in this way a metric with given systoles in dimensions ≥ 2 on every topological manifold or polyhedron (see [Berger]$_{\mathrm{Sys}}$ and Appendix D_+). But one's curiosity may not yet be satisfied, since these represent only a tiny slice of possible geometries, and one starts wondering about how typical the examples above actually are. And if one tries something different, one finds that it is unexpectedly hard to evaluate systoles of even the simplest manifolds, such as compact homogeneous spaces.

Test question. Determine the possible range of the systolic vector

$$s = s(g) = (s_1, s_2, s_3, \ldots, s_{n-3}, s_{n-2}, s_{n-1}, s_n),$$

$n = m^2$, for $s_i = \mathrm{syst}_i(U(m), g)$, where g runs over all *left-invariant* Riemannian metrics on the unitary group $U(m)$.

What is most interesting is to understand the asymptotic behavior of systoles as g goes to infinity by $g \mapsto A^k g_0$ for some left invariant endomorphism $A : T(U(m)) \to T(U(m))$ and $k \to \infty$. Nothing seems to be especially easy about the geometry of such spaces as $(U(m), A^k g_0)$ for large k (compare [Gro]$_{\mathrm{SII}}$ and [Gro]$_{\mathrm{CC}}$).

$\mathbf{F_+}$ Finer inequalities and systoles of universal spaces

4.46. Bounding syst_1 in the presence of extra topology. Let V be a closed, aspherical Riemannian manifold. We seek inequalities of the form

$$\mathrm{vol}(V)^{1/n} \geq |\,\mathrm{Top}\,|\, \mathrm{syst}_1(V), \tag{Top}$$

where $|\,\mathrm{Top}\,|$ is some measure of the topological complexity of V. We have already met such an inequality for surfaces of genus γ with $|\,\mathrm{Top}\,| = \mathrm{const}\sqrt{\gamma}/\log(\gamma)$. This generalizes to higher dimensions with *the simplicial volume* $\Delta = \Delta(V)$ defined in Ch. 5.F$_+$. Namely, Top is valid with $|\,\mathrm{Top}\,| = \mathrm{const}_n\, \Delta^{1/n}/\log(1+\Delta)$. To use this, we need examples of manifolds V with large Δ. These are provided by manifolds V admitting Riemannian metrics with sectional curvature ≤ -1 and volumes $\geq \delta$, where $\Delta \geq \varepsilon_n \delta$ for some universal $\varepsilon_n > 0$. Thus, *one cannot substantially increase the first systole of a manifold* (V, g_0) *with* $K(g_0) = -1$ *by deforming* g_0 *while keeping the volume* (V, g_t) *constant in* t (see [Gro]$_{\mathrm{FRM}}$).

Another Top-inequality relates syst_1 to the sum b of all Betti numbers of V, where Top holds true with

$$|\,\mathrm{Top}\,| = C_n b^{1/n}/\exp(C_n' \sqrt{\log(b)}),$$

while it is unknown with

$$|\,\mathrm{Top}\,| = C_n b^{1/n}/\log(1+b),$$

(see [Gro]$_{\mathrm{SII}}$ for further discussion).

4.47. Finding several short loops. If V is essential and $\mathrm{vol}(V)$ is small, one may expect that there are several "independent" closed curves (possibly loops based at the same point in V) of lengths $\ell_1, \ell_2, \ldots, \ell_k$, where all the ℓ_i are rather small, e.g.,

$$(\ell_1 \ell_2 \cdots \ell_k)^{n/k} \leq \mathrm{const}\, \mathrm{vol}(V)$$

(for $k \leq n \leq \dim(V)$). More generally, one may take some class of metric graphs, i.e., 1-polyhedra with path metrics, say $\Gamma = \{\gamma\}$, and ask whether there is a short map of some γ to V, such that the image of $\pi_1(\gamma)$ in $\pi_1(V)$ has rank $\geq k$ in a suitable sense. We refer to [Gro]$_{\text{FRM}}$ for results available in this direction, and here we look at this problem from the point of view of *universal spaces.*

4.48. Systoles syst_n of canonical metrics on $K(\Pi,1)$ spaces. Start with an abstract group Π and suppose we find some "canonical" *metric* space X with a (free) isometric Π-action. Then the n-th systole of the quotient space X/Π and the mass (or volume) norm on $H_n(X/\Pi)$ in general, can be regarded as some "algebraic" invariant of Π defined by geometric means.

Examples. (a) Consider the unit sphere S^∞ in the space of bounded functions $\mathbb{N} \to \mathbb{R}$ with the sup (i.e., ℓ_∞) norm. We give this sphere the induced path metric and make the group $\mathbb{Z}_2$ act on S^∞ by the involution $x \leftrightarrow -x$. Thus, we obtain the projective space $P^\infty = S^\infty/\mathbb{Z}_2$ with a "canonical" metric which, as we know, is a $K(\mathbb{Z},1)$ space, from which it follows that $H_i(P^\infty;\mathbb{Z}_2) = \mathbb{Z}_2$ for all $i = 0,1,\ldots$. Our basic systolic inequality trivially implies that

$$\text{syst}_k\, P^\infty \geq \varepsilon_k > 0,$$

and in fact, this is (trivially) equivalent to the bound on $syst_1$ in manifolds V homeomorphic to $\mathbb{R}\mathrm{P}^n$ by $\text{const}_n(\text{vol}\, V)^{1/n}$ (see [Gro]$_{\text{FRM}}$).

(a') **Remark.** Notice that this P^∞ is Finslerian rather than Riemannian, and so there are many possibilities for defining the notion of volume (or mass) for k-cycles in P^∞. These different volumes correspond to how we normalize the Lebesgue measure μ in the k-dimensional normed spaces $(H, \|\cdot\|)$. One can use, for example, our volumes vol and vol_+ from 4.25–27, or the Hausdorff measure for $\|\cdot\|$, etc. Of course, everything boils down to assigning a certain value to the measure $\mu(B(1))$ of the unit $\|\cdot\|$-ball in V, and there are many (too many) *canonical* assignments of this $\mu(B(1))$ (see [Gro]$_{\text{FRM}}$ for further discussion).

(a'') **Problem.** Determine the asymptotic behavior of $\text{syst}_k\, P^\infty$ for $k \to \infty$. (The answer will certainly depend on a particular canonical mass_k (vol_k) we choose, and we may expect nice asymptotics only for judicious choices of this mass_k.) The same question can be raised for P^∞ associated to an arbitrary "natural" infinite dimensional Banach space instead of $\ell_\infty(\mathbb{N})$, but the answer seems to be unknown for all except ℓ_2.

(*b*) Let X be the unitary group $U(\infty)$, and let a group Π act on X via some unitary representation, say the regular one. This X comes along with the operator norm metric, and so the systolic question is well-posed. One knows by the *Kuiper's theorem* that X/Π is a $K(\Pi;1)$ space, and the general filling argument of $[\mathrm{Gro}]_{\mathrm{FRM}}$ seems to apply to this space. Thus, apparently

$$\mathrm{syst}_n(X/\Pi) \geq \varepsilon_k > 0$$

for all k, where $H_k(\Pi) \neq 0$. Here again, one wishes to know more about the syst_k. Also, one may try different metrics on spaces of linear operators, such as the Hilbert-Schmidt or Schatten's ℓ_p, but then one must be aware of the extra topology harbored by certain (smaller) groups of operators.

(*c*) **Uryson spaces revisited.** Let us pick up a set D of left invariant metrics on a group Π and consider all metric Π-spaces X where the metric on each orbit $\mathrm{dist}_X|_{\Pi(x)}$ belongs to our distinguished set D. Then one can construct as in $3.11\frac{2}{3}_+$ a universal space, say $\mathcal{U} = \mathcal{U}(\Pi, D)$ such that

1. $\mathcal{U}$ is a Polish metric space
2. $\mathrm{dist}_{\mathcal{U}}\,\Pi(u) \in D$ for all $u \in \mathcal{U}$
3. Every Π-space X satisfying (1) and (2) admits an isometric Π-equivariant embedding into $\mathcal{U}$.
4. If X/Π is compact and $Y \subset X$ is an invariant subset, then every equivariant isometric embedding $Y \to \mathcal{U}$ extends to an equivariant isometric embedding $X \to \mathcal{U}$.

Next, one starts looking at $\mathrm{syst}_k(\mathcal{U}/\Pi)$ regarding these as certain invariants of (Π, D). In particular, if D consists of all metrics satisfying $\mathrm{dist}(\pi_1, \pi_2) \geq 1$ for $\pi_1 \neq \pi_2$, then our basic systolic inequality in 4.41 is essentially equivalent to

$$\mathrm{syst}_k(\mathcal{U}/\Pi) \geq \varepsilon_k > 0$$

whenever $H_k(\Pi) \neq 0$. And choosing different sets D often leads to finer results about short loops in essential polyhedra (see $[\mathrm{Gro}]_{\mathrm{FRM}}$ for an approach to this problem via a different class of universal spaces). And there is much left to be done in this field! (See 5.43_+ for different classes of universal spaces.)

Systolic reminiscences. I was exposed to metric inequalities in the late 60's by Yu. Burago, who acquainted me with the results of Loewner, Pu, and Besicovitch. These attracted me by the topological purity of their underlying assumptions, and I was naturally tempted to prove similar inequalities in a more general topological framework, where the first steps had already been made by M. Berger (see our citations of [Berger]). Since the setting was so plain and transparent, I expected rather straightforward proofs based entirely on the coarea formula and/or product inequalities for exterior forms dualized with Hahn–Banach (as in the above Ch. 4.D). Having failed to find such a proof, I was inclined to look for counterexamples but realized to my surprise that the geometrically rather shallow co-area idea gets an unexpected boost from the topological (surgical) induction on dimension once the perspective has been shifted from systoles to the (isoperimetric) filling problem. This gave the desired bounds on $\mathrm{syst}_1/(\mathrm{vol})^{1/n}$ (albeit not with sharp constants) and encouraged a search for similar inequalities for syst_k, $k \geq 2$.

On the other hand, if one wanted a counterexample, one needed a new geometric tool for bounding from below the volume of a k-cycle c in a riemannnan manifold where one had to forfeit the use of pointwise small closed k-forms with large integrals $\int_C \omega$, since this would also bound the $\mathbb{R}$-mass of C, in contradiction with the stable inequalities (see Ch. 4.D). Only belatedly (lecturing at the University of Maryland in the Spring of 1993) I realized that such a bound for $k \geq 2$ could be reduced to that for $k = 1$ with a variety of examples emerging via a geometric surgery applied to the inocuous-looking twisted product $S^3 \ltimes S^1$ (see 4.45) and further cartesian products $(S^3 \ltimes S^1) \times V'$ (see $[\mathrm{Berger}]_{\mathrm{Sys}}$ and Appendix D_+). So the surgery proved useful for obtaining lower bounds on syst_k in *specific examples* as well as for obtaining upper bounds on syst_1 for *general* (essential) (V, g), thus giving us a rather clear view on the possible range of the systolic vector

$$g \mapsto (\mathrm{syst}_1(g), \dots, \mathrm{syst}_n(g) = \mathrm{vol}(V, g)) \in \mathbb{R}^n$$

for a manifold V of a given homotopy type. And so the time came to look deeper into the geometry of spaces of metrics g with given systoles, or, even better, with a prescribed "volumic norm" on $H_*(V)$ (see $[\mathrm{Gro}]_{\mathrm{SII}}$ for some preliminary questions and conjectures).

Chapter 5

Manifolds with Bounded Ricci Curvature

A. Precompactness

In this chapter, we consider locally compact, pointed path metric spaces and the metric space structure on the collection of such spaces defined by the Hausdorff distance (or, more precisely, the uniform structure on this set defined by the family of Hausdorff distances on the balls of radius R).

5.1. Definition: A family (X_i, x_i) of path metric spaces is *precompact* if, for each $R > 0$, the family $B^{X_i}(x_i, R)$ is precompact with respect to the Hausdorff distance d_H.

Observation: For each $\varepsilon > 0$ and $R > 0$, we denote by $N(\varepsilon, R, X)$ the maximum number of disjoint balls of radius ε that fit within the ball of radius R centered at an $x \in X$. The function $X \mapsto N(\varepsilon, R, X)$ is almost continuous; indeed, if $d_H(B^X(x,R), B^Y(y,R)) < \delta$ and $N(\varepsilon, R, X) = N$, then there exist N points $x_1, \dots, x_N$ in $B^X(x,R)$ such that $\mathrm{dist}(x_i, x_j) > 2\varepsilon$ for $i \neq j$. This 4ε-net induces a net y_i in $B^Y(y,R)$ such that $\mathrm{dist}(y_i, y_j) > 2\varepsilon - 2\delta$ for $i \neq j$. Consequently, $N(\varepsilon - \delta, R, Y) \geq N$ and conversely, $N(\varepsilon - \delta, R, Y) \geq N(\varepsilon, R, Y)$. In particular, the functions $N(\varepsilon, R, \cdot)$ are bounded on every precompact family of path metric spaces.

5.2. Proposition: *A family (X_i, x_i) of pointed path metric spaces is precompact if and only if each function $N(\varepsilon, R, \cdot)$ is bounded on (X_i). In this case, the family is relatively compact, i.e., each sequence in the X_i admits a subsequence that Hausdorff-converges to a complete, locally compact path metric space.*

Proof. Fix $\varepsilon > 0$ and $R > 0$ and set $N = \sup N(\varepsilon, R, X_i)$. Using the hypothesis that N is finite, we must prove that there is a finite number of indices $i_1, \ldots, i_p$ such that, for each i, there exists j with

$$d_H(B^{X_i}(x_i, R), B^{X_{i_j}}(x_{i_j}, R)) \leq \text{const}\, \varepsilon.$$

To this end, it suffices to consider each subfamily $N(\varepsilon, R, X_i) = n$ for $1 \leq n \leq N$ individually, so that we may assume $N(\varepsilon, R, X_i) = N$ for all i. In each ball $B^{X_i}(x_i, R)$, the maximal disjoint filling is realized by balls of radius ε centered at points of a 2ε-net R_i. Indeed, if there exists a point y of $B^{X_i}(x_i, R)$ located at a distance greater than 2ε from each point of R_i, then the ball $B(y, \varepsilon)$ would be disjoint from the balls $B(x, \varepsilon)$ for $x \in R_i$, and therefore the net R_i would not be maximal. A choice of bijection $f : \{1, \ldots, N\} \to R_i$ induces a distance d_i on the set $\{1, \ldots, N\}$. This function $d_i : \{1, \ldots, N\}^2 \to \mathbb{R}_+$ only assumes values between 2ε and $2R$ on the subset A of pairs of distinct integers, which has $N^2 - N$ elements. Since the space $[2\varepsilon, 2R]^{N^2-N}$ is precompact in the product metric, there exist finitely many indices $i_1, \ldots, i_p$ such that, for each i, there is a j with

$$\sup_A |d_i(m, q) - d_{i_j}(m, q)| < \varepsilon^2,$$

and so

$$\sup_A \frac{d_i(m, q)}{d_{i_j}(m, q)} \leq \frac{1 + \varepsilon}{1 - \varepsilon},$$

that is, $d_L(R_{i_j}, R_i) \leq \log((1 + \varepsilon)/(1 - \varepsilon))$, where $d_H(R_{i_j}, R_i) \leq R \log((1 + \varepsilon)/(1\varepsilon))$. Finally, $d_H(B^{X_i}(x_i, R), B^{X_{i_j}}(x_{i_j}, R)) \leq (2R + 4)\varepsilon$ approximately, which proves precompactness.

To prove compactness, we first assume that X_i is a sequence of path metric spaces of radius R and argue as above: We fix $\varepsilon > 0$, extract a subsequence such that $N(\varepsilon, R, X_i) = N$, construct the 2ε-nets R_i, deduce from this the existence of functions d_i on $A = \{(m, q) : m, q \leq N, m \neq q\}$, and extract a subsequence d_{i_j} that converges to a distance on $\{1, \ldots, N\}$. In order to apply a diagonal argument, we set $Z_1 = X_{i_1}$ and $R_1^1 = R_i$, and note that R_1^1 is a 2ε-net in Z_1 having Lipschitz distance less than $\varepsilon/2$ from a distance d_1 on $\{1, \ldots, N_1\}$, where $N_1 = \text{card}(R_1^1)$. Now set $Y_j = X_{i_j} \setminus U_\varepsilon(R_{i_j})$, and let S_j be the set of centers of a maximal disjoint set of radius $\varepsilon/2$ balls in Y_j that are pairwise disjoint relative to X. Using the fact that the X_j are path metric spaces, it is easy to check that $R_{i_j} \cup S_j$ is an ε-net in X_{i_j} and that the radius $\varepsilon/2$ balls centered on $R_{i_j} \cup S_j$ are disjoint. It follows that $\text{card}(R_{i_j}) + \text{card}(S_j)$ is bounded and therefore equal to some constant N_2 for a subsequence X_{n_p}. For each p, we choose a bijection $\{1, \ldots, N_2\} \to R_{n_p'} \cup S_p$ extending the one from $\{1, \ldots, N_1\} \to R_{n_p}$ already chosen. This bijection induces a distance d_{n_p} on $\{1, \dots, N_2\}$, and we can extract a subsequence $X_{n'_p}$ such that the $d_{n'_p}$ converge. Since the restrictions of the $d_{n'_p}$ to $\{1, \dots, N_1\}$ converge, the limit distance d_2 extends d_1. We set $Z_2 = X_{n'_0}$, $R_2^1 = R_{n'_0}$, $R_2^2 = R_2^1 \cup S_0$, where n'_0 is chosen so that $\operatorname{dist}\{d_2, d_{n'_0}\} \leq \varepsilon/4$.

This procedure produces a subsequence Z_k of the X_i, a sequence N_k of integers, a distance d on $\mathbb{N}$, and a nested sequence of $(\varepsilon/2^{i-2})$-nets R_k^i in Z_k, having a Lipschitz distance less than $\varepsilon/2^i$ from $\{1, \dots, N_k\}$. Since $\{1, \dots, N_k\}$ is a $(\varepsilon/2^{i-3})$-net in $(\mathbb{N}, d)$, we conclude that Z_k tends to $(\mathbb{N}, d)$ with respect to the Hausdorff distance (see Proposition 3.5). Finally, the completion $\hat{\mathbb{N}}$ of $(\mathbb{N}, d)$ is also a limit of the Z_k and, by Theorem 1.8, is therefore a locally compact path metric space.

For a family (X_i, x_i) of pointed spaces satisfying the hypotheses of the proposition, we can, again by a diagonal argument, extract a subsequence i_j such that the balls $B^{X_{i_j}}(x_{i_j}, R)$ converge, which proves compactness for pointed convergence (Definition 3.14).

5.3. Theorem: *The set of pointed Riemannian n-manifolds satisfying the inequality*

$$\operatorname{Ricci}(g) \geq (n-1)rg,$$

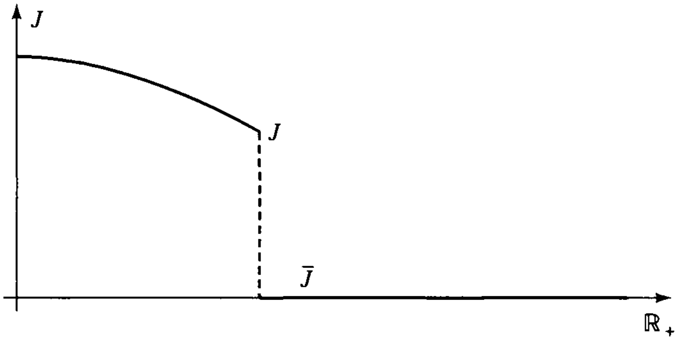
where g is the metric and r a real number, is precompact with respect to the pointed Hausdorff distance.

Indeed, Bishop's inequality enables us to bound the number of small balls within a large ball (cf. [Bish–Crit], p. 253): If V_r denotes the simply connected space of constant curvature whose Ricci curvature is equal to $-(n-1)rg$, then Bishop's inequality states that for $v \in V$, $x \in V_r$, if $U \subset V$ is diffeomorphic via $\exp_v$ to an open subset of T_vV, we can define a diffeomorphism f of U onto an open set of V_r by composing $\exp_v^{-1}$ with an isometry of T_vV onto T_xV_r and subsequently with $\exp_x \colon T_xV_r \to V_r$. Then the Jacobian J of f^{-1} decreases as we move away from v along a geodesic. From this, we can first deduce

5.3.bis Lemma: *Let V be a Riemannian n-manifold satisfying* $\operatorname{Ricci}(g) \geq -(n-1)rg$, *where g is the metric and r a real number. Let v be a point of V and choose real numbers $R \leq R'$. Then*

$$\frac{\operatorname{vol}(B^V(v,R))}{\operatorname{vol}(B^V(v,R'))} \geq \frac{\operatorname{vol}(B^{V_r}(x,R))}{\operatorname{vol}(B^{V_r}(x,R'))} = \frac{\int_0^R (\sinh(\sqrt{r}t))^{n-1}\,dt}{\int_0^{R'} (\sinh(\sqrt{r}t))^{n-1}\,dt}.$$

One interest of this lemma is that it holds even if one of (or both) the constants R, R' are greater than the injectivity radius at v. To see this, let U be the complement of the cut locus of V, let A (resp. A') be the subset of V_r that is the image under f of $U \cap B^V(v,R)$ (resp. $U \cap B^V(v,R')$).



With respect to the canonical measure $d\mu$ of V_r, we have

$$\mathrm{vol}(B^V(v,R)) = \int_A J\,d\mu, \qquad \mathrm{vol}(B^{V_r}(x,R)) = \int_{B^{V_r}(x,R)} d\mu,$$

$$\mathrm{vol}(B^V(v,R')) = \int_{A'} J\,d\mu, \qquad \mathrm{vol}(B^{V_r}(x,R')) = \int_{B^{V_r}(x,R')} d\mu.$$

Moreover, we can write

$$\mathrm{vol}(B^V(v,R')) = \int_{B^{V_r}(x,R)} J\,d\mu, \qquad \mathrm{vol}(B^V(v,R')) = \int_{B^{V_r}(x,R')} J\,d\mu$$

simply by extending J to $\hat{J}$ by zero! The function $\hat{J}$ decreases along geodesics, as does J.

Now we consider polar coordinates $(u,t) \in S \times \mathbb{R}_+$ based at x, where $S = S_xV_r$ is the unit sphere of tangent vectors at $x \in V_r$. We then have $d\mu = d\sigma \times d\nu$, where σ is the canonical measure of S and $d\nu$ is a measure on $\mathbb{R}_+$ given by $d\nu = (\sqrt{r})^{1-n}(\sinh(\sqrt{r}t))^{n-1}$. We then have

$$\mathrm{vol}(B^V(v,R)) = \int_S \left(\int_0^R \hat{J}(u,t)\,d\nu\right) d\sigma,$$

$$\mathrm{vol}(B^{V_r}(x,R)) = \left(\int_S d\sigma\right)(\sqrt{r})^{1-n}\int_0^R (\sinh(\sqrt{r}t))^{n-1}\,dt$$

$$\mathrm{vol}(B^V(v,R')) = \int_S \left(\int_0^{R'} \hat{J}(u,t)\,d\nu\right) d\sigma,$$

$$\mathrm{vol}(B^{V_r}(x,R')) = \left(\int_S d\sigma\right)(\sqrt{r})^{1-n}\int_0^{R'} (\sinh(\sqrt{r}t))^{n-1}\,dt.$$

However, the fact that $J(u,t)$ decreases with respect to t implies that

$$\frac{\int_0^R \hat{J}(u,t)\,d\nu}{\int_0^R d\nu} \geq \frac{\int_0^{R'} \hat{J}(u,t)\,d\nu}{\int_0^{R'} d\nu},$$

proving the lemma.

End of the proof of 5.3. If $v' \in V$ and $\operatorname{dist}(v,v') \leq R$, then $B(v,R) \subset B(v',2R)$, and so

$$\operatorname{vol}(B^V(v',\varepsilon)) \geq \varphi_n(\varepsilon,R,r)\operatorname{vol}(B^V(v',2R)) \geq \varphi_n(\varepsilon,R,r)\operatorname{vol}(B^V(v,R)),$$

and finally $N(\varepsilon,R,V) \leq \varphi_n(\varepsilon,R,r)$, where

$$\varphi_n(\varepsilon,R,r) = \frac{\int_0^R (\sinh(\sqrt{rn}))^{n-1}\,dt}{\int_0^\varepsilon (\sinh(\sqrt{rn}))^{n-1}\,dt}$$

depends only on $(\varepsilon,R,r;n)$.

5.4. Remark: The meaning of the preceding theorem is that the restriction

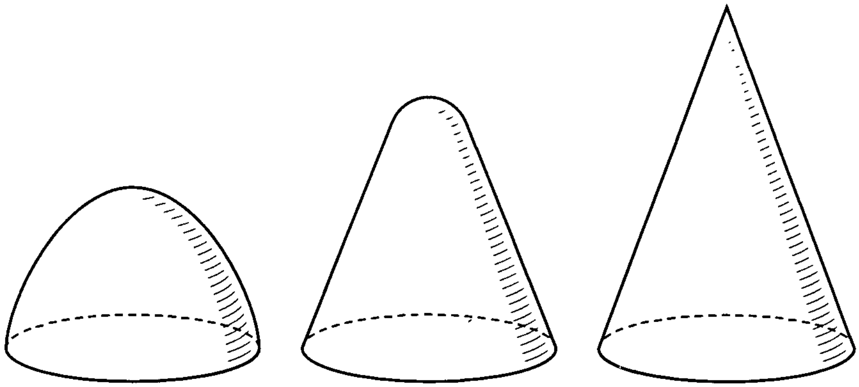
$$\operatorname{Ricci}(g) \geq -(n-1)rg$$

permits only a finite number of geometries in the following sense: For each $\varepsilon > 0$, we can choose a finite number of model spaces X_i such that a ball of radius R in a such a manifold "resembles" the radius-R ball in one of the X_i on the level of distances between points, up to ε.

If we considered manifolds that also satisfy $\operatorname{diam}(V,g) \leq d$, we would have a precompact space for the usual Hausdorff distance, and we ask ourselves, How much do these conditions restrict the underlying topologies? We will proceed further in this direction at the end of the present chapter and in Chapter 8.

5.5. Counterexamples: Manifolds of dimension n having bounded Ricci curvature form a relatively compact (but not compact) space, i.e., a sequence of these manifolds can converge to a path metric space that is not an n-manifold. For example, let (e_1, e_2) be an orthonormal basis of the euclidean plane, Λ_ε the lattice generated by the vectors e_1 and εe_2, and $\mathbb{T}_\varepsilon$ the flat torus $\mathbb{R}^2/\Lambda_\varepsilon$. Then $\operatorname{Ricci}(\mathbb{T}_\varepsilon) = 0$ and the family $\mathbb{T}_\varepsilon$ converges to the circle $\mathbb{R}e_1/\mathbb{Z}e_1$ of lower dimension.

Even when the dimension remains the same, singularities may appear in the limit space: The surfaces C_ε of truncated cones in $\mathbb{R}^3$, equipped



with the induced length structures, converge to the cone, which is not a Riemannian manifold.

More is said about these two limit phenomena in Chapter 8.

For now, we mention the following applications of Proposition 5.2.

5.6. Proposition: *If X, Y are path metric spaces separated by a finite Hausdorff-Lipschitz distance, and if the family $\{\lambda X : \lambda > 0\}$ is precompact, then the family $\{\lambda Y : \lambda > 0\}$ is also precompact.*

Proof. By hypothesis, there exist two path metric spaces X_1, Y_1 such that

$$d_H(X, X_1) = a < \infty, \qquad d_L(X_1, Y_1) = b < \infty \qquad d_H(Y_1, Y) = c < \infty.$$

The net M' consisting of the centers of a maximal disjoint family of ε-balls in $B^Y(y, R)$ induces a net in Y_1, then in X_1, and finally a net M in X, and the ε_1'-balls centered on M are disjoint in $B^X(x, R_1')$, where $\varepsilon_\lambda' = e^{-b}(\varepsilon - \lambda c) - \lambda a$ and $R_\lambda' = e^b(R + \lambda c) + \lambda a$, from which we can deduce that $N(\varepsilon, R, Y) \leq N(\varepsilon_1', R_1', X)$, and so, since the balls of λX are λ times smaller than those of X, $N(\varepsilon, R, \lambda Y) \leq N(\varepsilon_\lambda', R_\lambda', \lambda X)$. The second term is bounded (in λ) by hypothesis, and so the $N(\varepsilon, R, \lambda Y)$ are bounded.

Application: Let V be a compact manifold whose fundamental group is isomorphic to the discrete Heisenberg group, let Y be its universal cover, and let X be the continuous 3-dimensional Heisenberg group with its "Carnot" metric (see 1.1). Following Remark 2.23, the distance $d_{HL}(X, Y)$ is finite. The spaces λX are all identical and therefore form a precompact family, so that the same is true of the homotheties of Y. Thus, there exists a sequence $\lambda_n \to 0$ such that the sequence $\lambda_n Y$ Hausdorff converges to a complete, locally compact path metric space.

5.7$_+$. The above can be reversed and used for a geometric proof of the *Mil-*

nor conjecture for groups of *polynomial growth* (defined in 5.9 below), which claims that these are *virtually nilpotent*, i.e., contain nilpotent subgroups of finite index.

Exercise: Let Γ have polynomial growth. Show that the family $\{\varepsilon\Gamma : \varepsilon > 0\}$ is precompact. Study the Hausdorff limit H of $\varepsilon_i\Gamma$ for some sequence $\varepsilon_i \to 0$ and prove that it is a locally compact, finite-dimensional metric space with a transitive isometry group.

It follows via D. Montgomery's solution of Hilbert's fifth problem that H is *transitively* acted upon by a *connected Lie* group. Then one can show that Γ admits sufficiently many linear representations in order to apply

Tits' freedom theorem: *Every non-virtually solvable, finitely generated subgroup of $GL_n(\mathbb{R})$ contains a free subgroup on two generators.*

This shows that our Γ is a virtually solvable, and by Bass' argument, a virtually nilpotent group (see [Gro]$_{\text{GPG}}$ for details).

5.8$_+$ Problem: Is there an algebraic (or geometric) proof that avoids the limits? For example, one may try to develop a discrete version of the topological group techniques from [Mont–Zip].

Exercise: Prove directly that every group of polynomial growth of degree ≤ 2 is virtually abelian. (This is true by the above for degree < 4, but no direct argument is available, compare [Inc].)

B. Growth of fundamental groups

In this section, Γ will denote a discrete group with a finite set $\gamma_1, \ldots, \gamma_p$ of generators, $B(R)$ will be the ball of radius R with respect to the algebraic norm associated with this system (i.e., the set of elements of Γ that can be represented by a word having length less than R in this set), and $N(R)$ will denote the number of elements in $B(R)$.

5.9. Definition: The group Γ is said to have *polynomial growth* if there exists a real p such that $N(R) \leq R^p$ for sufficiently large R. If there is a $c > 0$ such that $N(R) \geq e^{cR}$ for large R, then Γ is said to have *exponential growth.*

5.10. Proposition: *The growth type is invariant under quasi-isometries.*

Proof. For a net M in a metric space X, we denote by $N(M, R)$ the number of points of M in the ball of radius R centered at x. Let Γ_1, Γ_2 be two quasi-isometric groups, so that there exist two compact manifolds V_1, V_2 with fundamental groups Γ_1, Γ_2, respectively, and metric spaces X_1, X_2 such that

$$d_H(\tilde{V}_1, X_1) = \rho_1 < \infty, \qquad d_L(X_1, X_2) = b < \infty, \qquad d_H(X_2, \tilde{V}_2) = \rho_2 < \infty.$$

These spaces are pointed at $\tilde{v}_1, x_1, x_2, \tilde{v}_2$. Let M_1, M_2 be the nets $\Gamma_1\tilde{v}_1, \Gamma_2\tilde{v}_2$. If d_1 is the diameter of V_1, then M_1 is a d_1-net of $\tilde{V}_1$. Since the algebraic and geometric norms on Γ_1 are equivalent (see Proposition 3.22), it follows that for some constant a, we have $N(M_1, R) \leq N_1(aR)$ (= the $N(R)$ of Γ_1). By the usual procedure, we associate with each point of M_1 a point of X_1 in order to obtain a $(d_1 + 2\rho_1)$-net M_1' of X_1, such that $N(M_1', R) \leq N_1(aR + \rho_1)$. There exists a homeomorphism of X_1 onto X_2 having dilatation less than e^b; the image of M_1' is a $e^b(d_1 + 2\rho_1)$-net M_2' of X_2 such that $N(M_2', R) \leq N(e^b(aR+\rho_1))$. Since $d_H(X_2, V_2) \leq \rho_2$, we have transformed the net M_2' into a $(e^b(d_1 + 2\rho_1) + 2\rho_2)$-net M_2'' in $\tilde{V}_2$ such that $N(M_2'', R) \leq N_1(e^b(aR + \rho_1) + \rho_2)$.

We now have two nets M_2, M_2'' of $\tilde{V}_2$, and we will show that their respective densities are comparable. Set $\delta = e^b(d_1 + 2\rho_1) + 2\rho_2$ and note that the number of elements of M_2 contained in any δ-ball is bounded (by a real number K) since Γ_2 acts by isometries. For each $x \in (M_2)^R = M_2 \cap B^{V_2}(v_2, R)$, we choose a $y = f(x) \in (M_2'')^{R+\delta}$ such that $d(x, y) \leq \delta$, which is possible since M_2'' is a δ-net. Then, each fiber of the map $f\colon M_2^R \to (M_2'')^{R+\delta}$ has at most K points, which implies that

$$N_2(R) = \operatorname{card}(M_2^R) \leq K\, N(M_2'', R + \delta) \leq K\, N_1(e^b(aR + \rho_1) + \rho_2). \quad (*)$$

For the polynomial case of the assertion, we set

$$p_i = \inf\{p \in \mathbb{R} : N_i(R) \leq R^p \text{ for sufficiently large } R\}.$$

The assumption that Γ_1 has polynomial growth means that p_1 is finite, but it follows from the inequality $(*)$ that for each $p > p_1$, we then have $N_2(R) \leq R^p$ for sufficiently large R, and so $p_2 \leq p_1$. Of course, by symmetry we can conclude that $p_1 = p_2$, and so both are simultaneously finite.

For the exponential case, we deduce from $(*)$ that

$$N_1(R) \leq \frac{1}{K} N_2\left(e^{-b}\left(\frac{R-\delta}{a} - \rho_1\right) - \rho_2\right).$$

If Γ_2 has exponential growth, then there exists $c > 0$ such that, for sufficiently large R, we have $N_2(R) \leq e^{cR}$, which implies that for $c' < c$ and large R, $N_1(R) \leq \exp(ce^{-b}a^{-1}R)$. Thus, Γ_1 has exponential growth.

This proof leads to the conclusion that

$$\liminf_{R\to\infty} \frac{\log(N(R))}{\log(R)}$$

is invariant of quasi-isometry, which gives meaning to the expression "*polynomial growth of degree p*." On the other hand,

$$\liminf_{R\to\infty} \frac{\log(N(R))}{R}$$

is not invariant, which justifies the following definition.

5.11. Definition: The *entropy* of a set $\{\gamma_i\}$ of generators is the number

$$h(\{\gamma_i\}) = \liminf_{R\to\infty} \frac{\log(N(R))}{R},$$

and the *entropy of the group* Γ, denoted $h(\Gamma)$, is the infimum of the entropies of all sets of generators.

5.12. Remark: If $h(\Gamma)$ is strictly positive, then Γ has exponential growth. On first sight, however, there seems to be no reason for the converse to be true. Nevertheless, we do not know of a counterexample.[1]

5.13. Example: *For the free group* Γ *on* k *generators,* $h(\Gamma) = \log(2k-1)$.

Indeed, if $\gamma_1, \ldots, \gamma_k$ are independent generators, then $N(R)$ is exactly the number of words of length less than R. But in order to create a word of length R, it is necessary to choose the first letter among one of the $2k$ letters γ_i, γ_i^{-1}, then to choose among the $(2k-1)$ letters for the second, etc., and so

$$N(R) = N(R-1) + 2k(2k-1)^{R-1} = \frac{2k}{2k-2}(2k-1)^R$$

and $h\{\gamma_i\} = \log(2k-1)$. If $\beta_1, \ldots, \beta_n$ is another set of generators, we select k independent generators $\beta_{i_1}, \ldots, \beta_{i_k}$. The minimal length of an element of Γ as a function of the β_{i_j} is larger than for $\beta_1, \ldots, \beta_n$, from which we conclude that $h\{\beta_i\} \geq \log(2k-1)$.

5.14. Conjecture: If Γ is a discrete group with k generators and p relations, then $h(\Gamma) \geq \log(2(k-p)-1)$.

[1] I have recently received a preprint discussing this subject matter, but to my shame I have lost it. My apologies to the author.

We have been able to resolve this question in several cases by exhibiting free subgroups. If $p = 1$, then the conjecture follows from the "Freiheit" theorem of Magnus (see [Lynd–Sch]). Stallings [Stall] has shown that, if $p \leq k - 2$, and if there is an integer q such that $H_4(\Gamma, \mathbb{Z}/q\mathbb{Z}) = \Gamma/[\Gamma, \Gamma] \oplus \mathbb{Z}/q\mathbb{Z}$ has rank $k - p$, then the conjecture is true. Finally, we have in general

$$h(\Gamma) \geq \text{const} \log(2(k-p) - 1),$$

which can be seen by looking at some subfactor group of Γ (see [Bau–Pri] and p. 83 in $[\text{Gro}]_{\text{VBC}}$).

Remark$_+$: It is known that the entropy of the *generators* $\gamma_1, \ldots, \gamma_k$ is at least $\log(2(k-p) - 1)$. Moreover, by a theorem of N. S. Romanovski (see [Rom]), *there are $k-p$ generators among the γ_i which are freely independent.*

For the universal cover of a compact Riemannian manifold, we will relate the growth of balls with the growth of the fundamental group.

5.15. Definition: The *entropy* of a compact Riemannian manifold V is the number

$$h(V) = \liminf_{R \to \infty} \frac{\log(\text{vol}(B^{\tilde{V}}(\tilde{v}, R)))}{R}.$$

Remark: This number is independent of the base point $\tilde{v} \in \tilde{V}$. Indeed, if $\tilde{v}, \tilde{v}' \in \tilde{V}$, then $B^{\tilde{V}}(\tilde{v}', R) \subset B^{\tilde{V}}(\tilde{v}', R + \delta)$, where $\delta = d(\tilde{v}, \tilde{v}')$, and so for each $c > h(V, \tilde{v})$ and sufficiently large R,

$$\frac{\log(\text{vol}\, B(v', R))}{R} \leq c + \frac{\delta}{R}.$$

Thus, $h(V, \tilde{v}') \leq c$ and finally $h(V, \tilde{v}') \leq h(V, \tilde{v})$.

Remark: In [Din] and [Mann], it is shown that the topological entropy of the geodesic flow on the unit tangent bundle of a manifold V is bounded below by $h(\pi_1(V))$, which justifies the use of the term *entropy* for this invariant group as well as the number related to the volume of balls.

5.16. Theorem (cf. [Šva] and $[\text{Milnor}]_{\text{CFG}}$): *If V is a compact manifold of diameter d, then $h(V) \leq 2d\, h(\pi_1(V))$.*

Proof. There exists an $\varepsilon > 0$ such that every closed curve in V of length $< 2\varepsilon$ is homotopic to 0. Thus, each element of $\Gamma = \pi_1(V)$ has geometric norm greater than ε (except the identity), which implies that the balls $B^{\tilde{V}}(\gamma\tilde{v}, \varepsilon/2)$ of radius $\varepsilon/2$ centered at points of the orbit $\Gamma\tilde{v}$ are disjoint. Let τ be the common volume of these small balls; the number $N'(R)$ of elements of $\Gamma\tilde{v}$ lying within the ball of center $\tilde{v}$ and radius R is then bounded above by $\tau^{-1}\operatorname{vol}(B^{\tilde{V}}(\tilde{v},R))$, and consequently,

$$h(V) = \liminf_{R\to\infty} \frac{\log(\operatorname{vol} B^{\tilde{V}}(\tilde{v},R))}{R} \geq \liminf_{R\to\infty} \frac{N'(R)}{R}.$$

However, we proved (Proposition 3.22) the inequality $\|\cdot\|_{geo} \leq K\|\cdot\|_{alg}$, where K is the largest geometric norm of a generator, which can be less than $2d$ for an appropriately chosen set of generators. Thus,

$$N(R) = \operatorname{card}\{\gamma \in \Gamma : \|\gamma\|_{alg} \leq R\} \leq N'(2dR),$$

and so

$$h(\Gamma) \leq 2d \liminf_{R\to\infty} \frac{N'(2dR)}{2dR},$$

i.e., $h(\Gamma) \leq 2dh(V)$.

5.17. Application: Suppose that a compact Riemannian manifold (V, g) satisfies $\operatorname{Ricci}(g) \geq -(n-1)rg$ for a constant $r > 0$. Then, by Bishop's inequality (see 5.3) the balls in $\tilde{V}$ grow more slowly than in the simply connected hyperbolic space that satisfies the inequality $\operatorname{Ricci} = -(n-1)rg$, i.e.,

$$\operatorname{vol}(B^{\tilde{V}}(\tilde{v},R)) \leq \operatorname{const} \int_0^R \left(\frac{\sinh(\sqrt{r}t)}{\sqrt{r}}\right)^{n-1} dt \leq \operatorname{const} e^{\sqrt{r}t},$$

and so $h(V) \leq \sqrt{r}$. Using the theorem, we can conclude that the entropy of the fundamental group of V is necessarily less than $\sqrt{r}/2d$. In particular, if $\pi_1(V)$ is free, it has at most $(\exp(\sqrt{r}/2d)+1)/2$ generators.

In this connection, we observe that sometimes the hypothesis $\operatorname{Ricci}(g) \geq -(n-1)rg$ enables us to bound the growth of balls from below as well as from above (cf. $[\mathrm{Yau}]_{\mathrm{ICFE}}$).

5.18. Proposition: *If V is a complete, noncompact Riemannian manifold with nonnegative Ricci curvature, then for each $\alpha < 1$, there exists a $c_\alpha > 0$ such that* $\operatorname{vol}(B^V(v,R)) \geq c_\alpha R^\alpha$ *for large $R \to \infty$.*

Proof. For each $v \in V$ and R, R', Remark 5.3 implies that

$$\frac{\operatorname{vol}(B^V(v,R))}{\operatorname{vol}(B^V(v,R'))} \leq \frac{R^n}{(R')^n}.$$

Choose a point $v \in V$. Then for each $t > 0$, there is a point $x \in V$ lying at exactly distance t from v. Indeed, since V is a complete, locally compact, noncompact path metric space, V is not bounded (by the Hopf–Rinow theorem 1.10); thus, there exists a point $y \in V$ such that $d(v, y \geq t)$, and if γ is a path connecting v to y, the continuous function $s \mapsto d(v, \gamma(s))$ assumes the value t.

Given $\alpha < 1$, fix $b \in (1, \alpha^{-1})$ and set $r_i = 2^{b^i}, R_i = \sum_{j=0}^{i} 2r_j$. Choose a point x_i at a distance $r_i + R_{i-1}$ from v; by construction, the balls $B(x_i, r_i)$ are disjoint and $\bigcup_{j=0}^{i} B(x_j, r_j) \subset B(x_i, r_i + R_{i-1})$. If we set $\tau_i = \sum_{j=0}^{i} \mathrm{vol}(B(x_j, r_j))$, then $\tau_i \leq \mathrm{vol}(B(x_i, r_i + R_{i-1}))$ and

$$\frac{\tau_i}{\tau_i - \tau_{i-1}} \leq \left(\frac{R_i + R_{i-1}}{r_i}\right)^n,$$

so that $\tau_i \geq (1 - (1 + R_{i-1}/r_i)^{-n})^{-1}\tau_{i-1}$. Note that for each i,

$$R_{i-1} = 2\sum_{j=0}^{i-1} 2^{b^j} \leq 2\sum_{j=0}^{i-1} 2^{[b^j]+1} \leq 2 \sum_{0 \leq k \leq [b^{i-1}]+1} 2^k \leq 2 \cdot 2^{b^{i-1}+2},$$

and so $R_{i-1}/r_i \to 0$ as $i \to \infty$. Moreover, $R_{i-1} = 2r_{i-1} + R_{i-2}$, and so $R_{i-1} \sim 2r_{i-1}$ as $i \to \infty$. It follows that, for sufficiently large i, $1 - (1 + R_{i-1}/r_i)^{-n} \geq 3n(r_{i-1}/r_i)$ and so $\tau_i/\tau_{i-1} \leq r_i/3nr_{i-1}$. However, $r_i/r_{i-1} = (r_{i+1}/r_i)^{1/b}$; since $1/b \geq \alpha$, it follows that for sufficiently large i,

$$\frac{1}{3n}\left(\frac{r_{i+1}}{r_i}\right)^{1/b} \geq \left(\frac{r_{i+1}}{r_i}\right)^{\alpha},$$

and so $\tau_i r_{i+1}^{-\alpha} \geq \tau_{i-1} r_i^{-\alpha}$, and $\tau_i \geq \mathrm{const}(r_{i+1})^{\alpha}$. Since $\bigcup_{j=0}^{i} B(x_j, r_j) \subset B(v, r_i + R_{i-1})$, it follows that $\mathrm{vol}(B(v, R_i)) \geq \mathrm{const}\, r_{i+1}^{\alpha} \geq \mathrm{const}'\, R_{i+1}^{\alpha}$ for sufficiently large i, since $R_i \sim 2r_i$ as $i \to \infty$.

To conclude the proof, we note that if R is sufficiently large, then there is an i such that $R_i \leq R \leq R_{i+1}$ and so

$$\mathrm{vol}(B(v, R)) \geq \mathrm{vol}(B(v, R_i)) \geq \mathrm{const}\, R_{i+1}^{\alpha} \geq \mathrm{const}\, R^{\alpha}.$$

C. The first Betti number

In the preceding section, we found an upper bound for the number of generators of a fundamental group, provided that it is free. We will now generalize this result by refining the technique used in Section 5.17 and by finding an upper bound for the first Betti number in terms of the diameter and Ricci curvature of a manifold.

5.19. Lemma: *Let X be a compact path metric space of diameter d and having a universal cover $\tilde{X}$. Then for each point $\tilde{x} \in \tilde{X}$ and each $\varepsilon > 0$,* *there is a finite covering $X' \to X$ and set of generators $\gamma_1, \dots, \gamma_k$ for the deck transformation (Galois) group of X' such that*

$$\text{I. } d(\tilde{x}, \gamma_i\tilde{x}) \leq 2d + \varepsilon \qquad \text{and} \qquad \text{II. } d(\gamma_i\tilde{x}, \gamma_j\tilde{x}) \geq \varepsilon$$

if $i \neq j$.

Proof. Consider the families $\{\gamma_i\}$ of elements of $\pi_1(X)$ that satisfy the properties I) $\|\gamma_j\|_{geo} \leq 2d + \varepsilon$ and II) if $i \neq j$; then $\|\gamma_i\gamma_j^{-1}\|_{geo} \geq \varepsilon$. Such families exist (e.g., the family $\{e\}$), so that we can choose one having the maximal number p of elements and denote by Γ' the normal subgroup of $\pi_1(X)$ that it generates. With this group is associated a covering space X' of X, defined as the quotient of $\tilde{X}$ by the action of Γ', whose automorphism group is isomorphic to $\pi_1(X)/\Gamma'$. Let x' denote the projection of $\tilde{x}$ into X' and suppose that there exists $z' \in X'$ such that $\operatorname{dist}(x', z') > d + \varepsilon$. If $\tilde{z}$ lies in the fiber over z', then there exists $\alpha \in \pi_1(X)$ such that $\operatorname{dist}(\tilde{z}, \alpha\tilde{z}) \leq d$, and so $\operatorname{dist}(z', \alpha x') \leq d$, which implies that $\varepsilon < \operatorname{dist}(x', \alpha x') < 2d + \varepsilon$, and so 1) there exists $\gamma_0' \in \Gamma'$ such that $\operatorname{dist}(\gamma_0'\tilde{x}, \alpha\tilde{x}) \leq 2d + \varepsilon$, i.e., $\|(\gamma_0')^{-1}\alpha\|_{geo} \leq 2d + \varepsilon$, and 2) for each $\gamma' \in \Gamma'$, $\operatorname{dist}(\gamma'\gamma_0'\tilde{x}, \alpha\tilde{x}) > \varepsilon$, i.e., $\|(\gamma')^{-1}(\gamma_0')^{-1}\alpha\|_{geo} > \varepsilon$. The system $\gamma_1, \dots, \gamma_k, (\gamma_0')^{-1}\alpha$ still satisfies conditions 1) and 2), which contradicts the maximality of the γ_i. We conclude that each point of X' lies at a distance less than $d + \varepsilon$ from x', and so X' is compact with diameter less than $2d + 2\varepsilon$. In particular, Γ' has finite index in $\pi_1(X)$.

5.20. Corollary: By Hurewicz's theorem (see [Span], p. 148), the quotient of $\pi_1(X)$ by its commutator subgroup is isomorphic to $H_1(X;\mathbb{Z})$. In particular, if the elements $\gamma_1, \dots, \gamma_p$ of $\pi_1(X)$ generate a normal subgroup of finite index in $\pi_1(X)$, then their images in $H_1(X;\mathbb{R})$ generate all of $H_1(X;\mathbb{R})$. In other words, $b_1 = \dim(H_1(X;\mathbb{R})) \leq p$. However, property II) of the system γ_i implies that the balls $B^{\tilde{X}}(\gamma_i\tilde{x}, \varepsilon/2)$ are disjoint, while property I) implies that they are all contained in $B^{\tilde{X}}(\tilde{x}, 2d + 3\varepsilon/2)$. If X is a Riemannian manifold that satisfies the identity $\operatorname{Ricci}(g) \geq -(n-1)rg$, then the inequality of Theorem 5.3 gives

$$\begin{aligned} b_1 \leq p \quad &\leq \quad \frac{\operatorname{vol}(B(\tilde{x}, 2d + 3\varepsilon/2))}{\operatorname{vol}(B(\tilde{x}, \varepsilon/2))} \\ &\leq \quad \int_0^{5d} (\sinh(\sqrt{r}t))^{n-1}\, dt \left(\int_0^{d} (\sinh(\sqrt{r}t))^{n-1}\, dt \right)^{-1}, \end{aligned}$$

where ε is a function of n, r, d only.

5.21. Theorem: *There is an integer-valued function $\varphi(n, r, d)$ such that for each Riemannian n-manifold (V, g) of diameter d and Ricci curvature*

$\mathrm{Ricci}(g) \geq -(n-1)rg$, the first Betti number of V satisfies the inequality $b_1 \leq \varphi(n,r,d)$. Moreover, when rd^2 is sufficiently small, the function φ equals n.

Proof. The preceding corollary shows that the function

$$\varphi(n,r,d) = \sup b_1(V;\mathbb{R})$$

is finite provided that V is a Riemannian n-manifold with diameter d and $\mathrm{Ricci}(g) \geq -(n-1)rg$. It remains to prove the second assertion.

By the case $\varepsilon = 0$ of Lemma 5.19, there exist p elements of $\pi_1(V)$ and a point $\tilde{x}$ of $\tilde{V}$ such that $\|\gamma_i\|_{geo} \leq 2d$, and such that the subgroup Γ' generated by the γ_i has finite index in $\pi_1(V)$. Let h denote the natural homomorphism from $\pi_1(V)$ into $H_1(V;\mathbb{R})$; then $h(\Gamma')$ generates the vector space $H_1(V;\mathbb{R})$.

We first extract a subgroup $\Gamma'' \subset \Gamma'$ whose image under h still generates $H_1(V;\mathbb{R})$, all of whose elements have norm $\geq d$. To begin, we extract a basis of H_1 from the vectors $h(\gamma_i)$, which we denote by $h(\gamma_1),\dots,h(\gamma_k)$, and restrict our attention to the subgroup Γ' generated by $\gamma_1,\dots,\gamma_k$. Note that only finitely many elements of Γ' have norm less than d. If γ and each of its powers have norm less than d, then γ has finite order, and so $h(\gamma)=0$, i.e., $\gamma = 1$, since h is injective by construction of the family $\gamma_1,\dots,\gamma_k$. Consequently, there is a power of γ having norm greater than d, and thus an integer m such that $d \leq \|\gamma^m\|_{geo} < 2d$. Choose a generator γ_j such that $h(\gamma)$ has a nonzero j-th component in the basis $h(\gamma_1),\dots,h(\gamma_k)$, and replace γ_j with $\gamma_j' = \gamma^m$ in the generating set. The new group no longer contains γ (if it did, then a nontrivial linear combination of the $k-1$ vectors γ_i $i \neq j$ would be zero). After a finite number of operations, we obtain the desired group: it is free abelian with b_1 generators of norm less than $2d$, and all of its elements have geometric norm $\geq d$.

The number of points in the orbit $\Gamma''\tilde{x}$ lying within the ball $B^{\tilde{V}}(\tilde{x},2pd)$ is at least equal to the number of points $(\lambda_i) \in \mathbb{Z}^{b_1}$ such that $\sum_{i=1}^{b_1} |\lambda_i| \leq p$, i.e., approximately $(1/2)(2\pi)^{b_1}/b_1!$. By construction, however, the balls $B^{\tilde{V}}(\gamma\tilde{x},1/2d)$ are disjoint and lie within $B^{\tilde{V}}(\tilde{x},2pd)$, and their number is at most

$$\frac{\mathrm{vol}(B(x,2pd))}{\mathrm{vol}(B(x,1/2d))} \leq \int_0^{2pd\sqrt{r}} (\sinh(t))^{n-1}\,dt \left(\int_0^{1/2d\sqrt{r}} (\sinh(t))^{n-1}\,dt\right)^{-1},$$

which tends to $(4p)^n$ as $rd^2 \to 0$, implying that $b_1 \leq n$. More precisely, suppose that $b_1 > n$. If c is a real number < 1, then there exists p_0 such that the number of points of $\Gamma''\tilde{x}$ within $B(\tilde{x},2p_0d)$ will be $\geq c(2p_0)^{b_1}/b_1!$**

and $p_0^{b_1-n} > 4^n b_1! c^{3-2n} 2^{1-b_1}$. There exists a $\varepsilon > 0$ such that for $|t| \leq \varepsilon$, we have $c \leq (\sinh(t))/t \leq (1/c)$. So, for $rd^2 \leq \varepsilon^2/4p_0^2$, the ratio of the integrals is less than $c^{2-2n}(4p_0)^n$, which implies that $p_0^{b_1-n} \leq 4^n b_1! c^{3-2n} 2^{1-b_1}$, a contradiction. Finally, we note that the bound ε_n such that $rd^2 \leq \varepsilon_n$ implies $b_1 \leq n$ decreases exponentially with n.

Example: The n-torus satisfies $b_1 = n$ and admits metrics with zero Ricci curvature, namely all the flat metrics.

5.22. Remark: From [Gro]$_{\text{AFM}}$, we can deduce the following more precise result: If the diameter and the sectional curvature of V are sufficiently close to 0, and if $b_1 = n$, then a finite covering of V is diffeomorphic to a torus.

Conjecture: If $d\sqrt{r}$ is sufficiently small, and if $b_1 = n$ (or even if only $b_1 \geq n-1$), then V is homeomorphic to a torus. (For $b_1 = n$, this has been proved by T. Colding, and for $b_1 = n-1$ disproven by Anderson, see [And] and [Col]$_{\text{LMPR}}$.)

Our argument is specific for the first Betti number over the reals but not over a finite field. This suggests that the analytical methods of Bochner could yield the same result.

Bochner introduced, for each $k = 1, \ldots, n$, a quadratic form R_k on the exterior power $\Lambda^k TV$ such that R_1 coincides with the Ricci curvature. He showed that (see [Boch–Yano]), if R_k is everywhere positive, then

$$b_k = \dim(H^k(V;\mathbb{R})) \leq \frac{n!}{(n-k)!k!}.$$

For a Riemannian manifold such that $R_k \geq -r_k g^{(k)}$, where $g^{(k)}$ denotes the k-th exterior power of the metric g, the Bochner formula show that the number of eigenvalues $\leq \lambda$ of the Hodge Laplacian on k-forms is bounded by the corresponding number for the *Bochner Laplacian* shifted by r_k,

$$N(\Delta_{\text{Hodge}}, \lambda) \leq N(\Delta_{\text{Bochner}}, \lambda + r_k)$$

for all $\lambda \geq 0$. Since the latter number allows a bound in terms of the Laplacian on functions via Kato's inequality (see $3\frac{1}{2}$.62), our result in Appendix C_+ implies that for each k,

$$b_k = N(\Delta_{\text{Hodge}}, 0) \leq (d/\sqrt{r_k})^n c_n^{1+d\sqrt{r}} + \frac{n!}{(n-k)!k!}.$$

For other results following from the Bochner formula, see [Gal]$_{\text{IIA}}$.

D. Small loops

In the first section of this chapter, it became apparent that Riemannian manifolds with bounded (above) diameter and bounded (below) Ricci curvature fall into a finite number of collections of manifolds having closely related geometric properties. We then promised some results addressing the topology of these manifolds. This is the topic of the present section.

5.23. Definition: For any positive constant c, we denote by $\mathcal{G}_c$ the family of groups Γ of finite type satisfying the following properties:

1. Γ is torsion-free.

2. Each $\gamma \in \Gamma$ is contained in a unique maximal cyclic subgroup Z_γ, and $Z_\gamma \neq \Gamma$.

3. If γ and δ generate a noncyclic subgroup of Γ, then this subgroup has exponentially growth and entropy c with respect to the generating set $\{\gamma, \delta\}$ (see Definition 5.11).

Remark: By (2) and (3), each element of $\mathcal{G}_c$ has trivial center. If $c \leq \log(3)$, then the class $\mathcal{G}_c$ contains the free noncyclic groups and the fundamental groups of manifolds of negative curvature (see Remarks 6.10(2) and 6.12). This definition enables us to give a new version of *Margulis' lemma* (compare $[\text{Gro}]_{\text{AFM}}$, p. 240, and see Remark 8.50).

5.24. Theorem: *Let (V, g) be a compact Riemannian manifold such that* $\text{Ricci}(g) \geq -(n-1)rg$. *If the group $\pi_1(V)$ is in the family $\mathcal{G}_c$, then there exists a point $\tilde{v} \in \tilde{V}$ such that, for each nontrivial $\gamma \in \pi_1(V)$, we have* $\text{dist}(\tilde{v}, \gamma\tilde{v}) \geq c/\sqrt{r}$.

The proof relies on the following two lemmas.

5.25. Lemma: *Let Γ be a group satisfying properties (1) and (2) of Definition 5.23 above. Then each element of Γ has at most one square root.*

Proof. Elementary.

5.26. Lemma: *Let X be a compact, connected manifold. Let $\Gamma^{min(x)}$ be the subgroup of $\pi_1(X, x)$ generated by those elements whose distance to the identity is minimal (with respect to the geometric norm 3.20). In other words, $\Gamma^{min(x)}$ is generated by nontrivial geodesic loops based at x having minimal length. If π_1 satisfies properties (1) and (2), and if for each x the* *group $\Gamma^{min(x)}$ is cyclic, then for each x this group is contained in the center of $\pi_1(X, x)$.*

Proof. Since for each curve u joining the two points x, y and for each $\gamma \in \pi_1(X, x)$, we have

$$|\|u^{-1}\gamma u\|_y - \|x\|_x| \leq 2\,\mathrm{length}(u),$$

the discrete and closed set (cf. 1.13) of lengths of closed, nontrivial geodesics based at x varies continuously with x. If x, y are sufficiently close and joined by a minimizing geodesic u, there exists $\gamma \in \pi_1(X, x)$, based at a closed minimizing geodesic at x, such that $u^{-1}\gamma u$ yields a closed minimizing geodesic at y. In other words, $u^{-1}\Gamma^{min(x)}u \cap \Gamma^{min(y)} \neq \{e\}$, thus $u^{-1}Z^x u = Z^y$ for the maximal cyclic subgroups Z^x, Z^y corresponding to the generators of $\Gamma^{min(x)}, \Gamma^{min(y)}$.

Now let δ be an element of $\pi_1(X, x)$ represented by a closed geodesic loop based at x. By applying the argument above to smaller and smaller subdivisions of this loop, we obtain $\delta^{-1}Z^x\delta = Z^x$. If α is a generator of Z^x, then $\delta^{-1}\alpha\delta$ is also, thus $\delta^{-1}\alpha\delta = \alpha^{\pm 1}$, the group being torsion-free. But if $\delta^{-1}\alpha\delta = \alpha^{-1}$, then $\delta = \alpha\delta\alpha$, and $\delta^2 = (\alpha\delta)^2$, contrary to the preceding lemma. The inner automorphism $\gamma \mapsto \delta^{-1}\gamma\delta$ therefore leaves Z^x pointwise invariant.

Proof of the theorem. Since π_1 has trivial center, Lemma 5.26 guarantees the existence of a point $a \in V$ and two elements $\gamma, \delta \in \pi_1(V, a)$ represented by minimal geodesic loops of length ε based at a and generating a noncyclic group Γ.

Since the loops have the same length, we will easily be able to compute the number $N(p)$ of points in the universal cover $(\tilde{V}, \tilde{a})$ of (V, a) belonging to $B(\tilde{a}, p\varepsilon) \cap \Gamma\tilde{a}$.

There are at least as many elements of Γ that can be written as words in γ, δ and their inverses that have length less than or equal to p. By property (3) of π_1, we have $N(p) \geq e^{cp}$.

There are at most $\mathrm{vol}(B(\tilde{a}, p\varepsilon))/\mathrm{vol}(B(\tilde{a}, \varepsilon))$, and the hypothesis for the Ricci curvature enables us to apply Bishop's inequality (cf. 5.3.bis)

$$\frac{\mathrm{vol}(B^V(\tilde{a}, p\varepsilon))}{\mathrm{vol}(B^V(\tilde{a}, \varepsilon))} \leq \int_0^{p\varepsilon} (\sinh(\sqrt{r}t))^{n-1}\,dt \left(\int_0^{\varepsilon} (\sinh(\sqrt{r}t))^{n-1}\,dt\right)^{-1},$$

the ratio of the volumes of balls of radii $p\varepsilon$ and ε in hyperbolic space whose Ricci curvature satisfies $\mathrm{Ricci}(g) = -(n-1)rg$. As $p \to \infty$, the ratio is on the order of $\exp(\sqrt{rp\varepsilon})$. All of the above implies for large p inequalities of the form $e^{cp} \leq N(p) \leq f(n, r, \varepsilon)e^{\sqrt{rp\varepsilon}}$, which are possible only if $\varepsilon \geq c/\sqrt{r}$.

5.26.bis Remark: (*a*) By $[\mathrm{Gro}]_{\mathrm{GPG}}$, we can replace condition (3) of Definition 5.23 by

(3)′ Γ has no subgroup isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$.

The conclusion of Theorem 5.24 then becomes $d(\tilde{v}, \gamma\tilde{v}) \geq c_n/\sqrt{r}$ for a universal constant c_n (see 5.32).

(b_+) The proof above makes essential use of a continuity argument (following in the steps of the original approach by Kazhdan–Margulis, see [Kaz–Marg] and [Ragun]), since we produce a pair of independent short loops of equal length at some point in V by bringing together the shortest loops from different points. This idea extends to k-parametric families of loops for all $k \geq 1$ and yields $(k+1)$-tuples of "independent" loops at certain points in V under suitable topological assumptions (see 6.6.D in $[\mathrm{Gro}]_{\mathrm{FRM}}$ where such a result applies to a refined systolic problem in the spirit of 4.47).

5.27. Corollary: *The class $\mathcal{G}_c$ contains only a finite number of groups isomorphic to the fundamental group of a Riemannian n-manifold having diameter $\leq d$ and satisfying* $\mathrm{Ricci}(g) \geq -(n-1)rg$, *for fixed n, r, d.*

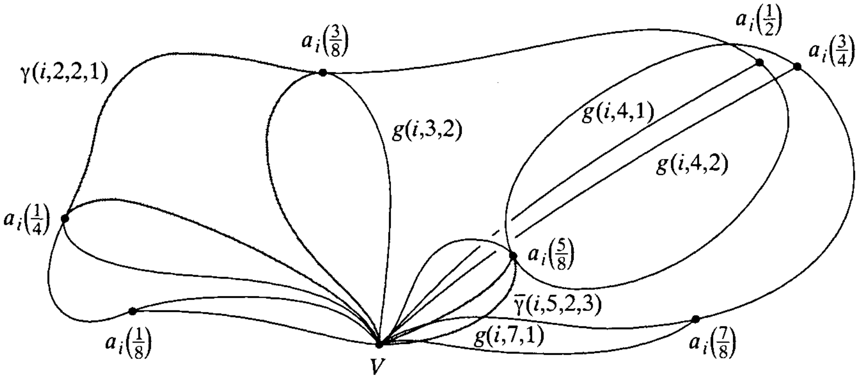
The corollary follows from the preceding theorem and the improved version of Proposition 3.22.

5.28. Proposition: *Let V be a Riemannian manifold of diameter d. Then for each $\tilde{v} \in \tilde{V}$, the group $\pi_1(V)$ admits a finite set of generators γ_i for which $\|\gamma_i\|_{geo} \leq 2d$ and such that all relations among the γ_i are of the form $\gamma_i\gamma_j\gamma_k^{-1} = 1$.*

Proof. As in the proof of Proposition 3.22, we choose a finite set of generators α_i of $\pi_1(V)$, and we represent each class α_i by a loop a_i based at v. There exists an $\varepsilon > 0$ such that the interval $(2, 2+\varepsilon)$ does not contain the geometric norm (relative to the point $\tilde{v} \in \tilde{V}$ of any element of $\pi_1(V)$, and there is an integer N such that $\mathrm{length}(a_i)/N < \varepsilon$ for each i. For $0 < k < N$, we connect each point $a_i(k/N)$ to v by all possible minimizing geodesics, which we denote by $g(i, k, j)$, $j \in J$, where the set of indices may be very large. Let $\tilde{\gamma}(i, k, j, j')$ be the homotopy class of the loop $g(i, j, k)g(i, k, j')^{-1}$ and $\gamma(i, k, j, j')$ is the homotopy class of the loop

$$g(i, k, j)a_i|_{[k/N, k+1/N]}g(i, j+1, j')^{-1}.$$

By construction, $\|\gamma\|_{geo} \leq 2d + \varepsilon$, thus $\leq 2d$ since ε is arbitrary. For each i, we can write $\alpha_i = \prod_{k=0}^{n-1} \gamma(i, k, j_k, j_{k+1})$ for every sequence j_k such that the family $\tilde{\gamma}, \gamma$ generates $\pi_1(V)$.



The relations that can exist between the $\tilde{\gamma}, \gamma$ are, on the one hand, all those which result from the numerous ways of writing the α_i, i.e., different choices of the sequence $k \mapsto j_k$. This reduces easily to the case in which the two sequences j_k and j_k only differ in the h-th term. Thus, the relation

$$\prod_{k=0}^{n-1} \gamma(i,k,j_k,j_{k+1}) = \prod_{k=0}^{n-1} \gamma(i,k,j'_k,j'_{k+1})$$

amounts to

$$\gamma(i,h-1,j_{h-1},j_h)\gamma(i,h,j_h,j_{h+1}) = \gamma'(i,h-1,j_{h-1},j'_h)\gamma'(i,h,j'_h,j_{h+1}),$$

which reduces to the two identities

$$\gamma(i,h,j'_h,j_{h+1}) = \tilde{\gamma}(i,h,j'_h,j_h)\gamma(i,h,j_h,j_{h+1}),$$

$$\gamma(i,h-1,j_{h-1},j_h)\tilde{\gamma}(i,h,j_h,j'_h) = \gamma(i,h-1,j_{h-1},j'_h).$$

Other relations arise from the relations among the α_i.

Let $\prod_{\ell=1}^{p} \alpha_{i_\ell}^{e_\ell} = 1$ be one such relation. The loop

$$a_{i_1} \overset{e_1\ times}{\cdots} a_{i_1}\, a_{i_2} \overset{e_2\ times}{\cdots} a_{i_2} \cdots a_{i_p}$$

is divided into segments of length less than ε, and the two expressions

$$\prod_{\ell=1}^{p} \prod_{k=0}^{N-1} \gamma(i_\ell,k,j_k^\ell,j_{k+1}^\ell)$$

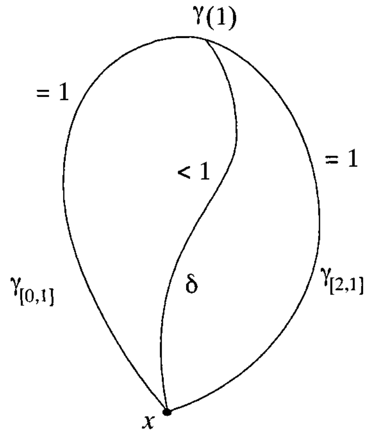
and 1 for the class $\prod_{\ell=1}^{p} \alpha_{i_\ell}^{e_\ell}$ are amenable to above-mentioned treatment. We have therefore shown that the group $\pi_1(V)$ admits a presentation with generators of norm less than $2d$ and relations of the form $\gamma_i\gamma_j\gamma_k^{-1} = 1$.

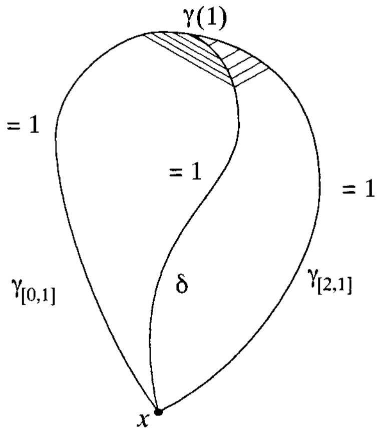
5.29. Proof of Corollary 5.27. Suppose that $\Gamma \in \mathcal{G}_c$ is isomorphic to the fundamental group of a Riemannian manifold (V, g) having diameter $\leq d$ and $\mathrm{Ricci}(g) \leq -(n-1)rg$. Then there exists a point $\tilde{v} \in \tilde{V}$ such that $\|\cdot\|_{geo}$ does not assume values less than $c/\sqrt{r}$ on $\Gamma \setminus \{1\}$. If $\gamma_1, \ldots, \gamma_p$ is the set of generators of Γ provided by the proposition, then the balls $B^V(\gamma_i\tilde{a}, c/2\sqrt{r})$ are disjoint in $B^V(\tilde{v}, 2d + c/2\sqrt{r})$; thus their number is bounded by a function $\psi(n, r, d)$. The relations among the γ_i are of the form $\gamma_i\gamma_j\gamma_k^{-1} = 1$, where the triplet (i, j, k) denotes a subset of R_Γ of $\{1, \ldots, p\}^3$. Since the group Γ is determined up to isomorphism by the number p and the subset R_Γ, the number of possible groups is thus $\psi(n, r, d)2^{\psi(n,r,d)^2}$.

Remark: It is tempting to ask whether, in the conclusion of Proposition 5.28, one can replace "≤ 2" by "< 2". The answer is that it can always be done, with the sole exception of the real projective space equipped with its canonical metric. More precisely,

5.30. Proposition: *Given an arbitrary point x in a Riemannian manifold (V, g) of diameter 1, we can always choose generators γ_i in Proposition 5.28 represented by loops of length < 2, except when V is diffeomorphic to $\mathbb{R}P^n$ and all geodesics based at x are simple, periodic, and have smallest period equal to 2. If, moreover, $(V, g) \neq (\mathbb{R}P^n, can)$, then there is always at least one $x \in V$ and generators γ_i of $\pi_1(V)$ that are representable by loops of length < 2.*

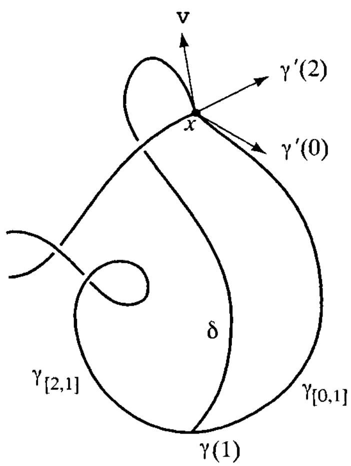
Proof. The second assertion follows from the first, combined with Corollary D.2 of [Besse], p. 236.





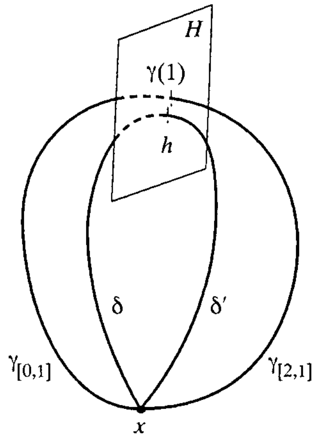
Fix $x \in V$ and a nontrivial loop γ based at x of length 2 and having minimal length in its homotopy class. We may assume that $\gamma(1)$ lies at distance 1 from x. Otherwise, there exists a segment δ joining x to $\gamma(1)$ with length strictly less than 1, and so γ can be represented by the product of $\gamma|_{[0,1]} \cup \delta$ and $\gamma|_{[1,2]} \cup \delta$, both of which have length less than 2. Similarly, there is no segment δ from x to $\gamma(1)$ other than $\gamma|_{[0,1]}$ and $\gamma|_{[2,1]}$, since then γ could be represented by the product of $\gamma|_{[0,1]} \cup \delta$ and $\gamma|_{[1,2]} \cup \delta$. These latter loops have length exactly equal to 2, but since they approach $\gamma(1)$ at an angle other than π, they can be shortened by homotopy.

From this proof and from the equality $d(x, \gamma(1)) = 1 = \operatorname{diam}(V)$, we conclude that the geodesic γ is periodic, i.e., the angle at x between $\gamma(0)$ and $\gamma(2)$ is zero. Indeed, if this were not the case, then by an application of Lemma 6.2 of [Che–Ebin] to the points $p = \gamma(1), q = x$ and to the vector $v = \gamma(2) - \gamma(0) \in T_v V$, we would obtain a segment γ from x to $\gamma(1)$ distinct from $\gamma_{[0,1]}$ and $\gamma_{[2,1]}$.



We will now show that all geodesics based at x are of the same nature as γ. For this purpose, we first note that $\gamma(1)$ is not conjugate to x on $\gamma_{[0,1]}$, nor on $\gamma_{[2,1]}$, since there would be a curve near γ having strictly smaller length.

Now choose a hypersurface H passing through $\gamma(1)$ and transverse to γ. Then H will intersect each geodesic based at an x that is sufficiently close to $\gamma_{[0,1]}$ or $\gamma_{[2,1]}$. Thus, for each point $h \in H$ sufficiently close to $\gamma(1)$, there is a segment δ near $\gamma_{[0,1]}$ passing from x to h, and a segment δ' near $\gamma_{[0,1]}$ passing from x to h. The union $\delta \cup \delta'$ is a loop based at x, homotopic to γ, and of length ≤ 2, since $\operatorname{diam}(V) \leq 1$. By the hypothesis on γ, it follows that $\delta \cup \delta'$ is a geodesic loop based at x. Thus, all geodesics sufficiently close to γ possess the same initial properties as γ and are therefore simple, closed geodesics having (smallest) period 2. These properties are stable under



closure, and the assertion of the theorem follows, since the unit tangent sphere at x is connected.

That V is necessarily diffeomorphic to $\mathbb{R}P^n$ follows from the last assertion of Theorem 5.3 of [Besse], p. 186–187.

E_+ Applications of the packing inequalities

5.31. The essential property of the manifolds (V, g) with Ricci curvature bounded from below used in the topological arguments above boils down to the following *packing inequalities*. Say that a metric space V has *packing type* P for a given $\mathbb{Z}_+$-valued function $P = P(R, r)$ if every ball of radius R in V contains at most P balls of radius r. (Compare with the *doubling property* in Appendix B_+.)

Now, the Bishop inequality provides a specific bound on P in terms of $\inf \mathrm{Ricci}(g)$. In particular, *if* $\mathrm{Ricci}(g) \le -(n-1)g$, *then the packing type* P *of* V *satisfies*

$$P(R, r) \le C(R/r)^m \tag{$*$}$$

for $R \le 1$ and $0 \le r \le R$, as well as

$$P(R, r) \le e^{\alpha R} \tag{$**$}$$

for $R \ge 1$ and $r \ge 1/2$, and where m happens to equal $n = \dim(V)$, where C and $\alpha = \alpha_n$ are universal constants. (In fact, $P(R, r) \le C_n(R/r)^n e^{\alpha_n R}$ for all positive R and r as we know very well by now.)

Exercises: (a) Check that the proofs of 5.3, 5.17, 5.18, 5.21, and 5.24 depend entirely on (the properties of) the packing types of the manifolds in question and of their covering spaces.

(*b*) Show that the packing bound (type) is implied by a bound on the ratio $\mu(B(R+r))/\mu(B(r))$, where μ is an arbitrary measure and the $B's$ are *concentric* balls. Then prove that a bound on the packing type P of V yields a similar bound on the (optimal) *covering numbers* of R-balls by smaller r-balls.

(*c*) **Questions:** Observe that the packing and covering functions of (balls in) metric spaces satisfy certain nontrivial relations. For example, if one can cover $B(R_3)$ by N_3 R_2-balls and each $B(R_2)$ by N_2 R_1-balls, then $B(R_3)$ is covered by N_2N_3 R_1-balls. For certain pairs (R,r), this imposes some relations among the values of the function $\mathrm{Cov}(V;R,r)$ whose value equals the infimum of the number of r-balls needed to cover each R-ball in V. Similarly, one defines the packing function $\mathrm{Pa}(V;R,r)$ and asks when a given function $C(R,r)$ (and $P(R,r)$) can be realized as Cov (respectively, Pa) for some metric space V.

Next, if we have a measure μ on V, we can define

$$\sup \mu = \sup_{v\in V} \mu(B(v,R)) \qquad\qquad \inf \mu = \inf_{v\in V} \mu(B(v,R)),$$

and

$$\sup \mu(V;R/r) = \sup_{v\in V} \frac{\mu(B(v,R))}{\mu(B(v,r))}.$$

Again, one wishes to know the possible restrictions on these functions. For example, what happens if $\sup \mu(V;R) = \inf \mu(V;R) = \alpha(R)$ for a given function $\alpha(R)$? (A special case of this problem is treated in $[\mathrm{Gro}]_{\mathrm{GPG}}$ under the heading of "regular growth.")

$\mathbf{F}_+$ On the nilpotency of π_1

Let X be a path metric space of packing type satisfying $(*)$ with some m (which is allowed here to be different from $\dim(X)$) and $C>0$. Consider some isometries $\gamma_1,\dots,\gamma_q$ on X, suppose they generate a *discrete* subgroup Γ in the full isometry group of X, and set

$$\delta_+ = \max_{1\le i\le q} \mathrm{dist}(x_0,\gamma_i(x_0))$$

for a given point $x_0\in X$.

Question: Does there exist a positive $\varepsilon=\varepsilon(m,C)>0$ such that the inequality $\delta_+<\varepsilon$ would imply that Γ is virtually nilpotent? If not, is Γ at least amenable? And what extra assumptions (if any) does one need to ensure the nilpotency (or amenability) of Γ?

The positive answer can be achieved with the aid of a *lower* bound on the *minimal* displacement

$$\delta = \inf_{\delta \in \Gamma \setminus \mathrm{id}} \mathrm{dist}(x_0, \gamma(x_0))$$

as follows.

Let $\lambda = \delta_+/\delta$. Then there exists a constant $\varepsilon = \varepsilon(m, C, \lambda) > 0$ such that the inequality $\delta_+ \leq \varepsilon$ makes Γ virtually nilpotent.

This is an easy corollary of the polynomial growth theorem (see §3.4 in $[\mathrm{Gro}]_{\mathrm{VBC}}$, the last section in $[\mathrm{Gro}]_{\mathrm{GPG}}$, and 6.6.B in $[\mathrm{Gro}]_{\mathrm{FRM}}$, where one should note that the definition of δ given in $[\mathrm{Gro}]_{\mathrm{FRM}}$ is incorrect).

A more difficult result is due to Cheeger and Colding (see [Che–Col]), who gave a positive answer to our question for Riemannian manifolds with Ricci $\leq -(n-1)$ and with no undesirable δ, just for $\delta_+ \leq \varepsilon$ with $\varepsilon = \varepsilon(\dim(X)) > 0$. (They assume that the action of Γ on X is free and cocompact, but this probably can be easily mended.)

The nilpotency of the group Γ allows a lower bound on the volume of V/Π for some (nonnilpotent) groups Π. For example, *if every virtually nilpotent subgroup $\Gamma \subset \Pi$ is contained in a unique maximal, virtually nilpotent subgroup $\Gamma' \subset \Gamma$, then*

$$\mathrm{vol}(X/\Pi) \geq \varepsilon' = \varepsilon'(m, C, \dim(X)) > 0,$$

provided that Π is **not** *virtually nilpotent, the action is free, and the classifying map $X/\Pi \to K(\Pi; 1)$ is nonhomologous to zero.* (See 6.6.C' in $[\mathrm{Gro}]_{\mathrm{FRM}}$, where it is stated for Ricci $\geq -(n-1)$.)

Exercise: Generalize the last claim in 6.6.D' of $[\mathrm{Gro}]_{\mathrm{FRM}}$ by replacing the lower Ricci bound by a packing bound.

5.32. Coverings and mappings to nerves. Pack a (locally) compact metric space V by a maximal system of R-balls $B_i(R)$ and then take the concentric $3R$-balls. These cover V with good overlaps, which trivially implies the existence of a partition of unity φ_i on V, all of whose functions φ_i (with supports in the $B_i(3R)$) are λ-Lipschitz with

$$\lambda \leq 3\,\mathrm{multi}/R,$$

where "multi" denotes the multiplicity of the cover. In particular, if the packing type $P(6R, R)$ is under control, i.e., it is bounded by a given constant N_0, then so is our multi. Thus, we get a Λ Lipschitz map Φ from V to the nerve Q of our cover, where $\Lambda = \sqrt{N_0}\lambda$, and where the metric we use in (the simplicial polyhedron) Q is piecewise Euclidean , built of the k-simplices

$$\{x_i > 0 : \sum_{i=0}^{k} x_i = 1\} \subset \mathbb{R}^{k+1} \subset \mathbb{R}^{N_0+1}$$

(compare pp. 85, 86 in [Gro]$_{\mathrm{VBC}}$).

We want to use the maps $\Phi\colon V \to Q$ to bound the topology of V. To do this, we need control over the contractibility function in V (see Ch. 3.E$_+$) in order to approximately reverse the map Φ, say by a $\Psi : Q \to V$ such that $\Psi \circ \Phi\colon V \to V$ is homotopic to the identity. If this is the case, and if V is compact, then we can bound, for instance, the Betti numbers of V by the number of simplices in Q. In fact, we can do better, since we need to take into account not all of Q, but only the image $\Phi(X) \subset Q$. Since Φ is Λ-Lipschitz, this image is at most Λ times greater than X. For example, if X is a Riemannian manifold, then the map Φ can be deformed to some Φ' which lands in the n-skeleton $Q_n \subset Q$ and which is Λ'-Lipschitz with $\Lambda' < N_0^{N_0}\Lambda$ by a (very) rough estimate. Then the volume of the image does not exceed $(\Lambda')^n \operatorname{vol}(X)$, and so we can push $\Phi'(X) \subset Q_n$ further to a subpolyhedron $Q'_n \subset Q_n$ consisting of at most N simplices with $N \leq n^n(\Lambda')^n \operatorname{vol}(X)$, thus bounding the Betti numbers of X by this N.

There are several possibilities for improvement:

I. Let all of V or some part V_0 of V be "thin" (or collapsed), which can be expressed by requiring smallness of some metric invariants of V or of V_0. For example,

(*i*) $\operatorname{FilRad} V_0 \leq \varepsilon$,

(*ii*) $\operatorname{wid}_m V_0 \leq \varepsilon$, where wid_m denotes the Uryson width of V_0 for some $m \leq n$.

(*iii*) $\operatorname{vol} B(R) \leq \varepsilon(R/\Lambda)^n$ for all R-balls in V.

(*iv*) each $B(R)$ can be covered by at most $\varepsilon(R/r\Lambda)^n$ balls of radius r for a given r which is significantly smaller than R.

Then we can push our map $\Phi' : V \to Q'_n$ to a smaller subpolyhedron containing few n-simplices (or m-simplices in the case of (ii)), thus getting a better bound on the topology (e.g., Betti numbers) of V.

II. It may happen that the balls $B_i(R)$ are noncontractible (in larger concentric balls) within V, but there is a map $f\colon V \to K$ into some space K where these $B_i(R)$ become contractible. This would allow the homotopy factoring of f to be

$$f = f' \circ \Phi$$

for some map $f'\colon Q \to K$. As a consequence, we could bound the topological invariants of f, such as the ranks of the homomorphisms $f_*\colon H_*(V) \to H_*(K)$.

III. It may happen that the (controlled) contractibility of balls in V (or in K receiving V) can be gained by passing to some covering of V. For example, if we start with a compact aspherical space V, then we do have a controlled contractibility of balls in the universal covering $X = \tilde{V}$ as well as in the finite coverings $\tilde{V}_i$ *converging* to $\tilde{V}$. (The existence of such $\tilde{V}_i$ is equivalent to the *residual finiteness* of $\pi_1(V)$, i.e., the existence of subgroups $\Phi_i \subset \pi_1(V)$ of finite index with $\bigcap_i \Phi_i = \{\mathrm{id}\}$). As we pass to such coverings, we need to *assume* packing inequalities in these (which follow, for example, from a lower bound on Ricci(V) and instead of the original invariants of V, we shall come up with a bound on those of $\tilde{V}$ or of $\tilde{V}_i$.

IV. Combine I-III.

Example: Suppose that V is an aspherical n-dimensional manifold with a residually finite fundamental group, where the packing type P of the universal covering $X = \tilde{V}$ is under control. Then the above allows a bound on the Betti numbers of $\tilde{V}_i$ in terms of $\operatorname{vol}\tilde{V}_i$. In particular, *one can thus bound the Euler characteristic and the signature of* $\tilde{V}_i$, and since these invariants are multiplicative under coverings, *one recaptures a bound on* $\chi(V)$ *and* $\sigma(V)$ *by* vol(V). The details can be found in [Gro]$_{\mathrm{VBH}}$, where this is proved in the framework of II.

Our next move is to drop the residual finiteness condition. This will require a more sophisticated formalism which we do not expect a beginner to be ready to swallow (at least not in one gulp).

5.33. Bounds on L_2-Betti numbers. Let us briefly recall the definitions in the special case of a polyhedron X with a discrete automorphism group Γ acting on X (where an important example is the Galois action on X covering some compact V). First we define the *ℓ_2-cohomology* of X (which needs no action) by just restricting to ℓ_2-cochains on X, i.e., those functions on the set of oriented simplices (or cells) in X which are square-summable. If X is uniformly locally finite, i.e., each k-cell has at most N_k neighbors for some $N_k = N_k(X)$, then the coboundary operator on ℓ_2 cochains $d_k : \ell_2 C^k \to \ell_2 C^{k+1}$ is ℓ_2-bounded, and then one defines *non-reduced* ℓ_2-cohomology as

$$\ell_2 H^k(X) = \ker d_k / d_{k-1}(\ell_2 C^{k-1}).$$

Of course, it gives you just ordinary cohomology for *finite* polyhedra X, but if X is *infinite*, then this reflects the rate of decay of cochains in question.

Also notice that the image $d_{k-1}(\ell_2 C^k) \subset \ell_2 C^k$ is not necessarily closed, which means that the quotient space (ℓ_2-cohomology) is not a separable space. This is not at all a pathology; yet, sometimes one feels better without it and defines the *reduced* ℓ_2-cohomology as

$$\overline{\ell_2 H}^k(X) = \ker d_k / (\ell_2\text{-closure of } d_{k-1}(\ell_2 C^{k-1})),$$

which is an honest Hilbert space.

At first sight, there is nothing especially interesting about these definitions. True, we get a bi-Lipschitz invariant of X (this is rather obvious as well as a stronger homotopy invariance in the Lipschitz category), but it carries little information, since all separable infinite-dimensional Hilbert spaces are mutually isometric (and for this reason the non reduced cohomology appears more promising as well as the ℓ_p- generalizations; see [Gro]$_{\mathrm{AI}}$ and the references therein).

But a miracle happens when there is a group Γ acting on X. This does not change the spaces $\overline{\ell_2 H}^k$ but adds *unitary* Γ*-actions* to them. Now, an infinite-dimensional Hilbert space H with a unitary Γ-action can be assigned a certain number, called the *Von-Neumann* Γ *dimension*, which can be finite since it measure the size of H/Γ in a certain sense. In particular, one can define the ℓ_2-Betti number

$$h^k(X : \Gamma) = \dim_\Gamma \overline{\Gamma H}^k(X),$$

which is finite for cocompact actions (i.e., if X/Γ is compact). Here is a direct

Definition of $h^k(X : \Gamma)$: Let $d = \oplus_k d_k : \ell_2 C^* \to \ell_2 C^*$, take the adjoint operator d^*, and let $\Delta = dd^* + d^*d$. Denote by $\ell_2\mathcal{H}^k \subset \ell_2 C^k$ the space of *harmonic* ℓ_2-cochains of degree k, i.e., those in $\ker \Delta$, and let P^k denote the normal projections of $\ell_2 C^k$ to $\mathcal{H}^k$. The projection operator, being ℓ_2-bounded, is given by a "kernel" which is a function $\Pi(b, b')$ on the pairs of oriented k-simplices (cells) in X, such that

$$(P^k(c(b)))(b') = \sum_b \Pi^k(b, b')c(b),$$

where Π^k is skew-symmetric for the change of orientation in the simplices (cells) b and b'. Finally, we bring in Γ by noting that the diagonal function $\mathrm{tr}_k(b) = \Pi^k(b, b')$ is Γ-invariant, and so we can set

$$\mathrm{Tr}_\Gamma P^k = \sum_{\underline{b}} (\mathrm{card}\,\Gamma_{\underline{b}})^{-1}\, \mathrm{tr}_k(b),$$

where the summation is taken over all Γ-orbits $\underline{b}$ of cells b in X, and where $\Gamma_{\underline{b}} \subset \Gamma$ denotes the stabilizer (subgroup) of b. (Notice that $\operatorname{card}\Gamma_{\underline{b}} = 1$ for free actions.)

These ℓ_2-Betti numbers are clearly finite for cocompact actions, and besides they satisfy the following (easy to prove) properties, which bring them close to the ordinary Betti numbers of X/Γ and especially close to $i^{-1} \operatorname{rank} H^i(X/\Gamma_i)$ for subgroups $\Gamma_i \subset \Gamma$ of large index i (preferably going to infinity, such that $\bigcap_i \Gamma_i = \{\mathrm{id}\}$).

(*a*) $h^k(X : \Gamma) = 0 \Leftrightarrow \overline{\ell_2 H}^k(X) = 0$.

(*b*) If X is contractible and X/Γ is compact, then the $h^k(X : \Gamma)$ do not depend on X, and moreover, the vanishing (or nonvanishing) of $h^k(\Gamma) \overset{\text{def}}{=} h^k(X : \Gamma)$ is a quasi-isometry invariant of Γ.

(*c*) $h^k(\Gamma_1 \times \Gamma_2) = \sum_{i=0}^k h^i(\Gamma_1)h^{k-i}(\Gamma_2)$.

(*d*) The *ℓ_2-Euler characteristic* $\chi(X : \Gamma) = \sum_i (-1)^i h^i(X : \Gamma)$ equals $\chi(X/\Gamma)$ if the action of Γ is cocompact and free.

(*e*) If $\Gamma' \subset \Gamma$ is a subgroup of finite index, then

$$h^k(X : \Gamma') = \operatorname{card}(\Gamma/\Gamma')h^k(X : \Gamma).$$

We suggest that the reader try to prove these properties and consult $[\mathrm{Gro}]_{\mathrm{AI}}$ and [Lück] for further information and references.

Now we want to explain the following

ℓ_2-bound. *Let V be a closed, aspherical Riemannian n-manifold where the packing type of the universal cover satisfies $P(b, 1) \leq N_0$ (e.g.,* $\operatorname{Ricci} V \geq -\delta_0$*). Then the ℓ_2-Betti numbers of the universal covering $X = \tilde{V}$ acted upon by $\Gamma = \pi_1(V)$ satisfy*

$$\sum_{k=0}^{n} h^k(X : \Gamma) \leq C \operatorname{vol}(V)$$

for some $C = C(N_0, n)$. In particular, the Euler characteristic and the signature of V are bounded by $C \operatorname{vol}(V)$ (where the signature is bounded by $h^{n/2}(X : \Gamma)$ according to the L_2-index theorem by Atiyah).

Idea of the proof. We can cover $X = \tilde{V}$ by balls of radius 3 with multi $\leq N_0$, but such a covering cannot, in general, be made Γ-periodic. But according to the philosophy of A. Connes, "periodic" can be relaxed to "quasi-periodic" without changing the outcome of the ℓ_2-computation of Betti numbers. This is done by taking some probability space M with a free, measure preserving action of Γ and then constructing a Γ-*invariant* covering of $X \times M$ for the diagonal action of Γ on $X \times M$ by subsets of the form $B \times M'$, where $B = B(3) \subset X$ are balls of radii 3 in X, and $M' \subset M$ are measurable subsets. The point is that we can now make such a covering with multiplicity $\leq N_0$ using the extra freedom given by the action on M. To see how it works, just look at $X = \mathbb{R}$ acted upon by $\mathbb{Z}$ in the usual way. Every equivariant covering here by 3-balls has multiplicity at least 6. But, if we twist it with an irrational action (rotation) of $\mathbb{Z}$ on S^1, then $\mathbb{R} \times S^1$ can be equivariantly covered by $B(3) \times M'$ with multiplicity at most 2, as a trivial argument shows.

Next, the nerve $\tilde{Q}$ of the covering of $X \times M$ by $B \times M'$ can be used, following Connes, to define the ℓ_2-cohomology and to bound the ℓ_2-Betti numbers in terms of the properly counted "number" of simplices in $\tilde{Q}$. Namely, we consider the Γ-orbits $\underline{\Sigma}$ of our simplices Σ in $\tilde{Q}$, where each Σ corresponds to some subsets $B_i \times M_i'$, $i = 0, \ldots, k$ meeting at some point. Every such Σ carries a weight, namely $\mu\left(\bigcap_{i=0}^{k} M_i'\right)$, and the "number" of k-simplices in $\tilde{Q} : \Gamma$ is understood as the sum of those weights over all Γ-orbits $\underline{\Sigma}$ in $X \times M$. This is an easy extension (made in a much more general framework by Connes) of the ℓ_2-discussion above.

Exercise: Carry out the details in the simplest case of a *trivial* group and then for a nontrivial *finite* Γ, where the von Neumann dimension satisfies

$$\dim_\Gamma = \operatorname{rank} H / \operatorname{card} \Gamma.$$

What remains is to bound the "number" of all simplices in $\tilde{Q} : \Gamma$ by $C \operatorname{vol} V$. Our earlier argument provides such a bound on the number of the *n-simplices*, and thus on the number of their k-faces. But the remaining k-simplices, $k < n$, which are not faces of n-simplices, do not contribute, due to Poincaré duality for ℓ_2-cohomology. The effect of this is similar to what happens to ordinary Betti numbers: if we have maps $\Phi : V \to Q$ and $\Psi : Q \to V$, where $\Phi \circ \Psi : V \to V$ is homotopic to the identity, then the Betti numbers of V in all dimensions are bounded by the number of the n-simplices in the image of Ψ times 2^n, provided that $\dim Q \leq n$ (as we can assume in our case). Also notice that the above fails to be true for general nonsimplicial cell complexes.

Exercises. (a) Generalize the above in the spirit of II in (5). Namely, consider a (non-aspherical) space V mapped to $K(\Gamma, 1)$, and bound the von Neumann rank of the (bilinear) cup-pairing on $\ell_2 H_*(\Gamma)$ associated to each homology class in $H_*(\Gamma;\mathbb{R})$ coming from $H_*(V;\mathbb{R})$. (Notice that the cup product of two ℓ_2-cohomology cochains on X is ℓ_1 on X and can be evaluated on each chain in X/Γ equivariantly lifted to X.)

(*b*) Extend all of the above to non-Riemannian metric spaces V where the n-volume is defined as the supremum of the n-dimensional Hausdorff measures of the 1-Lipschitz images of V in Euclidean spaces (compare vol in Ch. 4.C).

Questions. How should one modify the above estimates in order to make them (at least roughly) sharp? Can one make the ℓ_2-bounds above with the entropy of V instead of $P(6,1)$ of the universal covering? Can one bound the first ℓ_2-Betti number of a group Γ in terms of the entropy of Γ?

$\mathbf{G}_+$ Simplicial volume and entropy

5.34. Let us assign a *norm* to the real homologies of topological spaces. This means that we make a *functorial* assignment of a norm to the vector space $H_k(X;\mathbb{R})$ for *every* topological space X, where H_k refers to the *singular* homology of X, and where *functorial* means that our norm *decreases* under induced homomorphisms $f_* : H_k(X) \to H_k(Y)$ for all continuous maps $X \to Y$.

Examples. (*a*) Suppose that X admits a self-mapping $f : X \to X$ sending some $h \in H_k(X)$ to, say $2h$. Then, obviously, the norm of h must be zero. In particular, our (functorial) norm vanishes on the homology of the n-torus $H_k(\mathbb{T}^n)$ for $k > 0$. Consequently, it vanishes on the fundamental class of the sphere S^n, since this comes from $\mathbb{T}^n$ by some map $\mathbb{T}^n \to S^n$ (of degree 1). Furthermore, there is no nonzero norm on $H_1(X)$, since every homology class comes from $S^1 = \mathbb{T}^1$.

(*b*) Now look at $H_2(X)$ and observe that every (integral rather than real) 2-dimensional class (cycle) can be represented by a map of a surface S into X. If the genus of S is zero or one, this surface has zero norm and can be discarded. On the other hand, surfaces of genus ≥ 2 do admit a nonzero norm by the following trivial

5.35. Mapping lemma: *Let $f : S \to S'$ be a map of degree d between connected, oriented surfaces of negative Euler characteristics. Then*

$$|\chi(S)| \geq d|\chi(S')|.$$

Thus, $|\chi(S)|$ serves as a norm in the category of surfaces, and it extends to all X as the maximal norm compatible with the assignments above for surfaces. Namely, for a given $h \in H_2(X;\mathbb{Z})$, we consider all surfaces S and maps $f\colon S \to X$ with $f_*[S] = mh$ and define

$$\operatorname{norm}(h) = \inf_{m,S,f} m^{-1} \operatorname{norm}[S],$$

where the infimum runs over all $m = 1, 2, \ldots,$ all connected surfaces of genus ≥ 2, and all f above. Clearly, this makes a norm on $H_2(X;\mathbb{Z}) \otimes \mathbb{R}$ (which is somewhat different from $H_2(X;\mathbb{R})$ if H_2 is not finitely generated) with our starting $\operatorname{norm}[S] = |\chi(S)|$.

Exercise. Describe *all* possible functorial norms on H_2 in the category of surfaces.

Remarks: (*a*) The interpretation of $|\chi(S)|$ as a norm originates from a work by Thurston, who used this idea to define a norm on $H_2(X^3)$ using surfaces *embedded* into 3-manifolds.

(*b*) We shall meet below a wild variety of norms, all directly or indirectly associated to the fundamental group $\pi_1(X)$ (compare (4)(A) below). It is probable that every natural norm vanishes on $H_*(X)$ for all simply connected X.

5.36. Norm on $[S_1 \times S_2 \times \cdots \times S_k]$. Now it is tempting to define the norm on (the fundamental classes) of products of surfaces with $\chi < 0$ by

$$\|[S_1 \times S_2 \times \cdots \times S_l]\|_\chi = |\chi(S_1 \times S_2 \times \cdots \times S_l)|,$$

and then to extend it to $H_k(X)$, $k = 2l$, for all X by the above reasoning. Indeed, this is possible due to

Another mapping lemma: *Let*

$$f\colon S_1 \times S_2 \times \cdots \times S_l \to S_1' \times S_2' \times \cdots \times S_l'$$

be a map of degree d. Then

$$|\chi(S_1 \times S_2 \times \cdots \times S_l)| \geq d|\chi(S_1' \times S_2' \times \cdots \times S_l')|.$$

There are at least three different proofs of this lemma. One is based on the Lusztig-Meyer signature theorem (see $[\mathrm{Gro}]_{\mathrm{PCMD}}$); the second proof (which was probably the first chronologically) appeals to the proportionality principle for the simplicial volume (see below), and the third one (which is a variation of the first) was more recently suggested by Besson, Courtois, and Gallot. (See $[\mathrm{Pan}]_{\mathrm{VCE}}$ for an exposition of their work. Actually, I am not certain that their argument was carried over for products of surfaces.)

So, one can feel comfortable with $S_1 \times S_2 \times \cdots \times S_l$, but, as for general X, it is unclear which $2l$-dimensional classes in $H_{2l}(X)$ come from (mapped) products of surfaces (if there is not enough maps $S_1 \times S_2 \times \cdots \times S_l \to X$, we end up with an *infinite* norm on $H_{2l}(X)$). Probably most (interesting) $2l$-classes do *not* arise in this way (e.g., one suspects that the fundamental classes of irreducible, locally symmetric spaces admit no such maps), but *not a single* example has been worked out. In any case, one can be relieved with the following definition, which gives a finite (possibly zero) norm to every class.

5.37. Simplicial norm. Recall that every real singular k chain in X is a formal real combination of singular simplices

$$c = \sum_{i=1}^{p} r_i \sigma_i,$$

for $\sigma_i : \Sigma^k \to X$. Then $\sum_{i=1}^{p} |r_i|$ gives us a norm on the chains. This norm restricts to cycles and then passes to the quotient norm on homology classes. In other words, the simplicial norm $\|h\|_\Delta$ for an $h \in H_k(X)$ is the minimal "number" of simplices needed to represent h by a singular cycle. For example, if h is the fundamental class of an oriented manifold V, then $\|h\|_\Delta = \|[V]\|_\Delta$ is obviously majorized by the number of the top-dimensional simplices needed to triangulate V. In fact, one can think of $\|[V]\|_\Delta$ as the infimum of the number of simplices over all ideal (homotopy) triangulations of V, where one allows non-embedded simplices. This is especially transparent if V is a Riemannian flat (or almost flat) manifold, so that one can speak of standard Euclidean (or almost Euclidean) simplices isometrically *immersed* (not necessarily embedded!) into V and fitting into a cycle homologous to $[V]$. The simplest instance of this is given by a triangulation of a d-sheeted covering $\tilde{V}_d \to V$, where $[V]$ is represented by the simplices of some triangulation of $\tilde{V}_d$ with weights $1/d$. In fact, this procedure leads to a sharp evaluation of the simplicial norm for surfaces.

Exercise: Show that every oriented surface of genus ≥ 2 satisfies

$$\|[S]\|_\Delta = 2|\chi(S)|.$$

This is elementary. What is trickier is a similar *proportionality theorem* for other compact, locally homogeneous spaces V,

$$\|[V]\|_\Delta = c\,\mathrm{vol}(V),$$

where the constant c depends on the local geometry of V. This c is explicitly known for V of constant negative curvature, and it is nonzero for most (probably all) locally symmetric spaces with Ricci < 0 (see [Pan] and the references therein).

5.38. Upper bound on $\|\cdot\|_\Delta$. Now we have enough examples of nonzero simplicial norms to become interested in *upper* bounds.

Theorem: *Let the universal covering $X = \tilde{V}$ of V have packing type such that $P(b, 1) < N_0$. Then every homology class $h \in H_k(V; \mathbb{R})$ satisfies*

$$\|h\|_\Delta \leq \operatorname{const}(k, N_0)\|h\|_{mass}$$

for the mass norm defined in 4.15.

Idea of the proof. We proceed as in 5.34. In order to make that argument work, we shall need two essential properties of the simplicial norm.

(A) **π_1-essentiality.** *Let $f : V \to K(\Gamma; 1)$ be the classifying map for $\Gamma = \pi_1(V)$. Then $f_* : H_*(V; \mathbb{R}) \to H_*(K(\Gamma; 1); \mathbb{R})$ is isometric with respect to the simplicial norm.* (See $[\mathrm{Gro}]_{\mathrm{VBC}}$).

Now, if the group $\Gamma = \pi_1(V)$ is residually finite Γ, then we can pass to a finite d-sheeted covering $\tilde{V}_d$ of V and use the nerve of a suitable cover of $\tilde{V}_d$ by balls as a "triangulation" of the image $f_*(d[V]) \in H_*(K(\Gamma; 1))$ with the number of simplices controlled by $\operatorname{vol}(\tilde{V}_d) = d \operatorname{vol} V$ exactly as we did in 5.33. This takes care of $h = [V]$ and the rest of $H_*(V)$ is treated similarly.

Next, we must face the general case of non-residually finite Γ, and for this we need an extension of $\|\cdot\|_\Delta$ to *foliations with transversal measures.* This was suggested by A. Connes (unpublished) and presented in $[\mathrm{Gro}]_{\mathrm{FPP}}$. Here we explain the idea of Connes in the framework of our discussion in 5.34, and so we assume, to save notation, that V is aspherical and therefore plays the role of $K(\Gamma; 1)$ with f being the identity map. Then instead of triangulating some $\tilde{V}_d$, we equivariantly "triangulate" $X \times M$ (where $X = \tilde{V}$ and M is some probability Γ-space). The claim is that this gives the same value to the simplicial norm as the one we obtain with real cochains in $V = X/\Gamma$. In fact, there are *a priori* two distinct definitions of the *simplicial volume* of $(X \times M)/\Gamma$. The first one uses *integral* measurable chains: here *measurable k-simplices* are Γ-equivariant measurable maps

$$\sigma \colon \Delta^k \times \Gamma \times M' \to X \times M,$$

where each $\Delta^k \times \gamma \times m'$ goes to a single fiber $X \times m$, and where the resulting maps $M'_\gamma = \gamma \times M' \to M$ are measure preserving for each $\gamma \in \Gamma$. One builds the integral and the real chain complexes out of these σ with chains $c = \sum_i r_i \sigma_i$, where the norm is defined as

$$\|c\|_\Delta = \sum_i |r_i| \mu(M_i').$$

These integral and real chain complexes project to the corresponding singular chain complexes of V, and so one can speak of the representation of the homology of V by measurable chains. In particular, one applies this to the fundamental class $[V]$ and defines the integral and real simplicial M-volumes by taking the infima of the norms of such representation, denoted

$$\|[V]_{\mathbb{Z}}\|_\Delta^M \qquad \text{and} \qquad \|[V]_{\mathbb{R}}\|_\Delta^M.$$

(Notice that we could also define the simplicial norm $\|[V]_{\mathbb{Z}}\|_\Delta$ without any M by using *integral* singular chains representing $[V]$, but this would give us rather unruly objects.) It is clear that

$$\|[V]_{\mathbb{Z}}\|_\Delta^M \geq \|[V]_{\mathbb{R}}\|_\Delta^M$$

for all probability Γ-spaces M and by Connes' theorem (see $[\mathrm{Gro}]_{\mathrm{FPP}}$), we have

$$\|[V]\|_\Delta = \|[V]_{\mathbb{R}}\|_\Delta^M \tag{$*$}$$

for all M.

Exercise: Derive the above-mentioned proportionality of $\|[V]\|_\Delta$ from $(*)$.

We are now justified in defining the integral foliated simplicial volume as

$$\|[V]_{\mathbb{Z}}\|_\Delta^{\mathcal{F}} = \inf_M \|[V]_{\mathbb{Z}}\|_\Delta^M$$

over all probability Γ-spaces M. Clearly, this majorizes the ordinary simplicial volume $\|[V]\|_\Delta$ and, on the other hand, it can be majorized by $\operatorname{vol} V$ by the argument in 5.34. This concludes our sketch of the proof of the upper bound on $\|[V]\|_\Delta$.

Remarks and exercises: (a) Fill in the details in the argument (consulting the cited papers when necessary).

(b) Work out the case of $h \neq [V]$ and also of $V \neq K(\Gamma; 1)$.

(c) Obviously (with $(*)$),

$$\|[V]\|_\Delta \leq \|[V]_{\mathbb{Z}}\|_\Delta^{\mathcal{F}},$$

and so we bounded something better than $\|[V]\|_\Delta$. But the two invariants may be equal for all we know to date.

(*d*) One may relax the definitions of $\|[V]_{\mathbb{Z}}\|_{\Delta}^{M}$ and $\|[V]_{\mathbb{R}}\|_{\Delta}^{M}$ by allowing infinite chains $\sum_i r_i\sigma_i$ with $\sum_i |r_i|\mu(M_i') < \infty$. This does not change $\|[V]_{\mathbb{R}}\|_{\Delta}^{M}$ but may (?) diminish $\|[V]_{\mathbb{Z}}\|_{\Delta}^{M}$.

(*e*) Show that the sum of the ℓ_2-Betti numbers of $V = X/\Gamma$ satisfies

$$\sum_{k=0}^{n} h^k(X;\Gamma) \leq 2^n \|[V]_{\mathbb{Z}}\|_{\Delta}^{\mathcal{F}}$$

for all closed n-dimensional aspherical manifolds V. Extend this to non-aspherical V mapped to $K(\Gamma;1)$'s and thus generalize the ℓ_2-bound in 5.34. (One does not know if similar results hold with $\mathbb{R}$ in place of $\mathbb{Z}$.)

5.39. Bound for $\|[V]\|_{\Delta}$ by entropy. The following discussion sharpens the bound above for the simplicial volume, however it does not apply to $\|[V]_{\mathbb{Z}}\|_{\Delta}^{\mathcal{F}}$ any more.

Theorem.

$$\|[V]\|_{\Delta} \leq c_n (h(V))^n \operatorname{vol}(V) \tag{$\star$}$$

for some universal constant c_n (see p. 37 in [Gro]$_{\mathrm{VBC}}$ for the proof).

Notice that the constant given in [Gro]$_{\mathrm{VBC}}$ is not sharp. In fact, ($\star$) was originally proved by A. Katok (see [Katok]) *for surfaces* with *sharp* c_2 (and $\chi(V)$ instead of $\|[V]\|_{\Delta}$), who employed a conformal change of the metric to constant curvature with a subsequent use of the length-area method (compare [Gro]$_{\mathrm{FRM}}$). Recently, several sharp versions of ($\star$) were obtained by Besson, Courtois, and Gallot (see [Pan]$_{\mathrm{VCE}}$).

We conclude by observing that ($\star$) provides a *lower* bound on the volume growth of $X = \tilde{V}$ in terms of vol(V) and a purely topological invariant, the simplicial volume $\|[V]\|_{\Delta}$ of V.

H_+ Generalized simplicial norms and the metrization of homotopy theory

We start with several straightforward modifications of the definition of $\|\cdot\|_{\Delta}$.

5.40. Cubical norm, etc. It does not take much imagination to pass from the simplicial to *cubical* singular homology theory and to define the cubical norms $\|\cdot\|_{\square}$ on $H_k(X)$ accordingly. It is obvious that the two norms are *equivalent*, i.e.,

$$C_k \leq \frac{\|h\|_{\Delta}}{\|h\|_{\square}} \leq C_k'$$

for some C_k, C'_k, but the *exact* values of these constants are unknown. It is even possible (albeit unlikely) that $\|h\|_\Delta = A_k\|h\|_\square$ for some universal constant A_k.

One can go further in this direction and use more complicated cells to build up singular chains. The smallest such norm corresponds to the minimal number of cells in cellular decompositions of a space P. Using this number cell(h), $h \in H_k(X)$, realizing h by a map of an oriented pseudomanifold P with minimal possible number cell(P). This becomes a true norm after stabilization

$$\|h\|_{\text{cell}} = \lim_{i\to\infty} i^{-1}\,\text{cell}(ih).$$

One knows that this norm is nonzero on the fundamental classes of products of surfaces (see [Gro]$_{\text{PCMD}}$) and, more generally, is nonzero on the fundamental classes of locally symmetric spaces X with Ricci$(X) < 0$ and $\chi(X) \neq 0$ (see [Lustig]).

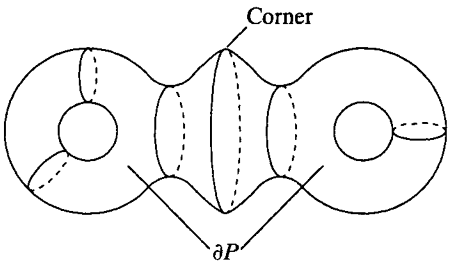
Rather than starting with cells, one can start with topologically more complicated manifolds (and pseudomanifolds V with boundaries subdivided (stratified) further in a cell-like fashion. For example, the standard simplex and cube are organized in this manner. But we can also admit some manifolds with or without (non-subdivided) boundaries or products of such with obviously subdivided boundaries, $W = V_1 \times V_2$, with

$$\partial W = (V_1 \times \partial V_2) \cup (\partial V_1 \times V_2).$$

Now, given such V, we try to assign to it some (simplicial-like) norm $\|V\|$ for which we want to satisfy the inequality $\|V_1\| \geq d\|V_2\|$ in the presence of a continuous map $V_1 \to V_2$ of degree d sending strata to strata. Furthermore, whenever V is assembled out of V_i with strata-preserving gluing maps on the pieces of the boundaries (as for pseudomanifolds being made out of the top-dimensional simplices), we want $\|V\| \leq \sum_i \|V_i\|$ (and if this fails, it should be attributed to some "norm" on the gluing maps). So, we may start with some collection of V_i, $i = 1, \dots, j$, assign a number (weight) α_j to each of them, and then define the maximal possible norm on all V's satisfying the requirements above. For example, if $\{V_i\}$ consists of the single simplex Δ^k with weight 1, then we recapture our simplicial norm. Or, we could start with V_i being products of surfaces with weights $|\chi(V_i)|$ as earlier.

Thus, one is led to a variety of combinatorially defined norms on the homology of spaces as well as on the (fundamental) homology classes of stratified spaces. (Notice that Δ^k admits *no face-preserving* map of degree $d \geq 2$, and so its "stratified norm" does not have to vanish.)

A particularly interesting class of examples is furnished by *hyperbolic polyhedra* P of the form $\tilde{P}/\Gamma$ for convex polyhedra $\tilde{P}$ in the hyperbolic space H^n. A case we want to emphasize is where neither P nor $\tilde{P}$ have vertices and perhaps not even 1-strata. For example, P may be a hyperbolic manifold with *totally geodesic* boundary, or a boundary made up of two hyperbolic $(n-1)$-manifolds meeting along *their* totally geodesic boundaries with a corner, and so on.



One can assemble more complicated spaces out of these polyhedra P by gluing them (isometrically, to be prudent) along the boundary strata. For example, some ramified coverings of hyperbolic manifolds with totally geodesic ramification loci come about this way (see [Gro-Thu]).

Problems: Compute the simplicial norms of manifolds assembled out of these polyhedra, e.g., of the above ramified coverings. Study the norms associated to specific collections of such polyhedra. Can two such norms ever be mutually proportional without being *obviously* proportional? Are all of these norms equivalent to the simplicial norm unless for some obvious reason to the contrary?

5.41. Norms on H_k defined via the geometry of cycles. Given a closed k-dimensional manifold or pseudomanifold V, one considers a class $\mathcal{G}$ of metrics on V and defines $\min_{\mathcal{G}} \operatorname{vol}(V)$ as the infimum of the volumes of all metrics $g \in \mathcal{G}$ on V. In what follows, this class $\mathcal{G}$ is essentially independent of V, and so one may speak of $\operatorname{vol}_{\mathcal{G}}(h)$, $h \in H_k(X)$, as the infimum of $\min_{\mathcal{G}} \operatorname{vol}(V)$ for all V realizing h (i.e., $h = f_*[V]$ for some continuous map $f\colon V \to X$). Finally, one stabilizes and sets

$$\|h\|_{\mathcal{G}} = \lim_{i\to\infty} i^{-1} \operatorname{vol}_{\mathcal{G}}(ih).$$

Examples: (a) Let $\mathcal{G}$ consist of piecewise Euclidean metrics such that each V is built of regular unit k-simplices Δ^k. Then the volume of such V equals the number of Δ^k's, and the norm $\|\cdot\|_{\mathcal{G}}$ is proportional to $\|\cdot\|_{\Delta}$.

(*b*) One can modify the above by insisting that there are at most N_ℓ k-simplices adjacent to each ℓ-face in (the triangulation) of V. This gives an *a priori* larger norm than $\|\cdot\|_\Delta$. Probably all such norms for sufficiently large N_ℓ are mutually equivalent (this seems easy), but it is unclear if these are equivalent to $\|\cdot\|_\Delta$. (This is related to the combinatorial filling problem from $5\frac{5}{7}$ of [Gro]$_{\text{PCMD}}$ solved for $k-1=2$, which implies a positive answer to the question above for $k=3$.)

(*c*) Now we take a more geometric stance and look at the following classes of metrics:

$$\begin{aligned}
\mathcal{G}_1 &= \{g : |K(g)| \le 1,\ \mathrm{InjRad}(g) \ge 1\},\\
\mathcal{G}_2 &= \{g : |K(g)| \le 1\},\\
\mathcal{G}_3 &= \{g : K(g) \ge -1\},\\
\mathcal{G}_4 &= \{g : \mathrm{Ricci}(g) \ge -(k-1)g\},\\
\mathcal{G}_5 &= \{g : \mathrm{Sc}(g) \ge -k(k-1)\},
\end{aligned}$$

where K denotes the sectional curvature and Sc is the scalar curvature. Obviously,

$$\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{G}_3 \subset \mathcal{G}_4 \subset \mathcal{G}_5.$$

It is also clear that the norm $\|\cdot\|_{\mathcal{G}_1}$ is equivalent to the one in (b) with sufficiently large (locally complex numbers) N_ℓ. Essentially all we know about these norms is that $\|\cdot\|_{\mathcal{G}_4}$ dominates $\|\cdot\|_\Delta$,

$$\|\cdot\|_\Delta \le \mathrm{const}_k \|\cdot\|_{\mathcal{G}_4}.$$

The most interesting and mysterious is the norm $\mathcal{G}_5$. One would be happy to prove that it is *not* identically zero. (This may follow for $k=3$ from the recent work by M. Anderson on the geometrization conjecture.)

One can continue with such questions further revealing our ignorance about the basic topology-geometry interaction. For example, one could look at the class $\mathcal{G}_4'$ consisting of the metric g with respect to which (V,g) has a small covering entropy, say $h \le \log 2$ (i.e., the volume growth of the R-balls in the universal cover is bounded by $\mathrm{vol}\,\tilde{B}(R) \le 2^R$ for $R \to \infty$). This again is known to dominate $\|\cdot\|_\Delta$, but one does not know if this $\|\cdot\|_{\mathcal{G}_4'}$ is equivalent to $\|\cdot\|_\Delta$. Also, one could restrict g by means of the systole $\mathrm{syst}_1(V,g)$. Namely, one may define $\mathrm{min}_{sy}\,\mathrm{vol}(V)$ as the supremal number $\sigma \ge 0$ such that

$$\mathrm{syst}_1(V,g) \le \sigma^{-1/n} \log(1+\sigma)\,\mathrm{vol}(V,g)^{1/n}$$

for *all* metrics g on V, and then set $\mathrm{vol}_{sy}(h)$ and $\|h\|_{sy}$ accordingly. Again, this definition is motivated by the bound $\|\cdot\|_\Delta \leq \mathrm{const}_k \|\cdot\|_{sy}$ (see [Gro]$_{\mathrm{FRM}}$).

Finally, one may introduce integral versions of the classes $\mathcal{G}_1 - \mathcal{G}_5$ and the corresponding norm on $H_*(X)$. Namely, one could minimize the integrals of $|K(g)|^{k/2}$, $|K_-|^{k/2}$, $\|\,\mathrm{Ricci}_-\,\|^{k/2}$ and $|\,\mathrm{Sc}_-\,|^{k/2}$ over (V, dg), where $f_-(v) = \min(0, f(v))$ at all $v \in V$, instead of the volumes of g. Again, we do not know if either of the resulting norms is *not* identically zero. The easiest case is that of the full sectional curvature, and yet we have no examples of a class $h \in H_k(X)$, $k \geq 3$, for some X, such that every Riemannian manifold (V, g) admitting a continuous map $V \to X$ sending $[V]$ to ih_k satisfies

$$i^{-1} \int_V |K(g)|^{k/2}\, dg \geq \varepsilon = \varepsilon(h) > 0$$

for all $i \in \mathbb{Z}$. (The situation is slightly more cheerful for *bordism* classes instead of homology, where the lower bounds on $\int |K|^{k/2}$ come from characteristic numbers. (See [Gro]$_{\mathrm{PCMD}}$ for more interesting geometric bordism invariants.)

5.42. Intrinsic metrics on homotopy types of spaces. Given a topological space X, we want to construct in a functorial way a metric space $\mathcal{X}$ which is homotopy equivalent to X so that continuous maps $X \to Y$ would transform into *distance decreasing* maps $\mathcal{X} \to \mathcal{Y}$. Then we could define the volumes and $\mathbb{R}$-masses on the homology $H_*(\mathcal{X}) = H_*(X)$, thus getting new norms of the above kind.

There is one standard construction which assigns to every X a simplicial polyhedron $\mathcal{X}$ as follows. Take an arbitrary countable subset $\mathcal{X}_0 \subset X$ for the vertex set of $\mathcal{X}$, where we assume that X is connected and consists of more than a single point. Then use the homotopy classes of paths in X between distinct $x_i \in \mathcal{X}_0 \subset X$ for edges. Next, fill all triangles of paths in X by all possible homotopy types of maps of the 2-simplex $\Delta^2 \to X$, and so on. Thus, one gets our $\mathcal{X}_0$ with a distinguished homotopy equivalence $\mathcal{X}_0 \to X$ (compare pp. 41-48 in [Gro]$_{\mathrm{VBC}}$). This $\mathcal{X}_0$ is built of infinitely many simplices Δ^k, $k = 0, 1, \ldots$, and one can give $\mathcal{X}_0$ a metric by assigning some metrics to the standard simplices Δ^k, $k = 1, 2, \ldots$, so that this assignment on Γ^{k1} agrees with the restriction of the metric on Δ^k to the $(k-1)$-faces.

There are innumerably many such metrics; here are a few which immediately come to mind.

(*i*) Take the standard Euclidean metric on each
$\Delta^k = \{x_i \geq 0 : \sum_{i=0}^k x_i = 1\} \subset \mathbb{R}^{k+1}$.

(*ii*) Use the spherical metric on the above Δ^k radially projected to $S^k \subset \mathbb{R}^{k+1}$.

(*iii*) Try the metric induced from the ℓ_1-norm on $\mathbb{R}^{k+1}$.

One can show in this regard that a suitably normalized $\mathbb{R}$-mass in (*iii*) equals the simplicial norm $\|\cdot\|_\Delta$ (compare pp. 91-92 in [Gro]$_{\mathrm{VBC}}$). We suggest this as an exercise for the reader. It also seems not hard to show that the norms coming from (*ii*) and (*iii*) are mutually equivalent, but (*i*), I am afraid, always leads to the identically zero norm.

Question: Do the norms coming from the function $X \rightsquigarrow \mathcal{X}_0$ exhaust all possible norms on homology with suitable metrics on Δ^k? Or, can one gain in generality by using more complicated $\mathcal{X}$'s? (e.g., built of cubes mapped to X.)

In practice, a space $\mathcal{X}$ may appear not exactly as the above $\mathcal{X}_0$. For example, the simplicial volume comes along with the natural action of a group Γ on the space $\Delta^\infty(\Gamma)$ of all probability measures on Γ with the ℓ_1 metric. The action of Γ on this Δ^∞ is not quite free (Γ may have torsion), but nevertheless the real homology of Γ can be realized by cycles in Δ^∞/Γ or, better, by Γ-equivariant cycles in Δ^∞. Thus, a suitable ℓ_1-volume on k-chains in Δ^∞ leads to the simplicial norm on $H_*(\Gamma)$ (compare 2.4 in [Gro]$_{\mathrm{VBC}}$). One wonders what happens here for other natural metrics on Δ^∞, e.g., the spherical one. Furthermore, one may try different actions of Γ on function spaces, e.g., the regular representation on the unit sphere $S^\infty \subset \ell_2(\Gamma)$. Also, if the group Γ is hyperbolic, then one has an action on the space of measures on the ideal boundary $\partial\Gamma$, and if this boundary happens to come with a C^1-structure (as for 1/4-pinched hyperbolic manifolds), then one can use the Hilbert space of 1/2-densities on $\partial\Gamma$ (as in [B–C–G], where the authors employ this idea for Γ action on the ideal boundaries of symmetric spaces of $\mathbb{R}$-rank one).

Problem: Study the volume "norm" and the $\mathbb{R}$-mass norm on the homology of natural function spaces X, especially those of the form $X = \tilde{X}/\Gamma$. Study the Plateau problem in these X and identify whenever possible the minimal subvarieties.

We have already encountered this problem in the systolic framework, but our present pool of spaces X seems different from the one in Ch. 4 where the geometry was tilted to L_∞ rather than to L_1, as in the simplicial volume picture.

5.43. Thurston's idea of straight simplices. The present conception of the simpicial volume has grown out of the following startling geometric observation made by Thurston about 20 years ago. Let V be a complete manifold with constant negative curvature, let P be a simplicial polyhedron of dimension k, and let $f_0 : P \to V$ be a continuous map. Then f_0 can be deformed to a map f such that *the k dimensional volume of the image $f(P)$ (counted with multiplicity if the need arises) is bounded by a constant times the number of k-simplices in P, provided that $k \geq 2$. For example, if both P and V are closed k-dimensional manifolds, then there is a bound on the degree of f in terms of the topology of P (see [Thurs]).*

This was, of course, well-known prior to Thurston in the case of $k = 2$, at least for maps of *surfaces* into V^n, $n \geq 2$, since one could deform f_0 to some *minimal* (say harmonic) map f and then use Gauss-Bonnet. But no tool was in sight for $k \geq 3$. In fact, it was not even clear why one could not significantly enlarge the volume of a hyperbolic manifold by deforming the metric while keeping the curvature below -1, although it had an irresistible appeal to one's geometric intuition. Then one day the following breathtaking argument (that only Thurston could come up with) was unloaded on me by Dennis Sullivan:

Straighten the map on every simplex Σ^i from P mapped to V. Namely, first deform each edge to the geodesic segment between the images of the vertices, then deform the curved 2-simplices to totally geodesic ones, and so on. Now we have an f sending each simplex Δ^k in P to a totally geodesic simplex $\underline{\Delta}^k$ in V where

$$\mathrm{vol}_k\, \underline{\Delta}^k \leq \sigma_k < \infty$$

for $k \geq 2$, as everybody since Lobashevskii knows.

Actually, Dennis brought up this argument to justify the question he posed:

Does the degree of a continuous map $f : V_1 \to V_2$ between equidimensional hyperbolic (i.e., with $K = -1$) manifolds satisfy

$$\deg f \leq \frac{\mathrm{vol}\, V_1}{\mathrm{vol}\, V_2}? \qquad (*)$$

(It could have been a slightly different question in the same spirit.) And then, he shot me with the "straightening" proof in order to arrest the flow of my objections to the feasibility of obtaining $(*)$, even in a non-sharp version, by means of the hyperbolic geometry available at the time.

The straightening argument, when applied to singular real chains in V, instantaneously bounds the $\mathbb{R}$-mass by the simplicial norm

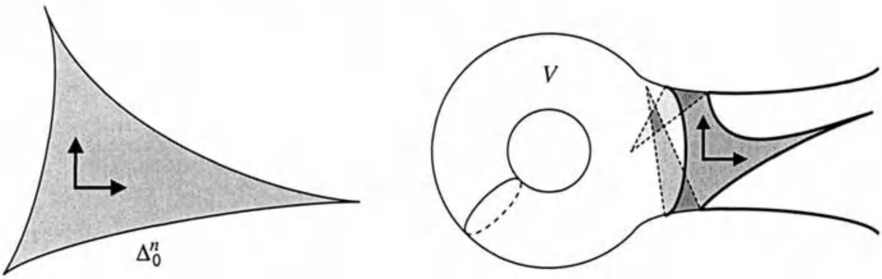
$$\|h\|_{\mathrm{mass}} \leq \sigma_k \|h\|_\Delta, \qquad (+)$$

where σ_k denotes the supremum of the volumes of the hyperbolic k-simplices. On the other hand, every such simplex, say $\Delta_0^n \subset H^n$, provides an *upper bound* on the simplicial volume of a closed hyperbolic manifold V of dimension n,

$$\|[V]\|_\Delta \leq \frac{\mathrm{vol}(V)}{\mathrm{vol}(\Delta_0^n)}. \qquad (-)$$

Ideally, one would like to prove $(-)$ by triangulating V with all n simplices of the triangulation isometric to Δ_0^n. The next best thing would be a triangulation of some finite covering $\tilde{V} \to V$ into isometric copies of Δ_0^n. In fact, it would (obviously) suffice to construct a sequence of triangulated i sheeted coverings $\tilde{V}_i \to V$ where almost all simplices of $\tilde{V}_i$ are nearly isometric to Δ_0^n, where "almost" and "nearly" disappear in the limit as $i \to \infty$.

In general, such triangulations do not exist; nevertheless, there is a measurable "triangulation" of the fundamental class $[V]$ into a family of simplices in V isometric to Δ_0^n. This family is parametrized by the bundle of the orthonormal frames in V, say $M(V)$ (which equals $\mathrm{Isom}\, H^n/\Gamma$ for $\Gamma = \pi_1(V)$, where we rigidly attach a frame m_0 at some point $x_0 \in \Delta_0^n$ and then moving m_0 to some frame $m = m_v \in M_v(V) \subset M(V)$ takes Δ_0^n along to the corresponding position in V.



Thus, we have a family of simplices Δ_m^n, $m \in M(V)$, in V where each is given the orientation induced by a fixed orientation in V. (These $\Delta_m^n \subset V$ are locally isometric to Δ_0^n but may have self-intersections.) Our family, albeit infinite, makes an *n-cycle*, since every $(n-1)$-face is contained in the exactly two copies of Δ_0^n, of which one is obtained from the other by the reflection in this face. What remains is to supress squeamishness and accept this measurable cycle c for a legitimate representative of $[V]$, or rather of some multiple of $[V]$. Here,

$$\|c\|_\Delta = \mu(M)$$

and

$$c_{\mathrm{mass}} = \mu(M)\,\mathrm{vol}(\Delta_0^n),$$

where the measure on M descends from the Haar measure on the group $\operatorname{Isom} H^n$ (since $M = \operatorname{Isom} H^n/\Gamma$), and where the second equality is due to the fact that each simplex in our family Γ^n_m contributes *positively* to $[V]$ being oriented according to the orientation of V defining the class $[V]$. In other words,

$$\|c\|_{\text{mass}}/\|c\|_\Delta = \operatorname{vol}(\Delta^n_0),$$

and conversely,

$$\|[V]\|_\Delta \leq \sigma_n^{-1}\|[V]\|_{\text{mass}} = \sigma_n^{-1}\operatorname{vol}(V) \tag{$-$}$$

since $\|[V]\|_\Delta \leq \|c\|_\Delta$ by the definition of the simplicial norm of V as the inf of these over all cycles representing $[V]$.

Finally, adding up (+) and (−), we obtain the inequality

$$\|[V]\|_\Delta = \sigma_n^{-1}\operatorname{vol}(V), \tag{$\circ$}$$

which implies (∗) by the functorality of the norm $\|\cdot\|_\Delta$. (Our intuition on measurable cycles comes from the geometry of flat and almost flat manifolds. Here one clearly sees how isometric copies of a Euclidean simplex "measurably triangulate" (the fundamental class of) the flat torus $\mathbb{T}^n$. More generally, every manifold V with $K(V)| \leq \varepsilon^2$ can be "measurably triangulated" into roughly Euclidean simplices of size $\approx \varepsilon^{-1}$, where these simplices may wrap around V if the injectivity radius of V is small. Such triangulations were originally applied to the bound on the Betti numbers of V,

$$\sum_{i=0}^n b_i \leq 2^n + \exp(\operatorname{const}_n \varepsilon \operatorname{diam}(V)).$$

But more interesting bounds on the homotopy invariants of V by the *volume* instead of the diameter were not practical without sufficiently many examples of manifolds with nonzero simplicial volume, which were not available prior to Thurston's idea.

It is not very hard to analyze the case of equality in (∗): this happens if and only if the map $f\colon V_1 \to V_2$ in question is homotopic to an *isometric covering*, provided $n \geq 3$. The easiest case here is where f is a homotopy equivalence with the conclusion equivalent to the *Mostow rigidity theorem* (see $[\text{Gro}]_{\text{HMTJ}}$, $[\text{Gro–Pan}]_{\text{Rig}}$, and the references therein). Probably the full Mostow rigidity for all locally symmetric spaces can be derived by a similar analysis of the equality case for the corresponding (∗)-inequality, compare [B–C–G].

Exercise: Prove the proportionality relation $\|[V]\|_\Delta = c \operatorname{vol} V$ for all compact locally homogeneous spaces V.

Finally, to get a fuller view of $\|\cdot\|_\Delta$, one should dualize and look at the cohomology. This shift in perspective was triggered by the challenging question put to me by P. Trauber: *Prove that the bounded cohomology of an amenable group vanishes.* This may not look very impressive, since everything "bounded" averages over "amenable" and kills cohomology without even knowing the specific definitions. (Trauber had something else up his sleeve, a conceptual proof of a theorem by Hirsch and Thurston about some affine manifolds with *non-amenable* fundamental groups.) But the idea of bounded cohomology of groups (which, I assume, is due to Trauber, compare [Johns]) proved very handy in the general framework of singular cohomology of general topological spaces X. Now one could interpret Thurston's straightening argument as a construction of a *bounded* (because $\operatorname{vol}\Delta^k \leq \sigma_k < \infty$) cocycle in a given cohomology class of a hyperbolic manifold. In fact, such a cocycle can also be seen as a pointwise bounded, Γ-invariant closed k-form on the above simplex Δ^∞ of measures on Γ. (This leads one to a problem of getting a clearer view on bounded (closed) forms on more general geometric models $\mathcal{X}$ of X.)

I_+ Ricci curvature beyond coverings

All applications of the condition Ricci $\geq -$const were reduced so far to the issuing bound on the packing type dependent on Bishop's inequality for balls. But there is more metric information in Ricci than just a bound on the volume of balls.

5.44. Volume growth of geodesic sectors. A subset S in a path metric space V is called *a geodesic sector with the focal point* $v \in V$ if it consists of a union of minimizing (geodesic) segments between the points $s \in S$ and v. That is, for each $s \in S$, there is a geodesic segment between s and v. (One gets a slightly different and also useful definition by requiring *all* geodesic segments $[v, s]$ to be in S.) We define, for such an S, the *annuli*

$$S(R_1, R_2) = \{s \in S : R_1 \leq \operatorname{dist}(s, v) \leq R_2\},$$

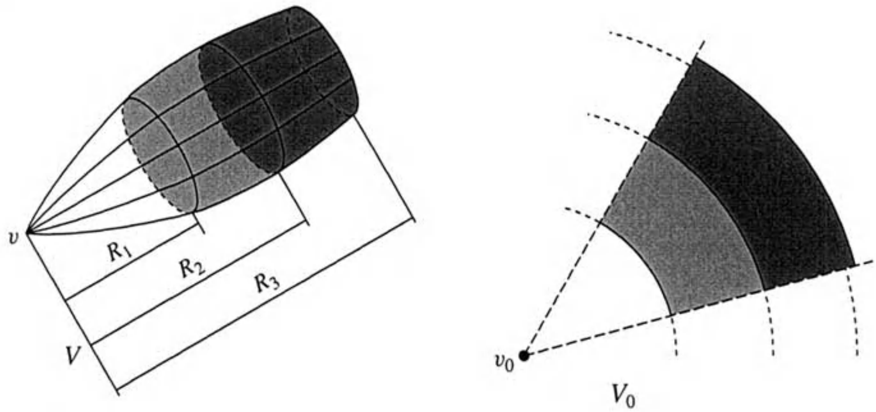
and we observe that the proof of Bishop's theorem provides a bound on the volume of such annuli in terms of what happens in the model space V_0, which is a complete, simply connected Riemannian manifold of *constant sectional curvature.* Namely, *if V is a complete Riemannian manifold with* $\dim V_0 = \dim V$, *and* $\operatorname{Ricci} V \geq \operatorname{Ricci} V_0$, *then*

$$\frac{\operatorname{vol}(S(R_2, R_3))}{\operatorname{vol}(S(R_1, R_2))} \leq \frac{\operatorname{vol}(A_0(R_2, R_3))}{\operatorname{vol}(A_0(R_1, R_2))}$$

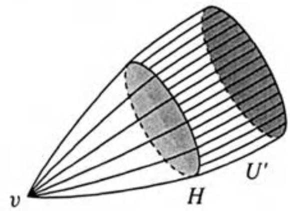
for arbitrary $R_1 \le R_2 \le R_3$*, where* $A_0(R, R')$ *denotes the annulus in* V_0*,*

$$A_0(R, R') = \{x \in V_0 : R \le \operatorname{dist}(x, v_0) \le R'\}$$

for a fixed point $v \in V_0$*.*



Notice that this property (as well as the inequality for the balls) is stable under the $\underline{\square}$-limits of mm-spaces and so could be used as an axiom of Ricci $\ge -\text{const}$ for singular spaces (compare [Che–Col]). Also, this property allows a bound on the volume of a region $U' \subset V$ shadowed by a hypersurface H from a point $v \in V$.

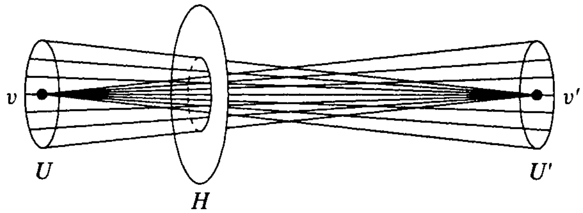


For example, if $\operatorname{Ricci}(V, g) \ge -(n-1)g$ (and so V_0 has $K(V_0) = -1$) and $\operatorname{diam} U' \le 1$, then

$$\operatorname{vol} U' \le \operatorname{const}_n (\operatorname{vol}_{n-1} H) \frac{\operatorname{diam} U'}{\operatorname{dist}(v_0, H)}. \tag{$*$}$$

This follows trivially from the above for $R_2 \to R_1$ and $S_2(R_1, R_2)$ degenerating to H.

Now, if we want to (and we often do) estimate the volume of a region U' bounded by a hypersurface H in V, we need a point v outside U far away from H. It may happen that we stay at a wrong point v in $U = V \setminus U'$ as



we observe some point $v' \in U'$. But then we can switch our point of view and look at v from v', thus bounding some measure in U.

One easily proves in this way the following *Poincaré–Cheeger* type *isoperimetric inequality* for the above V:

If V is divided into two domains U and U' by a closed hypersurface H, then

$$\min(\operatorname{vol} U, \operatorname{vol} U') \geq \operatorname{const}_n \operatorname{vol}_{n-1} H.$$

In fact, a similar inequality holds for all closed manifolds with given bounds $\operatorname{Ricci}(g) \geq R$ and $\operatorname{diam} V \leq D$ and yields (by Cheeger's inequality) a lower bound on the spectrum on the Laplace operator on V (see Appendix C_+).

5.45. Volume growth of distance tubes. A subset $T \subset V$ is called a *distance tube* with base $T_0 \subset V$ if, for every point $t \in T$, there is a segment $[t_0, t] \subset T$ with $t_0 \in T_0$ and such that $\operatorname{length}[t_0, t] = \operatorname{dist}(t, T_0)$. We define the annuli

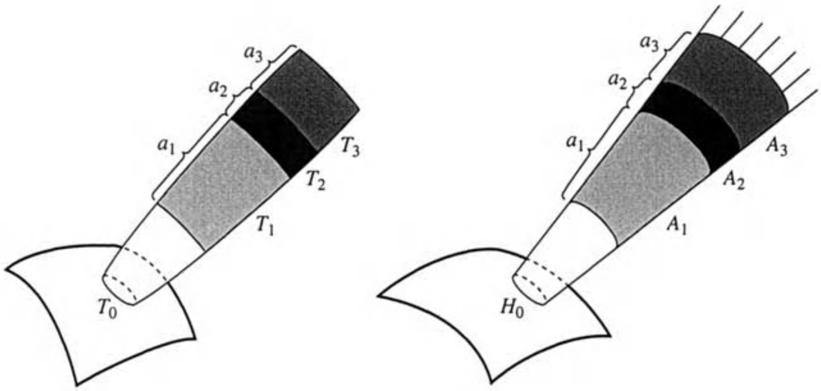
$$T(R, R + a) = \{t \in T : R \leq \operatorname{dist}(t, T_0) \leq R + a\},$$

and we observe the following inequality for the volume of *three* consecutive annuli, $T_i = T(R_i, R_{i+1})$, $i = 1, 2, 3$, $R_1 \leq R_2 \leq R_3 \leq R_4$, of *widths* $a_i = R_{i+1} - R_i$ relating these to the volumes of annuli A_i of the same widths in a model space V_0 of constant curvature with $\operatorname{Ricci} V_0 \leq \operatorname{Ricci} V$, where the base is a hypersurface $H_0 \subset V_0$ with all principal curvatures equal to some constant κ_0. (This H_0 is a sphere if $K(V_0) = \varepsilon_+ > 0$, H_0 is a sphere or a hyperplane if $V_0 = \mathbb{R}^n$, and it may also be a horosphere for $K(V_0) = \varepsilon_- < 0$.)

If $\operatorname{vol}(T_2)/\operatorname{vol}(T_1) \leq \operatorname{vol}(A_2)/\operatorname{vol}(A_1)$, *then*

$$\frac{\operatorname{vol} T_3}{\operatorname{vol} T_2} \leq \frac{\operatorname{vol} A_3}{\operatorname{vol} A_2}, \qquad (\star)$$

where we assume the geodesic tube in the model space V_0 is complete, i.e., contains a geodesic continuation of every geodesic segment insofar as it



remains distance-minimizing to H (while the original T does not have to be complete in this sense).

The relation $(\star)$ is equivalent to saying that the normal equidistant deformation of smooth hypersurfaces H in V increases their mean curvature *faster* than it happens for H_0 in V_0. This follows trivially from the definition of Ricci via the traced tube formula in 8.7 (compare $[\mathrm{Gro}]_{\mathrm{sign}}$ and the isoperimetric appendix in [Mil–Sch]). Also observe that $(\star)$ is stable under $\underline{\Box}_1$-limits.

Final remarks. The inequality $(\star)$ originates from the splitting paper by Cheeger and Gromoll, although it was known in some form to Paul Levy even earlier (compare Appendix C_+). More recently, it was used by Abresch and Gromoll in their excess inequality for *Ricci* $\geq -\mathrm{const}$ (see [Abre–Grom]) followed by a flow of new remarkable development of metric geometry with Ricci $\geq -$ const (see $[\mathrm{Col}]_{\mathrm{ARC}}$ as well as the other papers by Cheeger and Colding in our bibliography).

Chapter 6

Isoperimetric Inequalities and Amenability

A. Quasiregular mappings

6.1. Definition: Let $(V, g), (W, h)$ be oriented Riemannian manifolds. A mapping $f : V \to W$ is called *quasiregular* if it is locally Lipschitz, thus differentiable almost everywhere, and if its differential Df_x and Jacobian $J(x)$ satisfy the inequality $0 < \|Df_x\|^n \leq c\, J(x)$ for almost all x, where c is a constant.

Remark 1: We have included the local Lipschitz hypothesis for simplicity. Some specialists (see [Mar–Ri–Vais], Def. 2.20) use a more general definition in which f is only assumed to be ACL^n, from which they deduce that f is differentiable almost everywhere and that its Jacobian is strictly positive almost everywhere (see [Mar–Ri–Vais], Thm. 8.2).

A quasiregular mapping is called *quasiconformal* if it is a homeomorphism (cf. [Bers] for complete details).

6.2. Examples: Isometries and quasi-isometries are quasiregular, as are nonconstant holomorphic mappings of $\mathbb{C}$ into a Riemann surface. If we perturb a quasiregular mapping on a compact set, then the result is again quasiregular, provided that it preserves orientation. For this reason, we are mainly interested in quasiregular mappings of *noncompact* manifolds V.

6.3. Remark 2: Let p be the bounded, measurable function $V \to \mathbb{R}_+$ given by $x \mapsto \|Df_x\|$. The measurable tensor pg defines a length structure on V (indeed, we can compute the lengths of many paths using the formula $\mathrm{length}(\gamma) = \int p(\gamma(t))\sqrt{g(\gamma'(t),\gamma'(t))}\,dt$). For this length structure, the mapping $f : V \to (W, h)$ is quasi-isometric. Thus, a quasiregular mapping can be factored as a "conformal homeomorphism" and an arc-wise quasi-isometry whose Jacobian is positive almost everywhere, with the caveat that the intermediate space has a metric defined only a.e.

B. Isoperimetric dimension of a manifold

6.4. Definition: Let (V, g) be a Riemannian manifold and $I : \mathbb{R} \to \mathbb{R}$ an increasing function. We say that V satisfies *the isoperimetric inequality given by the function* I if there are constants K, K' such that, for each compact subset D of V with boundary, we have

$$\mathrm{vol}(D) \leq K\, I(K' \mathrm{vol}(\partial D)).$$

In particular, we say that V has isoperimetric dimension greater than m if it satisfies the isoperimetric inequality given by the function $t \mapsto t^{m/m-1}$. The *isoperimetric dimension* of V is then the supremum of the real $m > 0$ such that

$$\sup_D (\mathrm{vol}(D))^{m-1} < \infty.$$

By convention, manifolds of finite volume have zero isoperimetric dimension, and the isoperimetric dimension of no manifold lies strictly between 0 and 1. It is essential that we *not* assume that V is necessarily complete.

6.4$\frac{1}{2}$$_+$ Asymptotic remark. In what follows, we are mainly concerned with the *asymptotic* geometry of V and with the isoperimetric behavior of *large* domains D in V, say of those with $\mathrm{vol}\, D \geq 1$. This does not disturb the picture in most cases, since small D automatically satisfy the standard Euclidean inequality under rather mild assumptions on the local geometry of V. For example, if V has *locally bounded geometry,* i.e., $|K(V)| \leq C^2$ and $\mathrm{InjRad}\, V \geq \rho > 0$, then there exists a positive $\delta = \delta(C, \rho, n = \dim V)$ such that every $D \subset V$ with $\mathrm{vol}\, D \geq \delta$ satisfies $\mathrm{vol}\, D \leq \mathrm{const}_n(\mathrm{vol}(\partial D))^{n/(n-1)}$. This follows trivially from such an inequality for $V = \mathbb{R}^n$ by the cutting argument in Ch. 6C. Moreover, this conclusion remains valid under the milder assumptions $\mathrm{Ricci}\, V \geq -C^2$ and $\mathrm{vol}(B(1)) \geq \rho^n$ for the unit balls $B(1) \subset V$ (where it is safer to assume that V is complete). This follows from the argument in Appendix C_+. In fact, the manifolds V we are most interested in here are universal covers of compact manifolds, and these satisfy

$\operatorname{vol} D \leq I(\operatorname{vol}(\partial D))$ for *some* function I which is equal to $\operatorname{const}_V t^{n/(n-1)}$ for $t \leq 1$ as we shall see later on.

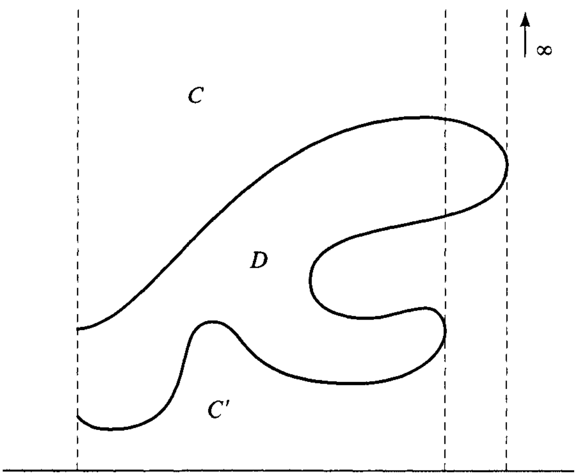
6.5. Examples: (a) The isoperimetric dimension of $\mathbb{R}$ is 1, and one can easily show that the isoperimetric dimension of a product of Riemannian manifolds is the sum of the dimensions of the factor spaces. It follows that the isoperimetric dimension of $\mathbb{R}^n$ equals n.

(b) The classical isoperimetric inequality in $\mathbb{R}^n$ with the optimal constant $(\operatorname{vol}(D))^{n-1} \leq (\operatorname{vol}(\partial D))^n / n^n \omega_n$ is more difficult but possible to obtain by using the splitting $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$ and induction on n (compare [Osser], [Bur–Zalg]).

(c) If V is an open ball of radius r in $\mathbb{R}^n$, then $\operatorname{vol}(D) \leq r^{n-1} \operatorname{vol}(\partial D)$ for each compact set D with boundary. Hyperbolic space shares the same property.

6.6. Proposition: *If A is a subset of $\mathbb{H}^n$ with boundary, then*

$$\operatorname{vol}(A) \leq \frac{1}{n-1} \operatorname{vol}(\partial A).$$



Proof. Choose a point at infinity of $\mathbb{H}^n$. Let C be the cone formed from the segments joining this point to the points of A, and let C' be the base of the cone. Then $A \subset C$ and $C' \subset \partial A$, and we will show that $\operatorname{vol}(C) \leq (1/n-1) \operatorname{vol}(C')$. In the half-space $x_n > 0$ of $\mathbb{R}^n$ equipped with the metric

$\sum dx_1^2/x_n^2$, the cone C is the set of points lying above a point of C', and there is a function f defined on a domain $D \subset \mathbb{R}^{n-1}$ such that

$$C = \{x_n > f(x_1, \dots, x_{n-1})\}.$$

Its graph equals C', and we have

$$\mathrm{vol}(C) = \int_C \frac{1}{x_n^n}\, dx_1 \cdots dx_n = \int_D \frac{1}{(n-1)} f^{n-1}\, dx_1 \cdots dx_{n-1}$$

$$\begin{aligned} \mathrm{vol}(C') &= \int_D [1 + (\partial_1 f)^2 + \cdots + (\partial_{n-1} f)^2]^{1/2} f^{1-n}\, dx_1 \cdots dx_{n-1} \\ &\geq \int f^{1-n}\, dx_1 \cdots dx_{n-1} = (n-1)\, \mathrm{vol}(C). \end{aligned}$$

6.7. Remark: S.T. Yau has shown that, more generally, if (V, g) is simply connected and has sectional curvature less than $-a^2$ for $a > 0$, then for every domain $D \subset V$, we have $\mathrm{vol}(D) \leq a\, \mathrm{vol}(\partial D)/(n-1)$ (cf. [Yau]$_{\mathrm{ICFE}}$, p. 498, and section 8.12 of this book).

Exercise: Recover the proof of this property from the note [Avez] of Avez.

These examples lead us, following Sullivan [Sul]$_{\mathrm{Cyc}}$, to make the following definition:

6.8. Definition: A manifold (V, g) is said to be *open at infinity* if there is a constant C such that $\mathrm{vol}(D) \leq C\ \mathrm{vol}(\partial D)$ for each domain $D \subset V$.

This property is *a priori* stronger than that of infinite isoperimetric dimension, since it could be the case that for each α a best constant $C(\alpha)$ such that $\mathrm{vol}(D) \leq C(\alpha)\, \mathrm{vol}(\partial D)^{\alpha/\alpha - 1}$ would satisfy $\lim_{\alpha \to \infty} C(\alpha) = \infty$.

$6.8\frac{1}{2}_+$ Example. If the fundamental group of a compact manifold is amenable without being virtually nilpotent, e.g., solvable nilpotent, then *the universal covering has infinite isoperimetric dimension* without being open at infinity. This result is due to Varopoulos, who has also shown that nonnilpotent, solvable, simply connected, unimodular Lie groups V with left-invariant Riemannian metrics satisfy

$$\mathrm{vol}\, \partial D \geq C\, \mathrm{vol}\, D/(\log \mathrm{vol}\, D)^2$$

(see Ch. 6.E$_+$).

6.9. The similarity in the behavior of the Euclidean disk and the hyperbolic plane suggests that the isoperimetric dimension could be a conformal invariant. This is half-true, thanks to

Lemma (Ahlfors, cf. [Ahl] and [Osser]). *Let (V, g) be a complete Riemannian n-manifold and assume that there is a point $a \in V$ with*

$$\int_1^\infty \mathrm{vol}(\partial B(a,r))^{-1/(n-1)}\, dr = \infty.$$

Then the isoperimetric dimension of every metric conformally equivalent to g is $\leq n$.

Proof. Given the metric fg, we must show that for each $\delta > 0$, there is a sequence D_i of domains such that

$$\lim_{i\to\infty} \frac{\mathrm{vol}_{fg}(D_i)}{(\mathrm{vol}_{fg}(\partial D_i))^{n/(n-1)-\delta}} = \infty.$$

As we shall see, we can take the D_i to be the balls $B(a, r_i)$ defined with respect to the metric g.

Set

$$X(r) = \mathrm{vol}_{fg}(B(a,r)), \qquad Y(r) = \mathrm{vol}_{fg}(\partial B(a,r))^{n/n-1},$$

and $A(r) = \mathrm{vol}_g(\partial B(a,r))^{-1/n-1}$. Since

$$X(r) = \int_{B(a,r)} f^{n/2}\nu_g = \int_0^r \left(\int_{\partial B(a,s)} f^{n/2}\nu_g\right) ds,$$

it follows from Hölder's inequality (with exponents $n/n-1$ and n), applied to $f^{(n-1)/2}$ and 1, that

$$X(r) \geq \int_0^r \left(\int_{\partial B(a,s)} f^{(n-1)/2}\nu_g\right)^{n/n-1} A(s)\, ds = \int_0^r A(s)Y(s)\, ds.$$

We now show that for each $\delta > 0$, $\lim_{r\to\infty} X(r)/Y(r)^{1-\delta} = \infty$. If this were false, then for some $\alpha > 0$, an inequality of the form $X^{1+\alpha}(r) \leq CY(r)$ would hold for large r, and so by setting $Z(r) = \int_0^r A(s)Y(s)\, ds$, we would obtain the inequalitites $Z'(r) = A(r)Y(r) \geq C^{-1}Z^{1+\alpha}(r)A(r)$. Thus,

$$-(Z^{-\alpha})'(r) = \alpha Z'(r)Z^{-1-\alpha}(r) \geq C^{-1}A(r),$$

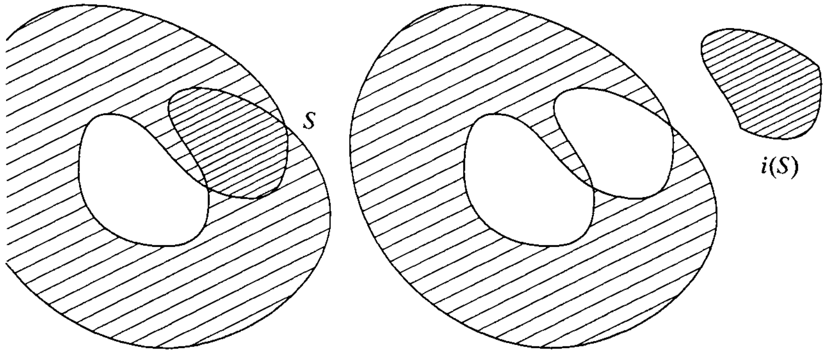
and integration on both sides gives

$$C^{-1}\int_R^\infty A(r)\, dr \leq Z^{-\alpha}(R) - Z^{-\alpha}(\infty),$$

contradicting the hypothesis.

To fully empower Ahlfors' lemma, it remains only to prove

6.10. Proposition: *If there is an arc-wise quasi-isometric map of V into a compact W whose Jacobian is a.e. positive, then the isoperimetric dimension of the universal covering $\tilde{V}$ of V is greater than or equal to that of $\tilde{W}$, provided that $\tilde{W}$ is noncompact.*



$f(A)$ $\qquad\qquad$ $f(A)\setminus S$

Proof. Let m be a real number strictly smaller than the isoperimetric dimension of W. Then, for each compact subset $B \subset W$ with boundary, we have $\mathrm{vol}(B) \leq \mathrm{const}(\mathrm{vol}(B))^{m/m-1}$. If $f\colon V \to W$ is an arc-wise quasi-isometry and $A \subset V$ is a compact set with boundary, then the set $f(A)$ is not always a compact set with boundary in W, but since f has a.e. positive Jacobian, the only phenomenon that could result in this happening is an "overlap" (see figure above). Using an isometry i of $\tilde{W}$ (of which there are many, since W is compact whereas $\tilde{W}$ is not) it is possible to move the self-intersection S a bit further, to apply the isoperimetric inequality to $f(A) \setminus S \cup i(S)$, and to conclude $\mathrm{vol}(f(A)) \leq \mathrm{const}(\mathrm{vol}(f(\partial A)))^{m/m-1}$. Since for a.e. $x \in V$,

$$\|Df_x\| \leq c \qquad \text{it follows that} \qquad \mathrm{vol}(f(\partial A)) \leq c^{\dim(V)}\,\mathrm{vol}(\partial A)$$

$$\|Df_x^{-1}\| \leq c^{-1} \qquad \text{it follows that} \qquad \mathrm{vol}(f(\partial A)) \geq c^{-\dim(V)}\,\mathrm{vol}(A),$$

and thus V satisfies the isoperimetric inequality

$$\mathrm{vol}(A) \leq \mathrm{const}(\mathrm{vol}(\partial A))^{m/m-1}.$$

We conclude that V has isoperimetric dimension greater than m.

The fundamental consequence of these results that will be used throughout the remainder of the chapter is the following:

6.11. Corollary: *Let $(V_1, g_1), (V_2, g_2)$ be Riemannian manifolds such that the former has dimension n and satisfies the hypotheses of Ahlfors' lemma and the latter has isoperimetric dimension strictly greater than n. Then there is no quasiregular map from (V_1, g_1) into (V_2, g_2).*

Proof. If there were such a map, then by Proposition 6.10 and Remark 6.3, the manifolds (V_1, pg_1) — for some function p — and (V_2, g_2) would satisfy isoperimetric inequalities with the same exponent.

6.12. Examples: The preceding hypotheses are satisfied if (V_1, g_1) is Euclidean space or a compact manifold with finite volume and if (V_2, g_2) is open at infinity.

In particular, by taking V_1 to be the complex plane and V_2 as the unit disk, we recover Liouville's theorem: "Every bounded entire function on $\mathbb{C}$ is constant." In fact, we have a more general version that applies to quasiconformal maps as well as conformal maps and uses no complex analysis.

Similarly, in anticipation of the remainder of the chapter, which is devoted to the computation of isoperimetric dimension, we can set $V_1 = \mathbb{C}$ and $V_2 = S^2 \setminus \{a, b, c\}$, for distinct a, b, c to obtain a quasiconformal version of the celebrated Picard theorem (cf. [Rudin]): "Every entire function on $\mathbb{C}$ which omits two values is constant." Our version of this theorem can, furthermore, be generalized to higher dimensions by taking $V_1 = \mathbb{R}^3$ and $V_2 = S^2 \setminus N$, where N is a nontrivial knot.

C. Computations of isoperimetric dimension

6.13. In this section, we will show that if V_2 is the universal cover of a compact Riemannian manifold V, then the existence of an isoperimetric inequality for V_2 depends only on the structure of the fundamental group of V. We must therefore introduce the notion of an isoperimetric inequality in a discrete group of finite type.

6.14. Definition: Let Γ be a group with a finite system of generators $\gamma_1, \dots, \gamma_p$, and let $\|\cdot\|$ be the algebraic (word) norm associated with this system (see 3.20). The *boundary* of a subset $A \subset \Gamma$ is defined as the set

$$\partial A = \{x \in A : \exists y \notin A \text{ with } \|yx^{-1}\| \leq 1\}.$$

If $I : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function, then we say that Γ satisfies the *isoperimetric inequality given by I* if there exist constants K, K' such that, for each finite subset $A \subset \Gamma$, we have

$$\operatorname{card}(A) \leq KI(K' \operatorname{card}(\partial A)).$$

In particular, Γ has isoperimetric dimension greater than m if Γ satisfies the isoperimetric inequality given by the function $I(t) = t^{m/m-1}$.

The group Γ is called *amenable* if it does not satisfy the isoperimetric inequality given by the function $t \mapsto t$.

In other words, a non-amenable group is the counterpart of a manifold that is open at infinity (see [Gre]).

6.15. Remark 1: The notion of boundary for subsets of Γ depends on the choice of generators, but the number of elements in the boundary is at worst multiplied by a constant when a different set of generators is chosen. It follows that the notion of isoperimetric inequality and related ideas do not depend on the choice of generating set, provided that it is finite.

6.16. Remark 2: Any non-amenable group must have exponential growth. Indeed, if $B(R) = \{x \in \Gamma : \|x\| \leq R\}$, then by hypothesis there is a constant c such that $\operatorname{card}(B(R)) = N(R) \leq c \operatorname{card}(\partial B(R))$, but $\partial B(R) \subset B(R) \setminus B(R-1)$, and so $N(R) \geq (c/c-1)N(R-1)$, hence $N(R) \geq (c/c-1)^R$. The converse is false in general: Solvable groups without nilpotent subgroups of finite index are amenable and have exponential growth (see [Wolf]$_{\text{Gro}}$, Thm. 4.8).

The following general lemma provides the best means for showing that a group is non-amenable.

6.17. Lemma: *Let Γ be a group of finite type equipped with the algebraic norm associated with a finite system of generators. Suppose that there is a mapping $f: \Gamma \to \Gamma$ such that, for each $x \in \Gamma$ we have i) $\|f(x)x^{-1}\| \leq 1$ and ii) $f^{-1}(x)$ contains at least two elements. Then Γ is non-amenable.*

Remark: The same conclusion holds if we replace "1" in condition (i) by an arbitrary constant K (taking $\{x : \|x\| \leq K\}$ as a new set of generators).

Proof. Given a finite subset $A \subset \Gamma$, set $A' = A \setminus \partial A$. If $x \in \Gamma \setminus A$ and $f(x) \in A$, then $f(x) \in \partial A$, since $\|f(x)x^{-1}\| \leq 1$, and thus $f(x) \notin A'$. In other words, $f^{-1}(A') \subset A$; thus by (2), $\operatorname{card}(f^{-1}(A')) \geq 2\operatorname{card}(A')$ so that $\operatorname{card}(A') \leq \operatorname{card}(A)/2$ and so $\operatorname{card}(A) \leq 2\operatorname{card}(\partial A)$.

6.17$\frac{1}{2}_+$ Exercise: Prove the converse: every non-amenable, finitely generated group Γ admits a self-mapping $f: \Gamma \to \Gamma$ such that $\operatorname{card} f^{-1}(x) \geq 2$ for all $x \in \Gamma$ and $\|f(x)x^{-1}\| \leq \text{const} = \text{const}_\Gamma$. (*Hint:* Show as a warm-up that every oriented manifold V open at infinity admits a *bounded* $(n-1)$-form η with $d\eta = (the volume form of V)$. This follows from the Hahn–Banach theorem and/or from Whitney-type duality between mass and comass, see Ch. 4.D. Then pass from $(n-1)$-forms to vector fields and from these to self-mappings.)

6.18. Corollary: *A free group having at least two generators is non-amenable.*

Proof. Indeed, the map f that erases the last letter of each word suffices.

Remark: A group containing a non-amenable subgroup is non amenable.

6.19. Theorem: *The fundamental group and the universal cover of a compact Riemannian manifold V satisfy the same isoperimetric inequality.*

Sketch of the proof. Let $\tilde{V}$ denote the universal cover of V and $\triangle$ a fundamental domain for the action of $\Gamma = \pi_1(V)$ on $\tilde{V}$. Suppose that Γ satisfies the isoperimetric inequality given by a function $I\colon \mathbb{R}_+ \to \mathbb{R}_+$. We first show (Lemma 1 below) that the isoperimetric inequality given by I for sets of the form $\bigcup_{\gamma\in B}\gamma\triangle$ follows immediately, and conversely, an inequality satisfied by $\tilde{V}$ is also satisfied by Γ. If A is now an arbitrary submanifold of V with boundary, then Lemmas 2 and 3 enable us to compare, for each $\gamma\in\Gamma$, $A\cap\gamma\triangle$ and $\partial(A\cap\gamma\triangle)$, and so to compare A and ∂A to a union of translates of $\triangle$.

6.20. Lemma 0: *There exist regular fundamental domains.*

Proof. Given a smooth triangulation of V, it is possible to choose disjoint liftings of each simplex of maximal dimension, thus providing a fundamental domain $\triangle$ (it is unimportant that $\triangle$ is not connected). The notions of volume for $\triangle$ and its boundary then make sense.

6.21. Lemma 1: *If $\tilde{V}$ satisfies an isoperimetric inequality, then the same is true for Γ. Conversely, if Γ satisfies the isoperimetric inequality given by a function I, then subsets of $\tilde{V}$ of the form $A = \bigcup_{\gamma\in B}\gamma\triangle$ satisfy* $\mathrm{vol}(A) \leq KI(K'\,\mathrm{vol}(\partial A))$ *for some constants K, K'.*

Proof. First we construct a good set of generators for Γ, setting

$$S = \{\gamma\in\Gamma\setminus\{1\} : \gamma\triangle \text{ and } \triangle \text{ share a common } (n-1)\text{-face}\}.$$

Note that $S' = \{\gamma\in\Gamma\setminus\{1\} : \gamma\triangle\cap\triangle \neq \emptyset\}$ generates Γ. Indeed, since Γ acts properly on $\tilde{V}$, it follows that for each $x\in\tilde{V}$ the set $\{\gamma\in\Gamma : B(x,1)\cap\gamma \neq \emptyset\}$

is finite, and so

$$r = \inf\{d(x, \gamma\triangle) : \gamma \in \Gamma, x \notin \gamma\triangle\} > 0,$$

and $B(x, r/2)$ only intersects $\gamma\triangle$ if $x \in \gamma\triangle$, i.e., $\gamma\triangle \cap \triangle \neq \emptyset$ for $\gamma \in S'$. It follows that $B(x, r/2) \subset S'\triangle$ and, if G' is the group generated by S', that $G'\triangle$ contains its $(r/2)$ neighborhood and is therefore open and closed. Since $\tilde{V}$ is connected, we have $G'\triangle = \tilde{V}$, and so $G' = \Gamma$. It suffices to show that the subgroup G generated by S contains S'. Let $\gamma \in S'$ and $x \in \triangle \cap \gamma\triangle$. There exist points $x_1 \in \triangle$ and $x_2 \in \gamma\triangle$ very close to x. If c is a path joining x_1, x_2, we can assume by transversality that c does not intersect any codimension-2 face of the translates of $\triangle$. Thus, c moves successively across the translates $\triangle, \gamma_1\triangle, \ldots, \gamma_k\triangle, \gamma\triangle$ by passing through a common $(n-1)$-face of $\triangle$ and $\gamma_1\triangle$, a common $(n-1)$-face of $\gamma_1\triangle$ and $\gamma_2\triangle$, etc..., which implies that $\gamma_1, \gamma_2\gamma_1^{-1}, \ldots, \gamma\gamma_k^{-1} \in S$, and thus $\gamma \in G$.

Let $A = \bigcup_{\gamma \in B} \gamma\triangle$. It is easy to check that for $\gamma \in B$,

$$\begin{aligned} &\{\gamma\triangle \text{ and } \partial A \text{ have an } (n-1)\text{-face in common}\} \\ &\qquad \Leftrightarrow \{\exists \gamma' : \|\gamma'\| = 1 \text{ and } \gamma'\gamma\triangle \not\subset A\} \\ &\qquad \Leftrightarrow \{\gamma \in \partial B\}. \end{aligned}$$

Set $K^{-1} = \min\{\mathrm{vol}(\triangle \cap \gamma\triangle) : \gamma \in S\}$. Then

$$\begin{aligned} \mathrm{vol}(\partial A) &\leq \mathrm{vol}(\partial\triangle) \cdot \mathrm{card}\{\gamma \in B : \gamma\triangle \text{ and } \partial A \\ &\qquad\qquad \text{have a common } (n-1)\text{-face}\} \\ &= \mathrm{vol}(\partial\triangle)\,\mathrm{card}(\partial B); \end{aligned}$$

$$\begin{aligned} \mathrm{vol}(\partial A) &\geq K^{-1}\,\mathrm{card}\{\gamma \in B : \gamma\triangle \text{and } \partial A \text{ have a common } (n-1)\text{-face}\} \\ &= K^{-1}\,\mathrm{card}(\partial B); \end{aligned}$$

$$\mathrm{vol}(A) = \mathrm{vol}(\triangle)\,\mathrm{card}(B).$$

In conclusion,

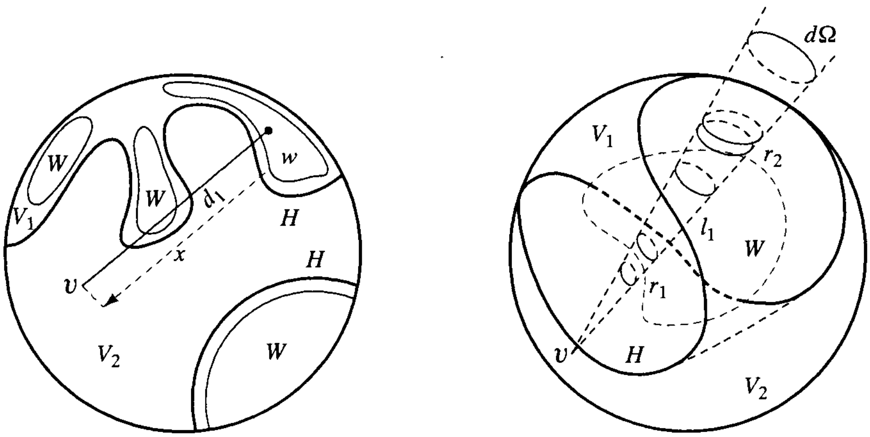
If $\mathrm{card}(B) \leq I(\mathrm{card}(\partial B))$, then $\mathrm{vol}(A) \leq \mathrm{vol}(\triangle) I(K\,\mathrm{vol}(\partial A))$,

If $\mathrm{vol}(A) \leq I(\mathrm{vol}(\partial A))$, then $\mathrm{card}(B) \leq (\mathrm{vol}(\triangle))^{-1} I(\mathrm{vol}(\partial\triangle)\,\mathrm{card}(\partial B))$.

Let B^n be the standard n-ball, V_1, V_2 two disjoint (not necessarily connected) open sets in B^n whose common boundary consists of a hypersurface H and such that $B^n = V_1 \cup V_2 \cup H$. We assume $\mathrm{vol}(V_1) \leq \mathrm{vol}(V_2)$.

6.22. Lemma 2: $\mathrm{vol}(V_1) \leq 2^{n+1}\,\mathrm{vol}(H)$.

Proof. We construct a subset $W \subset V_1$ and a point $v \in V_2$ such that



(*i*) $\mathrm{vol}(W) > \mathrm{vol}(V_1)/2$.

(*ii*) If $w \in W$ and if $d_v(w)$ is the distance between v and the last intersection of the segment $[v, w]$ with H, then $d_v(w) \geq d(v, w)/2$.

Set

$$U = \left\{(v, w) \in V_2 \times V_1 : d_v(w)/d(v, w) \geq \frac{1}{2}\right\}$$

$$T = \left\{(w, v) \in V_1 \times V_2 : d_w(v)h/d(v, w) \geq \frac{1}{2}\right\}$$

and define $i\colon V_2 \times V_1 \to V_1 \times V_2$ by $i(v, w) = (w, v)$. Note that $d_v(w) + d_w(v) \geq d(v, w)$, and so $\{d_v(w) \leq d(v, w)/2\}$ implies $\{d_w(v) \geq d(v, w)/2\}$, i.e.,

$$U^c \subset \left\{(v, w) \in V_2 \times V_1 : d_w(v)/d(v, w) \geq \frac{1}{2}\right\}$$

and $i(U^c) \subset T$. We conclude that

$$\mathrm{vol}(U) + \mathrm{vol}(T) \geq \mathrm{vol}(U) + \mathrm{vol}(i(U^c)) = \mathrm{vol}(U) + \mathrm{vol}(U^c) = \mathrm{vol}(V_2 \times V_1),$$

and one of the two volumes is $\geq \mathrm{vol}(V_1 \times V_2)/2$.

Suppose that $\mathrm{vol}(U) \geq \mathrm{vol}(V_2 \times V_1)/2$. Since $\mathrm{vol}(U) = \int_{V_2} \mathrm{vol}(U_v)\, dv$, the number

$$\int_{V_2} \left(\mathrm{vol}(U_v) - \frac{1}{2}\mathrm{vol}(V_1)\right)\, dv$$

is positive, where

$$U_v = \left\{w \in V_1 : d_v(w) \geq \frac{1}{2}\right\}.$$

Thus, the integrand is not always strictly negative; in other words, there exists $v \in V_2$ such that $\mathrm{vol}(U_v) \geq \mathrm{vol}(V_1)/2$, and we take $W = U_v$. Now we must find an upper bound for the volume of W as a function of the volume of H. Let $d\Omega$ be the volume form on the sphere of center v and radius 1. In spherical coordinates centered at v, at the distance r from v, the volume element of B^n is $r^{n-1}dr\, d\Omega$. Almost every ray intersects H transversely at a finite number of points $r_1, \dots, r_k$. Let I_i be the part of the ray within W lying between r_i and r_{i+1}. The volume element of W within the cone $d\Omega$ is

$$dW = d\Omega \sum_{i=1}^{k} \int_{I_i} r^{n-1}\, dr.$$

By Property *(ii)* of W, however, $r \geq 2d_V$, and, for a point of I_i, we have $d_V = r_i$, thus

$$dW \leq d\Omega \sum_{i=1}^{k} (2r_i)^{n-1} \int_{I_i} dr \leq 2^n \sum_{i=1}^{k} r_i^{n-1} d\Omega \leq 2^n ds,$$

the volume element of H lying in the cone $d\Omega$.

We conclude that $\mathrm{vol}(W) \leq 2^n \mathrm{vol}(H)$ and $\mathrm{vol}(V_1) \leq 2^{n+1} \mathrm{vol}(H)$. In the case when $\mathrm{vol}(T) \geq \mathrm{vol}(V_2 \times V_1)/2$, the same argument gives the inequality $\mathrm{vol}(V_2) \leq 2^{n+1} \mathrm{vol}(H)$, which provides the desired result.

6.23. Lemma 3: *If we define* $\partial_1 V = \partial V_1 \setminus \{H\}$, *then*

$$\mathrm{vol}(\partial_1 V) \leq n 2^{n+2} \mathrm{vol}(H).$$

Proof. For $x \in \partial_1 V$, we denote by $r(x)$ the distance from the origin to the first point of H along the ray issuing from x. Let α be the angle between the ray and the normal to H at this point. In this notation, we have

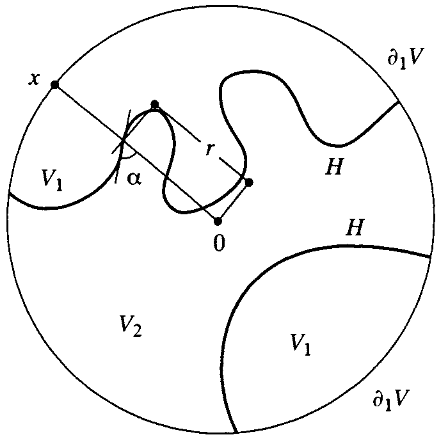
$$\mathrm{vol}(H) \geq \int_{\partial_1 V} \frac{r^{n-1}}{\cos(\alpha)}\, d\Omega \geq \int_{\partial_1 V} r^{n-1}\, d\Omega.$$

However,

$$\mathrm{vol}(V_1) \geq \int_{\partial_1 V} \left(\int_r^1 s^{n-1}\, ds \right) d\Omega = \int_{\partial_1 V} \frac{1 - r^n}{n}\, d\Omega,$$

and so

$$n\, \mathrm{vol}(V_1) \geq \mathrm{vol}(\partial_1 V) - \int_{\partial_1 V} r^n\, d\Omega \geq \mathrm{vol}(\partial_1 V) - \mathrm{vol}(H).$$



Since $\operatorname{vol}(V_1) \leq 2^{n+1}\operatorname{vol}(H)$ by Lemma 2, it follows that $\operatorname{vol}(\partial_1 V) \leq (n2^{n+1}+1)\operatorname{vol}(H)$, and so

$$\operatorname{vol}(\partial_1 V) \leq n2^{n+2}\operatorname{vol}(H).$$

6.24. End of the proof of Theorem 6.19. Note that the fundamental domain $\triangle$ constructed in Lemma 0 is the union of a finite number of simplices $\triangle_1, \ldots, \triangle_p$, each of which is the image of a ball under a quasi-isometric homeomorphism. Thus, Lemmas 2 and 3 are true for each $\triangle_i$ with the constants c_2, c_3.

Given a submanifold $A \subset \tilde{V}$ with boundary, we denote by A' the union of the translates of the $\triangle_i$, at least half of which are contained in A, i.e.,

$$A' = \bigcup_{i=1}^{p} \bigcup_{\gamma \in \Gamma} \left\{ \gamma\triangle_i : \operatorname{vol}(\gamma\triangle_i \cap A) \geq \frac{1}{2}\operatorname{vol}(\triangle_i) \right\}.$$

The sets $A, \partial A$ are divided between $\tilde{V} \setminus A'$ and A' into A_1, A_2 and $\partial_1 A, \partial_2 A$. Similarly, $\partial A'$ splits between A and $\tilde{V} \setminus A$ into $\partial_1 A$ and $\partial_2 A'$, which satisfy

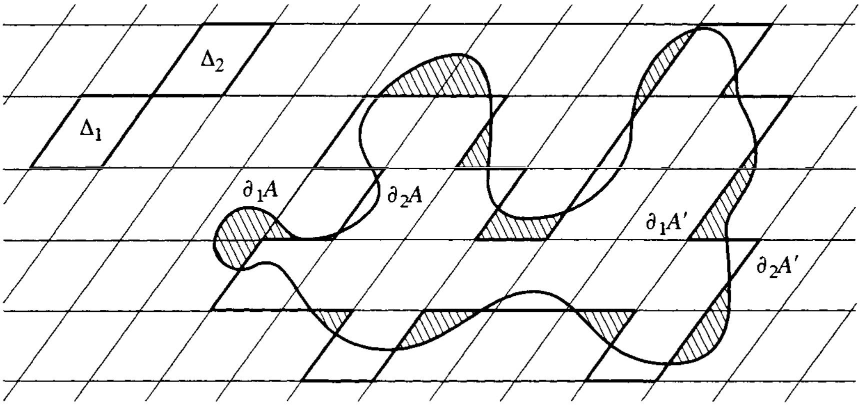
$$\operatorname{vol}(\partial_1 A') \leq c_3 \operatorname{vol}(\partial_1 A) \qquad \text{and} \qquad \operatorname{vol}(\partial_2 A') \leq c_3 \operatorname{vol}(\partial_2 A),$$

so that

$$\operatorname{vol}(\partial A') \leq c_3 \operatorname{vol}(\partial A)$$

$$\operatorname{vol}(A_1) \leq c_2 \operatorname{vol}(\partial_1 A) \leq c_2 \operatorname{vol}(\partial A) \qquad \text{and} \qquad \operatorname{vol}(A_2) \leq \operatorname{vol}(A').$$

If Γ satisfies the isoperimetric inequality given by the function I, then Lemma 1 implies $\operatorname{vol}(A') \leq KI(K' \operatorname{vol}(\partial A'))$ and so, since $\operatorname{vol}(A) \leq \operatorname{vol}(A')+$



$c_2 \operatorname{vol}(\partial A)$,

$$\operatorname{vol}(A) \leq KI(K'c_3 \operatorname{vol}(A)) + c_2 \operatorname{vol}(\partial A) \leq K''I(K''' \operatorname{vol}(\partial A)),$$

when I increases at least linearly, which is necessarily the case, for example, when I is an exponential function.

Examples:

6.25. The group $\mathbb{Z}^n$ has isoperimetric dimension n, since it is the fundamental group of a compact Riemannian manifold (flat torus) whose universal cover $\mathbb{R}^n$ has isoperimetric dimension n.

6.26. The fundamental group of a compact hyperbolic manifold is nonamenable, since the universal cover $\mathbb{H}^n$ is open at infinity. The same is true of the fundamental group of a compact, nonflat manifold with nonpositive sectional curvature, as follows from [Avez].

On the other hand, if V is simply connected, has $K(V) \leq 0$, and its Ricci curvature satisfies $\operatorname{Ricci}(g) \leq -\varepsilon g$ for some $\varepsilon > 0$, then V is open at infinity.

6.27. Conjecture: In fact, "nonpositive sectional curvature and strictly negative scalar curvature" should be sufficient.

6.28. Using only the hypothesis of nonpositive sectional curvature, D. Hoffman and J. Spruck have succeeded in proving that the isoperimetric dimension of V is greater than n (see [Osser], p. 1214, and [Hoff–Spru]).

Conjecture: *Under these hypotheses, we have the same inequality and the same constant as for Euclidean space.* This is known to be true for products of surfaces and manifolds of constant curvature.

$6.28\frac{1}{2}_+$. This conjecture has been solved in dimension 4 by Croke (see [Croke]$_{4DII}$) and for dimension 3 by Kleiner (see [Kle]). Consequently, it holds true for Cartesian products of these manifolds (we suggest that the reader actually work out the proof of the stability of the *sharp* Euclidean isoperimetric inequality under cartesian products). But the *sharp* isoperimetric inequality for general (complete, simply connected) manifolds V with $K \leq 0$ still remains conjectural for $\dim V \geq 5$, where "sharp" refers to the constant C_n in $\operatorname{vol} D \leq C_n(\operatorname{vol} \partial D)^{n/(n-1)}$, which is supposed to make this an equality for the Euclidean balls D. In fact, one expects even sharper results for particular classes of manifolds with $K(V) \leq 0$. For example, one wants to know the sharp inequality in symmetric spaces V (where it is unknown even for $\mathbb{R}\text{-rank}\, V = 1$), and one expects an isoperimetric comparison theorem between manifolds V with $K(V) \leq 0$ and $\operatorname{Ricci} V \leq -(n-1)$ and those with $K = -1$.

On the other hand, the non-sharp isoperimetric inequality of Hoffman and Spruck admits the following generalization (see p. 143 in [Gro]$_{FRM}$).

Let V be a complete Riemannian manifold where every $(k-1)$ dimensional cycle Y can be filled in by a k-chain Z such that

$$\operatorname{vol}_k Z \leq C(\operatorname{vol}_{k-1} Y)(\operatorname{diam} Y) \qquad (*)$$

for a given $C > 0$ and $k = 1, 2, \ldots, m$. Then every such Y can be filled in by Z satisfying

$$\operatorname{vol}_k Z \leq C'(\operatorname{vol}_k Y)^{k/(k-1)}, \qquad (**)$$

$k = 1, \ldots, m$, where $C' = C'(C, m)$.

In particular, if $m = n = \dim V$, then the inequality $(*)$ for all $k \leq n$ implies the ordinary isoperimetric inequality for domains $D \subset V$.

Remarks: (a) In order to have $(**)$, it is sufficient to prove $(*)$ only for "round" cycles Y, where $\operatorname{diam}(Y) \leq C_0(\operatorname{vol} Y)^{1/(k-1)}$ (see [Gro]$_{FRM}$).

(b) The sharp filling inequality $(**)$ is known for $V = \mathbb{R}^n$ from the work of Almgren (see [Alm]$_{Opt}$).

(c) The inequality $(*)$ (obviously) holds true for $K(V) \leq 0$ and, more generally, if V has *no focal points.* Additionally, it is likely that "no conjugate points" would be enough for $(**)$ (compare [Gro]$_{FRM}$).

6.29_+ Varopoulos' theorem: *A discrete, finitely generated group of finite type whose isoperimetric dimension is finite contains a nilpotent subgroup* *of finite index.* This was conjectured for solvable groups in the French edition of this book and proven by Varopoulos in full generality, see 6.E$_+$.

6.30. Remark: More precise results have been obtained for local homeomorphisms using *conformal moduli* of families of curves.

Theorem (Zoritch, see [Zor] or [Laur–Belin]): *Let V_1, V_2 be Riemannian manifolds of the same dimension* > 3, *and suppose that the fundamental group of V_1 is $\mathbb{Z}^n$ whereas V_2 is simply connected. If there is a quasiregular local homeomorphism $f : \tilde{V}_1 \to V_2$, then f is necessarily a homeomorphism of $\tilde{V}_1$ onto an open subset of V_2 whose complement has dimension* 0.

Theorem: *The same conclusion is true if $\tilde{V}_1$ is a Riemannian manifold that satisfies the hypothesis of Ahlfors' lemma, Lemma 6.9 (compare [Zor–Kes]).*

6.31. Corollary: *When $\tilde{V}_1 = \mathbb{R}^n$, then since homeomorphisms preserve ends (see [Bour]), and since $\mathbb{R}^n$ has only one end, the set $V_2 \setminus f(\tilde{V}_1)$ contains at most one point. Consequently, if V_2 is a simply connected manifold other than $\mathbb{R}^n$ and S^n, there is no quasiregular local homeomorphism $\mathbb{R}^n \to V_2$.*

D. Generalized quasiconformality

6.32. We are now tempted to define quasiregular mappings between manifolds of unequal dimensions.

There is a notion of a q-dimensional isoperimetric inequality within a Riemannian n-manifold where $n > q$ as follows: Given an increasing function $I : \mathbb{R}_+ \to \mathbb{R}_+$, the manifold V *satisfies the q-dimensional isometric inequality given by the function* $I = I(t)$ if, for every q-manifold S with boundary and for every map $f : \partial S \to V$, there is an extension of f to S such that $\operatorname{vol}(f(S)) \leq KI(K' \operatorname{vol}(f(\partial S)))$ for two constants K, K'.

For example, if, for each closed curve c of length ℓ in V, there exists an extension of c to a disk having area $\leq \operatorname{const}(\ell^{p/p1})$, then we say that V has *2-dimensional isoperimetric rank greater than p.* (Here, "rank" replaces "dimension," since the latter sounds awful here.)

6.33. Proposition: *Euclidean space has 2-dimensional isoperimetric rank equal to 2. Hyperbolic space has infinite 2-dimensional isoperimetric rank.*

Proof. Let c be a closed, connected curve of length ℓ in $\mathbb{R}^n$ and O a point of c. The cone k on c of origin O is an extension of $c\colon S^1 \to \mathbb{R}^n$ to

$$B^2 = \{re^{it} : 0 \le r \le 1, 0 \le t \le 2\pi\},$$

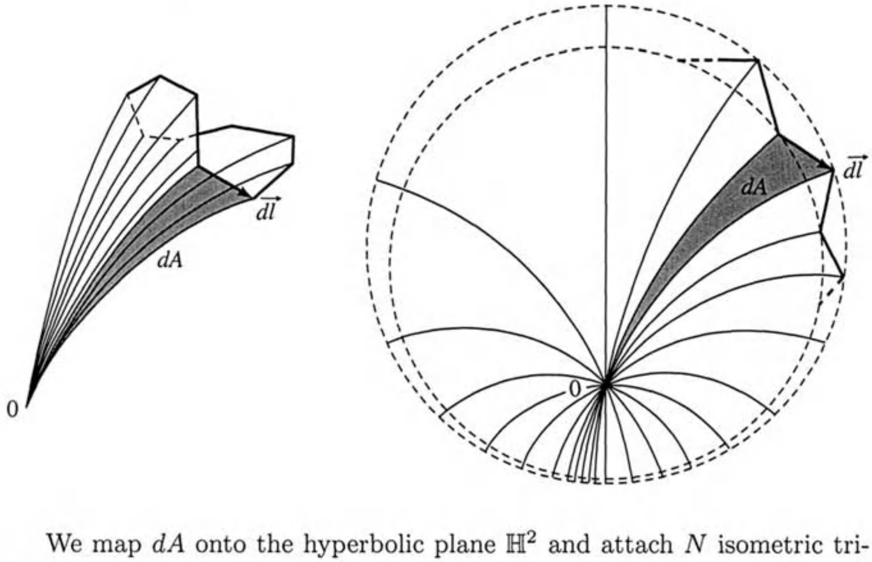
given by $re^{it} \mapsto K(re^{it}) = K(r,t) = O + rO(c(t))$. The area of K is

$$\begin{aligned}\int_0^{2\pi} dt \int_0^1 \left\| \frac{dK}{dr} \wedge \frac{dK}{dt} \right\| dr &\le \int_0^{2\pi} dt \int_0^1 \|Oc(t)\| \, \|c'(t)\| r \, dr \\ &\le \frac{\ell}{4} \int_0^{2\pi} \|c'(t)\| \, dt \le \frac{\ell^2}{4}.\end{aligned}$$

Thus, Euclidean space satisfies the inequality

$$\mathrm{vol}(K) \le \frac{1}{4} \, \mathrm{vol}(C)^2.$$

Given a curve c of length ℓ in $\mathbb{H}^n$, choose a point O of $\mathbb{H}^n$ lying at a large distance from c. The length element $d\ell$ on c determines an area element dA on the cone K with vertex O based at c.



We map dA onto the hyperbolic plane $\mathbb{H}^2$ and attach N isometric triangles to dA, in such a way to approximately cover a domain $D \subset \mathbb{H}^2$ whose boundary ∂D consists of N small mutually isometric segments in $\mathbb{H}^2$ (Proposition 6.6) $\mathrm{Area}(D) \le \mathrm{length}(\partial D)$, i.e., $N \, dA \le N \, d\ell$. We conclude that

$$\mathrm{Area}(K) = \int_c dA \le \int_c d\ell = \ell.$$

6.33.bis. If V^n is simply connected and has nonpositive sectional curvature, then the q-dimensional isoperimetric rank of V is greater than n for $q < n = \dim(V)$ (see [Hoff–Spru]).

An argument analogous to that of Lemma 6.9 enables us to define the notion of q-dimensional isoperimetric rank for a group Γ, provided that Γ admits a discrete cocompact action on a manifold $\tilde{V}$ with trivial homotopy groups in dimensions $< q$, i.e., $\pi_1(\tilde{V}) = 0$, $\pi_2(\tilde{V}) = 0, \dots, \pi_{q1}(\tilde{V}) = 0$. In particular, we have the following

6.34. Proposition-Definition: If V is a compact manifold with $\pi_i(V) = 0$ for $i = 2, \dots, q-1$, then its fundamental group and universal cover satisfy the same q-dimensional isoperimetric inequalities.

Corollary: *The universal cover of a compact manifold whose fundamental group is isomorphic to that of a hyperbolic manifold has infinite 2-dimensional isoperimetric rank.*

$6.34\frac{1}{2}{}_+$ Remarks. (a) Two dimensional isoperimetric inequalities make sense for all finitely generated groups Γ, where they measure the number of elementary steps needed to transform a word representing the identity element in Γ to the trivial word. For example, Γ has the *solvable word problem* if and only if the universal covering $\tilde{V}$ of V with $\pi_1(V) = \Gamma$ satisfies the 2-dimensional isoperimetric inequality with a *subrecursive* function I, which means that ent $I(\text{ent}\, t)$ is recursive, with "ent" standing for "integral part" (compare $7.8\frac{1}{2}$ and the references therein).

(b) If a group Γ (or a reasonable space) satisfies the *subquadratic* 2-dimensional inequality, i.e., where $I(t)/t^2 \to 0$ for $t \to \infty$ (which corresponds to 2-dimensional rank > 2 in our terminology), then, in fact, Γ satisfies the linear inequality with $I(t) = t$ (which makes the 2-dimensional isoperimetric rank infinite). The groups (and spaces) with this property are called *hyperbolic* and have been extensively studied in recent years.

(c) It is unclear how to approach the q-dimensional isoperimetric problem for $q \geq 3$ without the above condition $\pi_i = 0$ for $i \leq q-1$ (which was actually missing in the French edition of the book). Also, one should be more specific here on the allowed topology of manifolds (or rather pseudomanifolds) S involved in the definition. For example, one can stick to $(n-1)$-spheres which are allowed to be filled in (spanned) by q-balls. Or, we can look at $(q-1)$ tori filled by solid q tori, as suggested by the cartesian products $\Gamma_1 \times \Gamma_2 \times \cdots \times \Gamma_{q-1}$ (see [Al–Pr–Wa]).

These q-dimensional isoperimetric inequalities can yield nonexistence results for mappings in the spirit of the following

6.35. Theorem: *Let W be a compact Kähler manifold such that $\pi_2(W) = 0$ and such that $\pi_1(W)$ is hyperbolic (i.e., $\pi_1(W)$ has 2-dimensional isoperimetric rank strictly greater than 2). Then there is no nonconstant holomorphic map of $\mathbb{C}$ into $\tilde{W}$.*

Proof. Let ω denote the Kähler form of W. For a compact subset $A \subset \mathbb{C}$ and a C^1 mapping $f : A \to W$, the integral $\int_A \|f_*\xi\|$, where ξ is the constant 2-vector field $\frac{\partial}{\partial z} \wedge \frac{\partial}{\partial \bar{z}}$ on $\mathbb{C}$, is, by definition, the area of $f(A)$. In this notation, we have

$$\int_A f^*\omega = \int_A \langle \omega, f_*\xi_{f(x)} \rangle \, dx.$$

By Wirtinger's inequality, however,

$$\langle \omega, f_*\xi_{f(x)} \rangle \leq \frac{1}{2} \|f_*\xi\|$$

(see $[\text{Fed}]_{\text{GMT}}$, p. 40) with equality if and only if the simple 2-vector $f_*\xi$ can be written as $v \wedge iv$ for some tangent vector v. It follows that $\text{Area}(f(A)) \geq 2|\int_A f^*\omega|$. Since $\pi_2(W) = 0$, all extensions of a curve $\partial A \to W$ are homotopic, and so the integral $\int_A f^*\omega$ depends only on c. Thus, if f_0 is such an extension, then the area of f is necessarily no less than $|\int_A f^*\omega| = |\int_A f_0^*\omega|$, and this value is attained precisely by the holomorphic extensions. Given a nonconstant holomorphic map $f : \mathbb{C} \to W$, we denote by p the function $\|Df\|$. By Ahlfors' Lemma (Lemma 6.9), $(\mathbb{C}, pg_{eucl})$ has isoperimetric dimension less than 2; for each $m > 2$, there is a sequence A_i of compact sets in $\mathbb{C}$ such that

$$\text{vol}_{pg}(A_i) \to \infty \qquad \text{and} \qquad \frac{\text{vol}_{pg}(A_i)}{(\text{vol}_{pg}(\partial A_i))^{-m/(m-1)}} \to \infty$$

as $i \to \infty$. Since f is holomorphic,

$$\text{vol}(f(A_i)) = \text{vol}_{pg}(A_i) \qquad \text{and} \qquad \text{vol}(f(\partial A_i)) = \text{vol}_{pg}(\partial A_i)$$

and f is the extension of $f|_{\partial A_i}$ to A_i having minimal area. We conclude that W has 2-dimensional isoperimetric dimension less than m.

The preceding theorem suggests a way to define quasiregular mappings from one manifold into another manifold of larger dimension: The minimal property for holomorphic mappings between Kähler manifolds should be replaced by a "quasiminimal" property.

6.36. Definition: Let V_0 be a compact manifold with boundary. A mapping $f: V_0 \to W$ is called *C-quasiminimal* if, for each map $g: V_0 \to W$ such that $f|_{\partial V_0} = g|_{\partial V_0}$ and such that the chains $f(V_0), g(V_0)$ are homologous in W modulo the boundary $f(\partial V_0) = g(\partial V_0)$, we have $\mathrm{vol}(f(V_0) \leq C\,\mathrm{vol}(g(V_0))$.

If V is an arbitrary manifold, then a mapping $f: V \to W$ is called *quasiminimal* if there is a constant C such that the restriction of f to each compact submanifold of V with boundary is C-quasiminimal.

6.37. Example: If $f: V \to W$ is quasiregular (with $\dim(V) = \dim(W)$), then the graph $Gf: V \to V \times W$ given by $G(x) = (x, f(x))$ is quasiminimal.

To see this, let pr_1, pr_2 be the projections onto the factors of the product $V \times W$, and let ω_1, ω_2 be the volume forms on V, W. Then $\omega = pr_1^*\omega_1 + pr_2^*\omega_2$ is a closed form on $V \times W$ satisfying the inequality

$$\begin{aligned} |\langle \omega, X_1 \wedge \cdots \wedge X_n \rangle| &\leq \|pr_{1*}(X_1 \wedge \cdots \wedge X_n)\| + \|pr_{2*}(X_1 \wedge \cdots \wedge X_n)\| \\ &\leq 2\|X_1 \wedge \cdots \wedge X_n\|, \end{aligned}$$

for each simple n-vector $X_1 \wedge \cdots \wedge X_n$ of $V \times W$.

In particular, if $g: V_0 \to V \times W$ and if ξ denotes the n-vector of V dual to ω_1, in the sense that $\langle \omega_1, \xi \rangle = 1$, then

$$\begin{aligned} \mathrm{vol}(g(V_0)) &= \int_{V_0} \|g_*\xi(x)\|\,dx \\ &\geq \frac{1}{2}\int_{V_0} |\langle \omega, g_*\xi(x) \rangle|\,dx \\ &\geq \frac{1}{2}\left|\int_{V_0} g^*\omega\right| \\ &= \frac{1}{2}\left|\int_{g(V_0)} \omega\right| \\ &= \frac{1}{2}\left|\int_{Gf(V_0)} \omega\right|, \end{aligned}$$

when the chains $g(V_0)$ and $Gf(V_0)$ are homologous, since ω is closed. This integral equals $(1/2)\int_{V_0}(1 + J(f,x))\,dx$. We now write $\xi = v_1 \wedge \cdots \wedge v_n$. Then

$$\|Gf_*\xi\| = \|(v_1 + Df_{v_1}) \wedge \cdots \wedge (v_n + Df_{v_n})\|$$

is less than or equal to a polynomial of degree n in $\|Df\|$, and so there are constants A, B such that $\|Gf_*\xi\| \leq A + B\|Df\|^n$ so that $\mathrm{vol}(Gf(V_0)) \leq$

$A\,\mathrm{vol}(V_0) + B\int_{V_0} \|Df_x\|^n\,dx$. If f is C-quasiregular, then

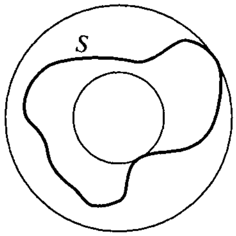
$$\mathrm{vol}(Gf(V_0)) \le C'\left(\frac{1}{2}\mathrm{vol}(V_0) + \frac{1}{2}\int_{V_0} J(f,x)\,dx\right) \le C'\,\mathrm{vol}(g(V_0)),$$

for $C' = \sup(2A, 2CB)$, and so Gf is quasiminimal.

6.37$_+$ It is unclear what a suitable notion of "quasiconformality" should be for maps $V \to W$, where $\dim W < \dim V$. One general possibility is to require the image of every small ball $B_V \subset V$ to be a *quasi-ball* in W, i.e., pinched between two concentric balls in W,

$$B_w(\rho) \subset f(B_v) \subset B_w(C\rho),$$

where $w = f(v)$ and $C = C(f) > 0$ is some constant.



For example, if $f\colon V \to W$ is a smooth *submersion*, then this is equivalent to a bound on the norm of the differential by the Jacobian of f on the subbundle $S \subset T(V)$ normal to the kernel $\ker Df \subset T(V)$ at each point of V,

$$\|D^*D\|_v^n \le \mathrm{const}\,\det(D^*D|_{S_v}),$$

$v \in V$. But to get off the ground, one should work out basic examples, e.g., to show that *no* such map $f: \mathbb{R}^{n+1} \to \mathbb{R}^n$ has a *bounded* image for $n \ge 2$ (or maybe 3). (Since this is unknown, one may strengthen the assumptions above by trying to rule out some particular maps $f: \mathbb{R}^{n+1} \to B^n \subset \mathbb{R}^n$, such as smooth fibrations where, moreover, the norm of the differential does not change much along each fiber $f^{-1}(x) \subset \mathbb{R}^{n+1}$.)

Another possibility is to fix some n-dimensional subbundle $S \subset T(V)$ *beforehand* and to study maps $f\colon V \to W = W^n$ which are *quasiconformal in* S. Here one may look at the contact subbundle S on the $(2n+1)$-dimensional Heisenberg group H_{2n+1} and again try to rule out *bounded* maps $H_{2n+1} \to \mathbb{R}^{2n}$.

Remark $_+$: It seems difficult (if at all possible) to make up a meaningful local metric definition of quasiconformality for maps $V \to W$, where

$\dim V < \dim W$. On the other hand, the above "quasi-ball definition" extends to arbitrary metric spaces, where it turns out to be quite productive, as for example in the case of Carnot–Caratheodory spaces.

6.38. Example: Let Ω be a pseudoconvex open subset of $\mathbb{C}^2$. In each tangent space $T_x\partial\Omega$, the set of vectors preserved under multiplication by $i = \sqrt{-1}$ is a 2-plane. Thus, we obtain a generally nonintegrable field of 2-planes on $\partial\Omega$ with which we can associate a "Carnot" metric (see 1.18 and 3.17).

> *If Ω is the ball B^4, then this metric on $S^3 = B^4$ is conformally equivalent to the metric introduced in 3.17 on the Heisenberg group.*

Indeed, let B^4 be equipped with its Bergman metric (see [Weil]), so that B^4 is isometric to the complex hyperbolic plane,

$$\mathbb{H}^2_{\mathbb{C}} = (U(1) \times U(2))\backslash U(1,2).$$

More precisely, $U(1,2)$ is the group of matrices or endomorphisms of $\mathbb{C}^3$ that preserve the Hermitian form

$$q(x) = -|x_1|^2 + |x_2|^2 + |x_3|^2,$$

and so $U(1,2)$ operates on the set

$$H = \{x \in \mathbb{C}^3 : q(x) = -1\},$$

and thus on the projection

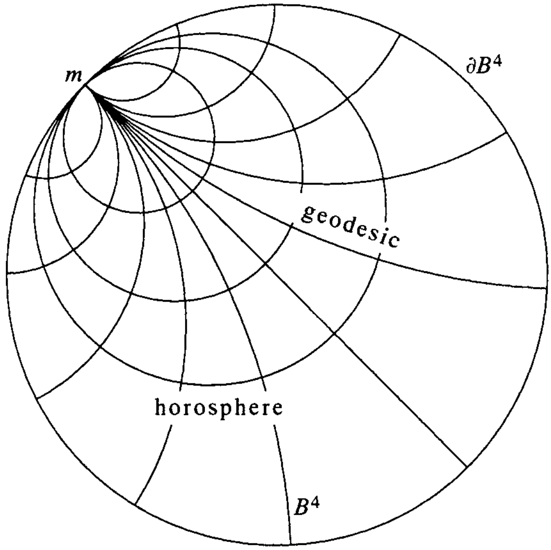
$$p(H) = B^4 \subset \mathbb{C}^2 \subset \mathbb{C}\mathrm{P}^2.$$

The action of $U(1,2)$ also preserves $\partial B^4 = p(\{q = 0\})$ and the field of 2-planes on ∂B^4 defined above. We now show that the stabilizer of a point of ∂B^4 contains a subgroup isomorphic to the Heisenberg group. Let J be the matrix of the quadratic form q, i.e.,

$$J = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The Lie algebra $\mathfrak{u}(1,2)$ of $U(1,2)$ is the set of 3×3 matrices M such that $M^*J + JM = 0$. Such a matrix can be written as

$$M = \begin{pmatrix} ir & a & b \\ a & is & -c \\ b & c & it \end{pmatrix},$$



where $a, b, c \in \mathbb{C}$ and $r, s, t \in \mathbb{R}$. Choose a point m in the sphere, for example $m = p(v)$, where $v = (1, 1, 0) \in \mathbb{C}^3$. The stabilizer F of m is the subgroup $\{g \in U(1,2) : gv \text{ is colinear to } v\}$; thus its Lie algebra is

$$\begin{aligned}\mathfrak{f} &= \{M \in \mathfrak{u}(1,2) : Mv \text{ is colinear to } v\} \\ &= \{M \in \mathfrak{u}(1,2) : b = -c, r = s, a \in i\mathbb{R}\}.\end{aligned}$$

Set

$$X = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & -i \\ i & -i & 0 \end{pmatrix} Z = \begin{pmatrix} i & -i & 0 \\ i & -i & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

so that $[X, Z] = [Y, Z] = 0$ and $[X, Y] = Z$, so that the real vector space generated by X, Y, Z is a Lie subalgebra $\mathfrak{n}$ of $\mathfrak{f}$ isomorphic to the Lie algebra of the Heisenberg group. From this we can deduce that the homomorphism φ from the Heisenberg group N into $U(1,2)$ whose differential at the origin sends the plane orthogonal to the center onto the plane spanned by X, Y, which is a complex line in $\mathfrak{u}(1,2)$. The mapping $j : N \to \partial B^4$ given by $j(n) = \varphi(n)(m')$ for a point $m' \in \partial B^4$.

It remains to show that j is a homeomorphism onto $\partial B^4 \setminus \{m\}$, which is simple, and that it is conformal, which follows from the formula $\|j_*(\alpha X + \beta Y)\|^2 = 3(\alpha^2 + \beta^2)$ at the origin and from the fact that N acts by homographies, which are conformal transformations of $\mathbb{C}\mathrm{P}^n$.

6.39. Remark: Since all contact structures in dimension 3 are locally isomorphic, it follows that for any pseudoconvex domain $\Omega \subset \mathbb{C}^2$, the Carnot–Caratheodory metric on $\partial\Omega$ is locally quasi-isometric to that of S^3.

6.40. Question: Under what conditions on Ω does there exist a locally quasiconformal homeomorphism from the Heisenberg group into $\partial\Omega$ when both are equipped with their Carnot–Caratheodory metrics?

6.41_+ Remark. A full-fledged quasiconformal theory for C-C geometry is developed in [Kor–Rei], see also $[\mathrm{Gro}]_{\mathrm{CC}}$.

6.42_+ The Bloch-Brody principle for quasiconformal maps. This principle, discovered I believe by A. Bloch a long time ago and revived by Brody in recent times, allows one to obtain, by some limiting process, a *uniformly continuous* map $\mathbb{R}^n \to W$ quasiconformal in some sense, starting from a nonuniformly continuous, quasiconformal map or a suitable sequence of maps $V_i \to W$. To see how it works, imagine for the moment that the maps in question are C^1-smooth (e.g. holomorphic from $\mathbb{R}^2 = \mathbb{C}$ to a compact complex manifold as in the original papers by Bloch and Brody) and try to make a given $f \colon \mathbb{R}^n \to W$ Lipschitz by moving the origin to the point v_0 where the norm of the differential $\|Df\|_v$ assumes its maximum. Of course, this does not work if the function $\|Df\|_v$ is unbounded on $\mathbb{R}^n$, but we can artificially make it bounded by restricting f to a large ball $B(R)$ in $\mathbb{R}^n$ about the origin and taking the complete hyperbolic metric g_R with curvature -1 on this $B(R)$. Since g_R blows up at the boundary of $B(R)$, the norm of the differential Df on $B(R)$ with respect to g_R does assume its maximum at some point, say $v_R \in B(R)$, and by rescaling $(B(R), g_R)$ with $\lambda = \|Df(v_R)\|_{g_R}$, we shall make $\|Df(v_R)\| = 1$ for the metric λg_R. Then we take a sequence $R_i \to \infty$ and go to the pointed Lipschitz limit

$$\lambda_i(B(R_i), g_{R_i}) \xrightarrow{\text{Lip}} \mathbb{R}^n$$

for $i \to \infty$ with 1-Lipschitz maps $f_i = f|_{B(R_i)}$ hopefully converging, perhaps after passing to a subsequence, to some (*nonconstant!*) map f_∞ : $\mathbb{R}^n \to W$ where $\|Df_\infty\|$ assumes its maximum at the origin. Now, in the situations we deal with, the nonconstancy of f_∞ is automatic. The only problem which needs attention is the possibility that $f_i(v_{R_i}) \to \infty$ in W. This, however, does not present itself if W is compact or, more generally, when $W/\operatorname{Isom}(W)$ is compact, since $f_i(v_{R_i})$ can be moved to a compact region of W. Furthermore, if $W/\operatorname{Isom}(W)$ is noncompact, but, for example, W is complete with *bounded local geometry*, then we also can go to the Lipschitz limit of (a subsequence of) the sequence (W, w_i), $w_i = f_i(v_{R_i})$ as $i \to \infty$ and come up with $f : \mathbb{R}^n \to W_\infty$. And even if W has unbounded geometry but still satisfies some asymptotic finiteness property, say $\sup_{w \in W} K_w(W) < \infty$, we can pass to the pointed Hausdorff limit of (W, w_i) with some "virtual quasiconformal map" f_∞ from $\mathbb{R}^n$ to some "virtual n-manifold" fibered over W_∞ of (possibly) $\dim W_\infty < n$.

This works, as it stands for holomorphic maps from $\mathbb{C}$, and more generally for *pseudoholomorphic* maps to quasicomplex manifolds. Here we are interested in quasiconformal maps where the above leads to the following conclusion.

$6.42\frac{1}{2}{}_+$ Proposition. *Let W be a complete Riemannian manifold such that $W/\operatorname{Isom}(W)$ is compact. If there exists a nonconstant quasi-regular (quasiconformal) map $f : \mathbb{R}^n \to W$, then there also exists a uniformly continuous quasiregular (respectively, quasiconformal) map $f_\infty : \mathbb{R}^n \to W$ which is necessarily uniformly Hölder for some exponent α.*

Remarks: This has several generalizations. For example, instead of having a single map f, we can start with a sequence $f_i\colon V_i \to W$ where all the V_i are complete manifolds with $|K(V_i)| \leq \text{const} < \infty$. Then, assuming that these f_i are not together (as a family) uniformly continuous (i.e., equicontinuous), we can produce $f_\infty : \mathbb{R}^n \to W$ as above. Furthermore, we can apply the same to quasiminimal quasiconformal non-equidimensional maps $\mathbb{R}^n \to W^{n+k}$. Finally, this can be used in the Carnot-Caratheodory framework, since we do not truly need hyperbolic metrics in domains (e.g., balls) $U \subset V$. Actually, a good enough metric g_U is obtained by blowing up $g_V|_U$ by the conformal factor $\delta(U) = \operatorname{dist}(u, \partial U)$ (or $\delta^\alpha(u)$ with any $\alpha > 0$ for this matter). All one needs to check is (local) precompactness of these metrics as well as the scaled $\lambda(U, \delta g_V)$ for $\lambda \geq 1$, in the pointed Lipschitz topology. (We suggest that the reader think through all of this by him/herself.)

$6.42\frac{2}{3}{}_+$ Problems. (a) Are closed manifolds $W = W^n$ receiving quasiregular maps $\mathbb{R}^n \to W$ elliptic in the sense of Ch. $2.\mathrm{E}_+$? Is the converse true?

(b) Can we derive the Rickman-Picard theorem (see [Rickm]) for quasiregular maps f by a purely isoperimetric argument? Notice that the Hölder property of f does follow in this way (see §2.6 in $[\mathrm{Gro}]_{\mathrm{CC}}$) and thus extends to quasiconformal quasiminimal maps $f\colon \mathbb{R}^n \to W^{n+k}$, but it is unknown if Rickman's theorem does as well.

(c) Let V and W be equidimensional Riemannian homogeneous spaces. When does there exist a quasiconformal (quasiregular) map $f : V \to W$? Does the existence of this make the isoperimetric dimension of V greater than that of W? The same question arises for V with compact $V/\operatorname{Isom} V$ and an arbitrary W.

E_+ The Varopoulos isoperimetric inequality

6.43. There is the following quite popular general strategy for bounding the volume of ∂D. Suppose we have means of transportation of mass from D to the complement $V \setminus D$. In the course of this, we necessarily cross ∂D, and we can sometimes estimate the amount of mass passing through ∂D by $\operatorname{vol} \partial D$. Thus, we can bound $\operatorname{vol} D$ provided that we have a lower bound on the percentage of the total mass (volume) of D that is being transported.

Examples: (a) Suppose our "transportation" is given by a vector field X on D. This is seen as an "infinitesimal transport," where the mass crossing the boundary per unit time is obviously bounded by $\int_{\partial D} \|X\|$. On the other hand, the infinitesimal percentage of the transported mass equals $\int_D \operatorname{div} X$, and by Stokes' formula,

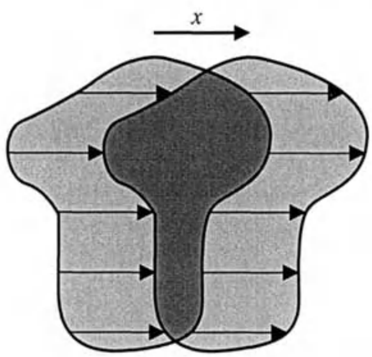
$$\int_D \operatorname{div} X \le \int_{\partial D} \|X\| \le \sup \|X\| \operatorname{vol}(\partial D).$$

This gives a bound on $\operatorname{vol} D$ if $\operatorname{div} X \ge \varepsilon > 0$, and $\|X(v)\| \le C$, that is,

$$\operatorname{vol} D \le \varepsilon^{-1} C \operatorname{vol}(\partial D)$$

(we already saw how this applies to domains in hyperbolic space, for example).

(b) Take some $D \subset \mathbb{R}^n$ and translate it by a vector $x \in \mathbb{R}^n$.



The amount of mass transported through ∂D in this way is obviously bounded by $\|x\| \operatorname{vol} \partial D$, and to bound $\operatorname{vol} D$ by $\operatorname{vol} \partial D$, we need a vector x which moves much of D outside of D. To find such an x, we consider *all* x with $\|x\| \le d$ and observe that the average transport of mass can be computed as follows. Let φ denote the characteristic (indicator) function of D, and let φ_d be the smoothing of φ by means of the *smoothing kernel*

$$S(x,y) = \begin{cases} \alpha_d^{-1} & \text{for } \operatorname{dist}(x,y) \leq d \\ 0 & \text{for } \operatorname{dist}(x,y) \geq d, \end{cases}$$

where $\alpha_1 d^n$ denotes the volume of the d-ball $B(d)$ in $\mathbb{R}^n$. Then the above transport, say T_x, averaged over $x \in B(d)$ (obviously) equals

$$\int_D |\varphi(y) - \varphi_d(y)|\, dy, \tag{$*$}$$

where

$$\varphi_d(y) = \alpha_d^{-1} \operatorname{vol}(B_y(d) \cap D)$$

for $y \in \mathbb{R}^n$. Now let us choose d such that $\alpha_d = \operatorname{vol} B(d) = 2 \operatorname{vol}(D)$. Clearly, $\varphi_d(y) \leq 1/2$ for all $y \in D$, and so at least half of the mass of D is transported away from D by x with $\|x\| \leq d$. It follows that some x with $\|x\| \leq d$ gives us the inequality

$$d \operatorname{vol} \partial V \geq \frac{1}{2} \operatorname{vol} D,$$

where

$$2 \operatorname{vol} D = \alpha_d = \nu_n d^n$$

for $\nu_n = \operatorname{vol}(B(1))$. Thus,

$$(2\nu_n^{-1} \operatorname{vol} D)^{1/n} \operatorname{vol}(\partial V) \geq \frac{1}{2} \operatorname{vol}(D),$$

or

$$\operatorname{vol} D \leq \gamma_n (\operatorname{vol} \partial D)^{n/(n-1)} \tag{$+$}$$

for $\gamma_n = (2(2\nu_n^{-1}))^{1/(n-1)}$, which is a rather neat proof (due to Minkowski, I guess) of the isoperimetric inequality in $\mathbb{R}^n$, albeit with a nonsharp constant.

Exercise: Improve the above constant by a better evaluation of the mass crossing through ∂D *averaged* over $x \in B(d)$ with the angle between x and ∂D taken into account.

Now we come to the subject matter of this section, i.e., to a finite subset D in a group Γ equipped with the word metric associated to a finite generating subset in Γ. We transport D by using *left* translates by $\gamma \in \Gamma$, where we recall that our metric $\|\alpha\beta^{-1}\|$ in Γ is *right* invariant. Every left translate $\delta \mapsto \gamma\delta$, $\delta \in D$, moves δ by the distance $\|\gamma\|$, since

$$\operatorname{dist}_\Gamma(\delta, \gamma\delta) = \|\delta(\gamma\delta)^{-1}\| = \|\gamma\|,$$

and the amount of mass transported away from D does not exceed $\|\gamma\|\operatorname{card}(\partial D)$. Indeed, γ can be decomposed into $\|\gamma\|$ "elementary moves" corresponding to the generators of Γ, where each elementary transport lands in the boundary of D, since these move the points by distance one (where we recall that $\partial D = \{\alpha \in \Gamma : \operatorname{dist}(\alpha, D) = 1\}$). Then we average these γ-transportations over all $\gamma \in B(d) \subset \Gamma$ for the minimal d-ball $B(d)$ having $\operatorname{card} B(d) \geq 2 \operatorname{card} D$, as in (b) above. This immediately leads to the inequality

$$d \operatorname{card}(\partial D) \geq \frac{1}{2} \operatorname{card} D, \tag{$\star$}$$

where $\operatorname{card} B(d) \approx 2 \operatorname{card} D$. For example, if the group Γ is nilpotent and the balls have polynomial growth of degree n,

$$\operatorname{card} B(d) \simeq d^n,$$

then our d in $(\star)$ is $\simeq (\operatorname{card} D)^{1/n}$, which gives the expected inequality

$$\operatorname{card} D \leq \operatorname{const}(\operatorname{card} \partial D)^{n/(n-1)} \tag{$\star$}_n$$

proven earlier by Pansu for the Heisenberg group (see [Pan]$_{\text{IIGH}}$).

Furthermore, if Γ is not virtually nilpotent, and hence the balls grow faster than polynomially (see [Dri–Wil]), then $(\star)_n$ holds for *all* n, which means that Γ has infinite isoperimetric dimension. And, if Γ has exponential growth (e.g. being solvable but not virtually nilpotent), then we take $d \simeq \log(\operatorname{card}(D))$ and get the isoperimetric bound

$$\operatorname{card}(\partial D) \geq \operatorname{const} \operatorname{card}(D)/\log(\operatorname{card}(D)).$$

Remarks. (*a*) All these isoperimetric inequalities in Γ are due to Varopoulos, who originally used a more complicated transportation associated to the random walk in Γ (see [Va–Sa–Co] and [Pit–Sal], where stronger results can be found). The simple proof above appears in [Coul–Sal] and in [Coul]. Another simple transportation argument can be found in [Gro]$_{\text{CC}}$.

(*b*) One can transport the measure away from D following the geodesic flow in the unit tangent bundle $U(V) \to V$, where D is first lifted to $U(V)$, then transported according to geo_t for some $t \in [0, d]$ and finally projected back to V. This idea goes back to Santalo and leads to interesting corollaries whenever we can show that a sufficient amount of mass goes away from D for a given d. For example, if V is complete, simply connected, and without conjugate points, then every geodesic of length $\geq \operatorname{diam} D$ starting in D leaves D. It follows easily that

$$\operatorname{vol} D \leq \operatorname{const}_n(\operatorname{diam} D)(\operatorname{vol} \partial D).$$

Then, if we want to exclude diam D, we need *a priori* an extra lower bound on the amount of mass transported by geo_t, which can be achieved, for example, with the Jacobian of the exponential map $E: T_v(V) \to V$. For example, if

$$\| \,\mathrm{Jacobian}\, E_v(x) \| \geq C_0,$$

then

$$\mathrm{vol}\, D \leq \mathrm{const}(n, C_0)(\mathrm{vol}\, \partial D)^{n/(n-1)}$$

as the logic of our argument shows (see [Gro] for better results).

(c) There is nothing sacred about the geodesic flow in this argument. One could take an arbitrary measure space U fibered over V and try some measure-preserving flow in U. This works very nicely, for example, for the time-shift flow in the space of (Brownian) paths in V with the Wiener measure (as was observed by J. Moser years ago) and relate this isoperimetric inequality to the rate of decay of the heat flow in V.

Problem: Is there a *necessary and sufficient* condition for a given isoperimetric inequality in V in terms of the existence (or nonexistence) of the above flow or just some (kernel) functions $S_d(x, y)$ in V? (We do know in this regard that being open at infinity, i.e., satisfying $\mathrm{vol}\, D \leq C \,\mathrm{vol}\, \partial D$ is equivalent to the existence of a $(n-1)$-form η with $\|\eta\|_v \leq C$ such that $d\eta$ is a volume form. This follows by an application of the Hahn- Banach theorem (as in the Whitney mass-comass duality of our section 4.D), which was indicated to me by D. Sullivan about twenty years ago.)

Chapter 7

Morse Theory and Minimal Models

A. Application of Morse theory to loop spaces

7.1. In Chapter 6, we introduced the notion of 2-dimensional isoperimetric rank (6.32) as the largest number p, such that each simple curve of length ℓ is bounded by a disk of area at most $C\ell^{p/p-1}$. This definition only makes sense in noncompact manifolds, and we have shown that the 2-dimensional isoperimetric rank of the universal cover of a compact manifold V depends only on the fundamental group $\pi_1(V)$.

We can also ask if it is possible to deform a given loop ℓ to a point without increasing its length, or without increasing its length beyond $f(\ell)$, where the function f would arise as an invariant of the geometry of the (universal cover of the) manifold. The answer to this question may vary depending on V: For the universal cover of a manifold of nonpositive curvature, we can take $f(\ell) = \ell$. These manifolds, although cocompacts of very different groups, are the simplest possible from the point of view of the invariant f. On the other hand, we will see that, when a universal cover or fundamental group is so complicated that the word problem cannot be solved, then the function f is not computable. Between these two extremes, it is unclear whether we can say much about the invariant f in the noncompact case. Nevertheless, the invariant f makes sense for a compact manifold V, as illustrated in Figure 2.26, for which we already have several results: Scolie 2.27 shows that $f(\ell) \leq C(V)\ell$, while Proposition 2.26 gives

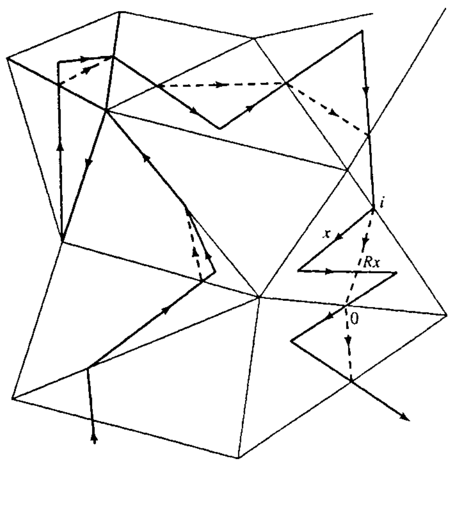
$f(\ell)/\ell \to 1$ as $\ell \to \infty$.

It is interesting to compare the topology of all the spaces X_ℓ of loops of length $\leq \ell$ and not just that of their connected components.

7.2. We now introduce some notation. Let X (resp. Y) be the space of piecewise C^∞ paths connecting w to w' in a compact Riemannian manifold W (resp. the space of closed, piecewise C^∞ curves without base point in W), both equipped with the C^∞ topology, so that the length function L is continuous. We denote by $X_t = L^{-1}([0,t])$ (resp. Y_t) the space of paths (resp. curves) of length at most t. Finally, we let $dm(t)$ be the largest integer d such that for each i with $0 \leq i \leq d$, the homomorphism in homology induced by inclusion $X_t \hookrightarrow X$ (resp. $Y_t \hookrightarrow Y$) is surjective.

7.3. Theorem: *If W is a compact, simply connected Riemannian manifold with path space X and space Y of unbased loops, there are constants c, C such that $ct \leq dm(t) \leq Ct$.*

Proof. Fix a triangulation Tr of W and let X' denote the space of piecewise linear paths for this triangulation, where we assume that the endpoints w and w' are vertices (and where the case Y of closed curves is left to the reader so that we do not have to carry along the omnipresent "resp."). Every path in X' is given by a sequence of points $w = w_1, w_2, \ldots, w_k = w'$ in W, where every two neighboring points w_i and w_{i+1} lie in one simplex Δ of Tr and are joined in Δ by the (unique) straight segment.



Thus,

$$X' = \bigcup_{k=1}^{\infty} X'_k,$$

where X'_k denotes the space of paths with at most k vertices. Also observe that for every sequence of simplices $\Delta_1, \dots, \Delta_k$ in Tr, the paths with the vertices $w_i \in \Delta_i$ make a cell in X'_k, namely $\Delta_1 \times \Delta_2 \times \cdots \times \Delta_k$. Thus, X' comes along with a natural cell decomposition Tr' with X'_k being contained in the k_n-skeleton Tr_{kn} of Tr'.

7.3$\frac{1}{2}$$_+$ Lemma. *Every continuous map f of a d-dimensional polyhedron P into a compact X can be homotoped to a map whose image lies in the d-skeleton $Tr'_d \subset X' \subset X$.*

Proof. Maps $f\colon P \to X$ correspond to maps $F\colon P \times [0,1] \to W$, and these admit simplicial approximations $F' : P \times [0,1] \to W$ for sufficiently fine triangulations of $P \times [0,1]$. But, every simplicial map F' of $P \times [0,1]$ to W gives us a map from P to X' which lands in some d'-skeleton $Tr'_{d'}$ from where it can be homotoped to Tr'_d.

All this is quite trivial and has nothing to do with the simple connectedness of W. But, for the following lemma, this is crucial

7.3$\frac{2}{3}$$_+$ Lemma. *If $w = w'$, then there is a homotopy of Tr'_d in X moving it to $X_t \subset X$ with $t \leq \operatorname{const} d$ for $\operatorname{const} = \operatorname{const}(W, Tr)$ independent of d.*

Proof. Let $\alpha\colon W \to W$ be a smooth map homotopic to the identity and sending the 1-skeleton of the triangulation Tr of W to the point $w \in W$. This induces a map, say $\alpha'\colon X \to X$, also homotopic to the identity. Let us see what this α' does to a d-cell in Tr'. Such a cell is the product $\Delta_1 \times \Delta_2 \times \cdots \times \Delta_k$, where $\sum_{i=1}^{k} \dim \Delta_i = d$. It follows that the number of these simplices of dimension > 1 does not exceed d, and those which have dimension zero correspond to pairs of vertices, i.e., edges in Tr. These are contracted to a single point in W, and so the path corresponding to any point in Tr'_d is transformed to a new path consisting of at most d pieces, each of the form $\alpha(I_i)$, where I_i is a straight segment in some simplex $\Delta_i \subset W$. Thus, the length of this path does not exceed $\operatorname{const} d$.

This proves the inequality $ct \leq dm(t)$ for c equal to the $\operatorname{const} = \operatorname{const}(W, Tr)$ above (which depends on α in practice), since every i-cycle for $i \leq d$ is moved to X_{cd}, provided that $w = w'$, and then the case $w \neq w'$ follows trivially.

Finally, in order to prove the opposite inequality, we observe that the paths in X_t can be replaced by piecewise geodesic paths with $d \leq C_0 t$ segments, which makes this space homotopy equivalent to a simplicial complex of dimension $d' \leq Ct$ (compare [Milnor]$_{\mathrm{MT}}$, where again we do not need $\pi_1(W) = 0$). And then, we just refer to the nonvanishing of $H_i(X)$ for a set of $i \in \mathbb{N}$ containing an arithmetic progression (see [Klin], [Sul]$_{\mathrm{ICT}}$). This trivially implies that $dm(t) \leq Ct$.

Remarks$_+$. (a) If a manifold W has positive Ricci curvature, then the index of a geodesic is roughly proportional to its length, which leads to the conclusion $dm(t) \geq ct$. This could have been useful as a necessary criterion for the existence of a metric on W with Ricci > 0 if not for Theorem 7.4. Now, in view of 7.4, one might think that there are hardly any stable homotopy restrictions preventing W from having a metric with Ricci > 0. For example, one may expect that, for every finite (or, more generally, locally finite), simply connected, finite dimensional polyhedron P, there exists a complete W homotopy equivalent to P (or at least to a suspension of P) with Ricci $W > 0$. And even if one looks at *closed* simply connected manifolds W, it seems that most of them are homotopy equivalent to those with Ricci > 0, at least after a suitable stabilization, such as passing to $W \times S^2$, or by taking a connected sum of two copies of W with opposite orientations, etc. In fact, all known (C^∞ as well homotopy) obstructions for Ricci > 0 in the simply connected category are the same as for positivity of the scalar curature.

Notice in this regard that Yau's celebrated solution of the Calabi conjecture provides many examples of (Kähler) manifolds with Ricci > 0 which, as $\dim W \to \infty$, display a wide spectrum of "stable" homotopy types. On the other hand, there are elementary examples of closed manifolds with Ricci > 0 and with arbitrarily large homotopy types (see [Sha–Ya], [Col]$_{\text{ARC}}$).

(b) Our trick of taming the loop spaces by contracting the 1-skeleton of W to a point is not new. It already appears, as was pointed out to me by I. Babenko, in an old paper by Adams and Hilton (see [Ad–Hi]).

7.4. Corollary: *For two generic points x, y in W, the number of geodesic arcs from x to y of length less than t is at least*

$$\sum_{i=1}^{ct} b_i(X).$$

The same inequality is true for periodic geodesics, provided that the metric has the (generic) property of being "bumpy" in the sense of [Abr].

Proof. The condition that the metric be bumpy guarantees that all critical orbits of the energy function are nondegenerate. Thus, Morse theory (cf. [Klin], p. 63) shows that Y_t has the homotopy type of a CW complex having no cells of dimension p such that of geodesics of index p and length less than

t. Thus, there are at least

$$\sum_{i=1}^{\infty} b_i(Y_t)$$

periodic geodesics of length less than t. By Theorem 7.3, however, we have $b_i(Y_t) \geq b_i(Y)$ for $i \leq ct$, and the desired inequality follows.

7.5. Remark: Using connected sums, one can construct compact, simply connected manifolds W for which $b_n(X)$ grows exponentially as a function of n. For any bumpy Riemannian metric on such a manifold, the number of geometrically distinct geodesic loops at a point grows exponentially with the length (and "bumpy" is probably unneeded in most cases).

7.6. Remark: Usually the condition that a metric be bumpy required by Corollary 7.4 is not satisfied by homogeneous spaces. When such spaces have a closed geodesic, they have families of closed geodesics. For manifolds all of whose geodesics are closed, see [Besse].

$7.6\frac{1}{2}_+$ Remarks. (*a*) The above implies that if $b_i(X)$ grows exponentially, then the singular metric on $\mathbb{R}^n = T_w(W)$ induced by the exponential map $T_w(W) \to W$ has exponential volume growth. This, combined with Yomdin's solution to the C^∞-version of the *Shub entropy conjecture*, shows that the entropy of the geodesic flow for every C^∞ metric on such W has positive topological entropy (compare $[\mathrm{Gro}]_{\mathrm{Yom}}$ and [Bur–Pat]).

(*b*) All this suggests that having zero entropy of the geodesic flow and/or a slow (subexponential or less) volume growth of $T_w(W)$, with the metric induced from (W, g) by $\exp_w : T_w(W) \to W$ severely restricts the geometry (as well as the topology) of W. But how severely? For example, can one classify in some sense the Riemannian manifolds (W, g) for which $T_w(W)$ has polynomial growth of a given (e.g., small) degree d at every point $w \in W$?

7.7. Exercise: Using Theorem 7.3, the reader can prove the following result of D. Kan (see [Kan]): Given a compact, simply connected manifold W via a triangulation, one can effectively compute the homotopy groups of W, showing in the course of this that they are finitely generated (where the actual effectiveness depends on a given contraction of the 1-skeleton of W to a point, compare [Serre]).

7.8. Remark: Theorem 7.3 can be immediately generalized to the case of a compact manifold whose fundamental group is finite by passing to the universal cover. But, we can already see some restrictions on this generalization: Suppose that the conclusion of Theorem 7.3 is true for a manifold W. Then the word problem can be solved effectively in the fundamental group $\pi_1(W, w)$. Indeed, let us realize a word given by a loop ℓ of length t in W, where W is described by a triangulation. It then suffices to consider loops in W of length less than Ct to decide whether ℓ is homotopic to 0. In fact, it suffices to examine the list of loops consisting of $C't$ edges of the triangulation and the contiguity relations among these loops (see [Span], Ch. 3, Sec. 6) which represents a finite number of operations. For this reason, if the word problem in $\pi_1(W)$ is algorithmically unsolvable, then for every subrecursive function $f(t)$ there exist contractible loops in W of lengths $t_i \to \infty$ such that none of them can be deformed to a point by passing through loops of lengths $\leq f(t_i)$ (where "recursive" for real t means $\mathrm{ent}(f(\mathrm{ent}(t)))$ recursive).

7.8$\frac{1}{2}_+$ Dehn functions. The minimal function $f(t)$ such that every contractible loop in W can be homotoped to zero within loops of lengths $\leq f(t)$, is essentially (up to an obvious equivalence) determined by $\pi_1(W)$ and may be called the *Dehn function* of the group $\Pi = \pi_1(W)$. Much work has been done recently to evaluate Dehn's functions of specific groups Π, but it is nevertheless unknown for many rather elementary examples (see $[\text{Gersten}]_{\text{DF}}$, $[\text{Gersten}]_{\text{II}}$, $[\text{Gro}]_{\text{AI}}$ and the references therein). On the other hand, there is a general construction due to Rips and Sapir (unpublished) realizing "almost every" recursive function by the Dehn function of some group Π (compare [Brid], $[\text{Pap}]_{\text{SQII}}$). This Dehn function is closely related to the optimal two-dimensional isoperimetric function $I(t)$ from 6.32, and some people associate the name of Dehn with $I(t)$ rather than with $f(t)$. We note here that, obviously,

$$f(t) = O(I(t)) \tag{$*$}$$

and

$$I(t) = O(\exp\exp f(t)), \tag{$**$}$$

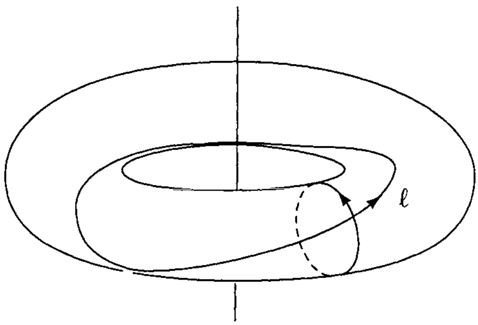
where $(*)$ easily improves to

$$f(t) = o(I(t)) \tag{$*+$}$$

for nonhyperbolic groups.

A closely related problem lies in finding groups Π with a controlled growth of the number N_t of contractible geodesics of length $\leq t$. Here there are examples of groups Π with sufficient complexity (Kolmogorov type) of the word problem to imply a lower bound $N_t \geq \exp(\lambda t)$, $\lambda > 0$, for every W with $\pi_1(W) = \Pi$, (see [Nab]).

7.9. Corollary 7.4 proves the existence of many geodesics on a generic compact, simply connected manifold. For other cases in which the existence of closed geodesics is known, see [Gro–Mey], [Klin], [Sul–Vig]. Does there exist a closed geodesic in a given complete, noncompact, simply connected manifold? The answer is "yes" when the manifold is not contractible by the theory of Lusternik–Fet (see [Klin] and [Lus–Fet]). But if the manifold has no topology, then anything can happen. One way to proceed regardless of the topology is to consider metrics arising from compact quotients. For example, it is natural to ask whether a metric on $\mathbb{R}^n$ that is induced by a metric on the torus and without closed geodesics is necessarily flat. The answer is "no" by the following example of Y. Colin de Verdière: The round 2-torus embedded in $\mathbb{R}^3$ has no geodesic loops homotopic to 0. Indeed, such a loop ℓ, if it is not a meridian, must intersect meridians transversally, since they are all geodesics. By continuity, ℓ intersects all oriented meridians in the same sense, and so the number of intersections of the cycle defined by ℓ and the cycle of the meridians is strictly positive or negative, so that ℓ is not homotopic to zero.



Problem: Study closed Riemannian manifolds without contractible closed geodesics. Then isolate from among them those which, in addition, have a unique closed geodesic in every homotopy class.

B. Dilatation of mappings between simply connected manifolds

Theorem 2.18 admits the following generalization:

Theorem 7.10 *Let V, W be compact, simply connected Riemannian manifolds. If V has the rational homotopy type of a sphere, then the growth of*

$\#(D)$ (the number of homotopy classes of mappings from V into W containing a representative with dilatation less than D) is polynomial. More precisely, $\#(D) \leq CD^{\alpha(V,W)}$, where α depends only on the rational homotopy type of V, W.

Remarks: (a) The examples of 2.17 show that the simple connectivity (in fact, the finitude of π_1) is essential.

(b) Of course, the constant C (about which we would like to know more) depends on the geometry of V, W.

We will prove the theorem in several steps. First, we note that it suffices to consider C^∞ mappings, since by [Dieu] every Lipschitz map f is homotopic to a C^∞ mapping whose dilatation is arbitrarily close to that of f.

7.11. The case when $V = S^3$ and $W = S^2$ indicates the general line of proof. Since the homotopy class $[f]$ of $f: S^3 \to S^2$ is determined by the Hopf invariant $h(f)$, it suffices to control $h(f)$ as a function of the dilatation.

7.12. Lemma: *There is a constant $C > 0$ such that for each f, we have $h(f) \leq C(\mathrm{dil}(f))^4$.*

Proof. We recall the computation of $h(f)$ in terms of differential forms, as described, for example, in [Godb], p. 221. Let ω be a volume form on S^2, normalized by the condition $\int_{S^2} \omega = 1$. Its pullback $f^*\omega$ is a closed 2-form on S^3 and is therefore exact. If $f^*\omega = d\alpha$, the number

$$h(f) = \int_{S^3} f^*\omega \wedge \alpha$$

depends only on $[f]$ and is an integer.

If, as in Proposition 2.11, we equip the space $\Omega(S^n)$ of differential forms with the sup norm, we have

$$\|f^*\omega\| \leq C(\mathrm{dil}(f))^2 \|\omega\|,$$

and the desired inequality is a consequence of a metrical Poincaré lemma formulated as follows.

7.13. Lemma: *There exists a constant C (depending on the metric on S^n) such that for each closed form in $\Omega^i(S^n)$ $(0 < i < n)$, there exists a form α in $\Omega^{i-1}(S^n)$ such that $d\alpha = \omega$ and $\|\alpha\| \leq C\|\omega\|$.*

Proof. For a form ω defined on an open subset of $\mathbb{R}^n$ that is star-shaped relative to 0, the primitive given by

$$\alpha_x(\xi) = \int_0^1 \omega_x(tx, \xi)\, dt$$

satisfies such an inequality. From this remark, we obtain the proof from the fact that $H^i(S^n; \mathbb{R}) = 0$ $(0 < i < n)$, as done in [Ber–Gost] for example (i.e., using differential forms), controlling the norm at each step.

Some results of Chen–Sullivan (See [Fr–Gr–Morg], [Sul]$_{\mathrm{DF}}$, and [Sul]$_{\mathrm{ICT}}$) By 7.11, the map $f \mapsto h(f)$ defines a linear form on $\pi_3(S^2)$ that can be expressed in terms of differential forms on S^2. More generally, if V is a simply connected manifold, the graded vector space $\mathrm{Hom}(\pi_*(V), Q)$ is isomorphic to $M\Omega(V)/[M\Omega(V), M\Omega(V)]$, where $M\Omega(V)$ is the *minimal model* of the graded differential algebra $\Omega(V)$ of differential forms on V defined as follows:

7.15. If A is a commutative-graded differential algebra (cf. [Leh], p. 18, or [Fr–Gr–Morg], Ch. 12) such that $H^0(A) = \mathbb{R}$ and $H^1(A) = 0$, it can be shown (ibid) that there exists a comutative-graded differential algebra M, unique up to isomorphism, such that

1. M is a free algebra, i.e., it has no relations other than associativity and commutativity $ab = (-1)^{\deg(a)\deg(b)} ba$.
2. dM_n is contained in the subalgebra generated by the M_k such that $k \leq n$ (here we denote by M_k the set of elements of degree k).
3. There is a morphism $\rho\colon M \to A$ of graded differential algebras that induces an isomorphism on cohomology.

7.16. Examples: The minimal model of ΩS^{2n} is the free $\mathbb{R}$-algebra $\mathbb{R}\{a, b\}$ generated by two elements a, b such that

$$\deg(a) = 2n \quad \deg(b) = 4n - 1 \quad da = 0 \quad db = a^2.$$

The element a gives H^{2n}, and the adjunction of the generator b of degree $4n - 1$ kills the cohomology that the powers of a would give otherwise.

In contrast, the minimal model of $\Omega(S^{2n+1})$ is simply $\mathbb{R}\{c\}$ with $\deg(c) = 2n + 1$, where for reasons of parity $c^2 = 0$ (cf. [Leh], Ch. 5).

7.17. Minimal models and the Hopf invariant. Under these assumptions, a C^∞ map $f : S^3 \to S^2$ gives rise to the following commutative diagram (cf. [Leh], p. 19),

$$\begin{array}{ccc} M_{S^2} & \xrightarrow{\tilde{f}^*} & M_{S^3} \\ {\scriptstyle\rho}\downarrow & & \downarrow{\scriptstyle\rho} \\ \Omega S^2 & \xrightarrow{f^*} & \Omega S^3 \end{array}$$

From the properties of degree, it follows that $\tilde{f}^* b$ is proportional to c and one can show (exercise or [Fr–Gr–Morg], p. 210) that $\tilde{f}^* b = h(f)c$.

7.18. Proof of Theorem 7.10 for $V = S^n$. We must extend the geometrical and algebraic considerations above to arbitrary simply connected manifolds. The minimal model of $\Omega(W)$ is constructed step by step (cf. $[\mathrm{Sul}]_{\mathrm{ICT}}$, p. 38, [Leh], p. 29, and [Fr–Gr–Morg], Ch. 12 for more details). Given a differential algebra M^n satisfying 7.15(1) and (2), and a morphism $M^n \to \Omega(W)$ inducing an isomorphism of $H^p M^n$ onto $H^p(W;\mathbb{R})$ if $p \le n$, and a surjection for $p = n+1$, we obtain a new algebra M^{n+1} having the same properties up to stage $n+1$ by adjoining to M^n some generators of degree $n+1$ in order to kill the excess cohomology in $H^{n+1}M^n$.

Next, we must find an integral representation of $\mathbb{R}$-linear forms on $\pi_n(W)$. Given a representative $f : S^n \to W$ of an element of $\pi_n(W)$, we begin with a system of representatives ω_i for the deRham cohomology of W of degree less than or equal to n, and we consider their inverse images $f^*\omega_i = \omega_i^*$.

(a) For an n-form ω_i, the integral $\int_{S^n} \omega_i^*$ gives a linear form on $\pi_n(W)$. In this way, we obtain those forms that factor through the Hurewicz homomorphism (see 4.20 and [Hur–Wall], p. 148) to forms on $H_n(W;\mathbb{Z})$.

(b) If $\deg(\omega_i) < n$, then the form ω_i^* is exact, so we can write $\omega_i^* = d\alpha_i$ and consider all linear combinations $\sum c_{ij}\omega_i \wedge \alpha_j$ that give closed forms. By the preceding discussion, these are the images under ρ of the generators of the minimal model of $\Omega(W)$. We integrate those of degree n over S^n. The others are exact, of the form $d\alpha'_k$, and we recommence with the combinations

$$\sum c'_{ik}\omega_i \wedge \alpha'_k.$$

After finitely many steps (since $H^1(W;\mathbb{R}) = 0$), we obtain n-forms, which we integrate to show (cf. [Fr–Gr–Morg], Ch. 12, and $[\mathrm{Sul}]_{\mathrm{ICT}}$, §11) that all linear forms L on $\pi_n(W)$ are obtained in this way. It now suffices to note that by Lemma 7.12, we have $L(f) \leq c(\mathrm{dil}(f))^r$, where r depends essentially on the degree of the forms involved and the number of necessary steps.

N.B. A more careful examination of the above gives for r the bound $2(n-1)rg(\pi_n(W))$ (see [Gro]$_{\mathrm{HED}}$).

7.19$_+$ Remarks. (*a*) The method of minimal models applies to maps $V \to W$, where V is an arbitrary manifold, and this seems to bound the number of mutually nonhomotopic maps of dilatation less than d by d^r for some r depending on the rational homotopy types of V and W. (Actually, this appears to be obvious and my use of the word "seems" just reflects my reluctance to trust the intuition for minimal models which has grown rusty from years of nonuse.)

(*b*) One can use weaker dilatation measures of maps, such as the L_p norm of the differential of Df on $T(V)$ or on some exterior power $\bigwedge^i T(V)$ to evaluate the morphisms of the minimal models (see [Gro]$_{\mathrm{HED}}$ and [Gro]$_{\mathrm{CC}}$).

(*c*) Instead of applying minimal models to general V, one could try an induction by skeletons of some triangulation of V, thus reducing the general case to that of the spheres. This approach, however, needs *a priori* bounds on the dilatation of homotopies contracting our spheres mapped to W. This point (which was overlooked in [Gro]$_{\mathrm{HED}}$ and pointed out later to me by P. Pansu) is briefly discussed below.

7.20$_+$ Basic problems of quantitative homotopy theory. Topological spaces naturally appearing in the homotopy theory come along with particular metrics, or rather Lipschitz (or more general) classes of metrics. Such a class is essentially unique for spaces represented by finite polyhedra. Of course, two homotopy equivalent polyhedra P_1 and P_2 do not have to be bi-Lipschitz equivalent, but they are *Lipschitz homotopy equivalent* (in an obvious sense), and this is all we need. Next, if we look at the space of maps between two compact Riemannian manifolds V_1 and V_2, for instance, this space $\mathrm{Map}(V_1 \to V_2)$ carries many (perhaps too many) natural (classes of) metrics coming from metrics on V_1 and V_2. Here one can use the uniform metric between $f_1, f_2 \colon V_1 \to V_2$, that is,

$$\mathrm{dist}(f_1, f_2) = \sup_{v \in V_1} \mathrm{dist}(f(v_1), f(v_2)),$$

but also more interesting metrics measuring the dilatations of homotopies $h \colon V_1 \times [0, \ell] \to V_2$ between maps (where the choice of ℓ may be important) and also such characteristics of h as generalized Dirichlet functionals as

$$\int \| \wedge^i D | \wedge^i T(V_1 \times [0,\ell]) \|^p$$

mentioned in (b) above. And also some infinite-dimensional classifying spaces have particularly nice geometric realizations (Grassmannian manifolds, spaces of operators on Hilbert space, etc.) with interesting metrics or even families of those.

Now, given any two metric spaces X and Y, we have a way to measure dilatation of maps $f: X \to Y$ which may be our old friend the Lipschitz constant, or the dilatation of the k-volume or k-dimensional subsets, or something like a Dirichlet functional. Then we ask:

> What is the asymptotic behavior of the number $N(d_1)$ of maps $f: X \to Y$ with "dilatation"$(f) \leq d$?

Conjecture: If X and Y are finite polyhedra with $\pi_1(Y)$ finite, then $N(d)$ is asymptotic to d^r for the (necessarily integral!) number r predicted by the computations with minimal models.

The problem here is to actually construct maps $X \to Y$ with not very large dilatations in homotopy classes $\to \infty$.

A closely related problem is that of constructing a homotopy $h: X \times [0,\ell] \to Y$ between two given maps $f, f': X \to Y$ of "dilatation" $\leq d$, where we are given *a priori* knowledge of the existence of some h, and we want an h with a minimal possible "dilatation" $\delta = \delta(d)$. So, we have something here like the Dehn function, and we expect $\delta(d) \sim d^\alpha$ for all simply connected finite polyhedra X and Y, and where quite often α may be just equal to one. Finally, generalizing individual homotopies, we can look at the whole space of maps $\mathcal{F} = \{f: X \to Y\}$ and study the asymptotics of the homotopy invariants of the levels of suitable geometric functionals on $\mathcal{F}$ (such as dilatation) in the spirit of Morse theory. So, the possibilities are many (and one can easily imagine further questions along the same lines), but almost nothing is known beyond the results indicated in the sections above (see $[\mathrm{Gro}]_{\mathrm{N}\ell\mathrm{S}}$, $[\mathrm{Gro}]_{\mathrm{CC}}$, and $[\mathrm{Gro}]_{\mathrm{PCMD}}$ for further aspects of these problems).

Example. The first space to look at is $\mathbb{C}P^\infty$. One knows that every map $f: S^i \to \mathbb{C}P^\infty$ is contractible for $i \geq 3$, and we want to estimate the necessary dilatation δ of such a contraction h in terms of the dilatation d of f. The usual way to contract this f is by lifting it to the sphere S^∞ Hopf-fibered over $\mathbb{C}P^\infty$. In other words, we take the induced S^1-fibration over S^i, say $p: \Sigma \to S^i$, where we look for a section $\varphi: S^i \to \Sigma$ of a small dilatation. We use the standard connection in the Hopf fibration $S^\infty \to \mathbb{C}P^\infty$, which we pull back to $p: \Sigma \to S^i$. Next, we observe that the curvature of this equals the 2-form $f^*(\omega)$ on S^i pulled-back from the canonical (Kähler) form ω on $\mathbb{C}P^\infty$. This $f^*\omega$ has norm $\approx d^2$ for $d = \operatorname{dil} f$, and this allows a radial section in a hemisphere $S^i_+ \subset S^i$ with dilatation $\approx d^2$. Thus, for $i \geq 3$, two of these sections, one on S^i_+ and the other on S^i_- can be joined by a "small" homotopy (essentially in S^1) which gives us a section $S^i \to \Sigma$ of dilatation $\approx d^2$ and hence a contraction of $f: S^i \to \mathbb{C}P^2$ with dilatation $\approx d^2$. Probably, a (formal) generalization of this argument (to something like Postnikov towers) could yield the above conjecture along with a polynomial bound on the higher dimensional Dehn function $\delta(d)$ for simply connected spaces.

Chapter 8_+

Pinching and Collapse

A. Invariant classes of metrics and the stability problem

8.1. The group of diffeomorphisms $\mathrm{Diff}(V)$ of a smooth manifold V naturally acts on the space of Riemannian metrics g on V; various classes of metrics one studies in geometry are usually invariant under $\mathrm{Diff}(V)$. In fact, we tend not to distinguish isometric manifolds, and a diffeomorphism $f : V \to V$ establishes an isometry between (V, g) and $(V, f^*(g))$ for each metric g. Furthermore, the geometric dictum "from local to global" suggests the study of *locally* defined classes of metrics on V which are moreover Diff-invariant.

Example: Fix a germ g_0 of a Riemannian metric on a manifold V_0 at a point $v_0 \in V$ and consider the class $\mathcal{G} = \mathcal{G}(g_0)$ of metrics g on V, where $\dim V = \dim V_0$, defined by the following condition: for each point $v \in V$, there exists an arbitrarily small neighborhood $U \subset V$ of v isometric to (V_0, g_0) (i.e., to some neighborhood U_0 of v_0 in V_0 where g_0 is defined). In other words, (V, g) is *locally isometric* to (V_0, v_0, g_0) for every $g \in \mathcal{G}$.

8.2. Theorem (Singer, see [Sing]). *Let V be a C^∞-smooth Riemannian manifold which is locally isometric to some (V_0, v_0, g_0). Then V is locally homogeneous in the following sense: there exists a continuous function $\varepsilon = \varepsilon(v) > 0$ on V such that the ε-balls $B(v_1, \varepsilon(v_1))$ and $B(v_2, \varepsilon(v_2))$ in V are mutually isometric for all $v_1, v_2 \in V$, where the implied isometry $B(v_1, \varepsilon(v_1)) \leftrightarrow B(v_2, \varepsilon(v_2))$ sends $v_1 \leftrightarrow v_2$. Consequently, if V is complete and simply connected, then it is homogeneous, i.e., the isometry group is transitive on V.*

8.3. Problem: What is the minimal degree of regularity needed for the conclusion of the theorem to hold true? One knows that C^r with $r \approx \dim V$ suffices (see [Gro]$_{\text{PDR}}$, p. 165), but one may expect this for $r = 0$ or even for rather general locally compact path metric spaces. (The difficulty stems from the possibility that the sizes of the mutually isometric ε-balls at v_1 and v_2 get smaller as $\operatorname{dist}(v_1, v_2) \to 0$, since we do *not* assume *a priori* that our local isometries $(U, v) \to (V_0, v_0)$ send $v \to v_0$. This problem becomes even more pronounced if we only require that for every neighborhood $U_0 \subset V_0$ of u_0, where g_0 is defined, there is a neighborhood $U \subset V$ of v which admits an isometry onto an *open subset* $U_1 \subset U_0$. Notice that Singer's theorem applies in this case, thus showing that the new condition is as good as the old one for C^∞ metrics.)

8.4. Example: Fix a "curvature" tensor R on $\mathbb{R}^n$, i.e., a 4-tensor in the variables x, y, z, t, symmetric with respect to pairs, alternating in $x \wedge y$ and $z \wedge t$, etc., (see [Kulk]), and consider the (class of) metrics whose curvature tensor lies in the $O(n)$ orbit of R for every tangent space $T_v(V)$, $v \in V$. The family thus obtained consists of those manifolds V whose curvature is constant, in the sense that for each $v, w \in V$, there exists an isometry i of T_vV onto T_wV such that $i^*R_w = R_v$. (The reader should not confuse this with the expression *constant curvature*, which means that the tensor R is of the form $R(X,Y)Z = K(\langle Y, X\rangle Z - \langle X, Z\rangle Y)$, nor with the stronger hypothesis of *parallel curvature*, or $DR = 0$, which characterizes symmetric spaces (see [Helg])). The manifolds in this family *are not all locally homogeneous* since there exist non-isometric manifolds having the same curvature (see [Fer–Kar–Munz] and [Tri–Van]).

8.5. Existence, uniqueness, and stability of global developments of locally homogeneous spaces. Let V_0 be a *locally homogeneous path metric space* for which there is an $\varepsilon > 0$ such that all ε-balls in V_0 with centers sufficiently close (e.g., ε close) to a given point $v_0 \in V_0$ are all isometric to $B_{v_0}(\varepsilon) \subset V_0$, where we assume that the implied isometries $B_v(\varepsilon) \leftrightarrow B_{v_0}(\varepsilon)$ move $v \leftrightarrow v_0$ (which is automatic in many cases). Then we ask the following questions:

I. **Existence of a complete development.** Does there exist a *complete* metric space V locally isometric to V_0 having all ε-balls $B_v(\varepsilon) \subset V$ isometric to $B_{v_0}(\varepsilon) \subset V_0$ with $v \leftrightarrow v_0$? Such a space is called a *development* of V_0.

Counterexample: Let G be a compact, connected Lie group and let H_0 be a connected *local* subgroup whose global development is *non*-closed in G.

Then the local quotient G_0/H_0 usually admits no complete development.

II. **Homogeneity and uniqueness.** When is V homogeneous? unique up to isometry? Does *simple connectivity* ensure these properties?

The essential requirement needed for the uniqueness and homogeneity of the simply connected development is the uniqueness of extensions of local isometries (in V_0), which is known to be satisfied by locally homogeneous topological manifolds by D. Montgomery's theorem, but there may exist some pathological infinite-dimensional counterexamples.

III. **Stability.** Now we arrive at the problem we are truly interested in. Suppose V' is a complete path metric space which is everywhere "locally close" to V_0. Is then V' "*globally* close" to some development V of V_0?

To make sense of this, we need to specify the notion of local closeness. Here are several possibilities:

(A) There is an $\varepsilon > 0$ such that all ε-balls in V' are δ-close to $B_{v_0}(\varepsilon) \subset V_0$ in one of our metrics, e.g., Hausdorff, Lipschitz, or $\underline{\square}_1$, where $\delta \leq \delta(\varepsilon)$ is small. Then, we naturally expect V' to be close to V in the same metric with the implied distance between V' and V going to zero as $\delta \to 0$ for every fixed $\varepsilon > 0$.

(B) The same as in (A), but with $\varepsilon = \varepsilon(v) > 0$ depending on v and with $\delta = \delta(\varepsilon) > 0$ depending on ε.

(C) It may easily happen that even a space V_1 locally *isometric* to V_0 is not close to it in the sense of (A). In fact, we can divide a homogeneous space V by a discrete isometry group Γ which moves some points in V less than ε, thus creating small balls in $V_1 = V/\Gamma$ non-isometric to those in V_0. For example, if we divide $\mathbb{R}^n$ by a lattice Γ spanned by vectors of norm $\leq \varepsilon_1 \sim \varepsilon/n$, then each ε-ball in $V_1 = \mathbb{R}^n/\Gamma$ is isometric to all of V_1, which is a torus whose geometry is far from that of a Euclidean ball $B(\varepsilon) \subset \mathbb{R}^n$. With this in mind, we modify (A) by requiring that for each $v \in V'$, there is a (possibly non-injective!) map $e\colon B_{v_0}(\varepsilon) \to V'$ for our ball $B_{v_0}(\varepsilon) \subset V_0$, such that the path metric in $B_{v_0}(\varepsilon)$ induced from V' is δ-close to the induced V_0 metric in $B_{v_0}(\varepsilon)$. Here "close" typically means Lipschitz close, i.e., e is $(1+\delta)$-biLipschitz with respect to V_0 and V'-metrics. But one can invoke the Hausdorff (and/or $\underline{\square}_1$) metric as well. In the latter case, one could use a slightly different setting with an intermediate "ball" $\tilde{B}_v$, that is, a connected space coming equipped with a local homeomorphism $\tilde{e}\colon \tilde{B}_v \to V'$ such that a distinguished point $\tilde{v} \in \tilde{B}_v$ goes to a given point $v \in V$ and $\tilde{v}$ lies ε-far from the boundary of $\tilde{B}_v$. Here we refer to the induced path metric $\widetilde{\mathrm{dist}}$ in $\tilde{B}_v$ and "ε-far from the boundary" means that each ε'-ball in $\tilde{B}_v$ with the center $\tilde{v}' \in \tilde{B}_v$ positioned δ-close to $\tilde{v}$ is sent *onto* the ε'-ball in V' around $\tilde{e}(\tilde{v}') \in V'$, provided that $\delta + \varepsilon < \varepsilon$. We call such $\tilde{B}_v \xrightarrow{\tilde{e}} V$ an (open) *ε-multiball around* $v \in V'$ if in addition $\widetilde{\mathrm{dist}}(\tilde{v}, \tilde{v}') < \varepsilon$ for all $\tilde{v}' \in \tilde{B}$. Now, the condition "V' is locally close to V_0" reads as (A) above with "ball" replaced by "multiball."

8.6. Obstructions to stability: Overgrown size and collapse. Suppose V' is compact and δ-close on the ε-balls to V_0 as in (A) above. Then, if δ is sufficiently small compared to the universal number of ε'-balls needed to cover V', with ε' slightly smaller than ε, say $\varepsilon' = \varepsilon/3$, then in most cases one can show that V' is close to some development V of V_0. We shall discuss particular examples later on and indicate here what can go wrong if some of the conditions above are not met.

(i) V' is too big compared to δ, for example, V' has large diameter and/or dimension. Then local errors of order δ appearing in the course of an assembly of V' from ε-balls may accumulate to something of order $\varepsilon \exp(N)$, where N is the number of the balls needed to cover V'. This is closely related to another kind of instability, called *Ulam instability.* A (homogeneous) metric space V is called *Ulam stable* (compare §III.5 in [Ulam]) if every δ-almost isometric self-mapping $V \to V$ is δ'-close to some isometry with $\delta' = \delta'(\delta) \xrightarrow{\delta \to 0} 0$. Here again, one should specify the meaning of "δ-almost" and "δ-close." One knows, for example, that most symmetric spaces (essentially, with the exception of $\mathbb{R}^n$, H^n, and $\mathbb{C}H^n$) are Ulam stable in the quasi isometric sense (see [Pan]$_{CC}$, [Kle–Le], [Esk–Farb]), but one does not know if the Hilbert sphere is Ulam bi-Lipschitz stable (as was recently pointed out to me by V. Zoritch), and where the expected answer is "no."

Exercise: Show that for every $\varepsilon > 0$ and $\delta > 0$, there exists an N, an ε-net Σ in the unit sphere $S^N \subset \mathbb{R}^{n+1}$, and a δ-almost isometric (i.e. $(1+\delta)$-biLipschitz) map $\Sigma \to S^N$ which is ρ_0-far from any isometry, for some fixed ρ_0', say $\rho_0' = 0.01$. (*Hint:* Observe that random $2N$-vectors in S^N are almost mutually orthogonal for large N and recall the Kirszbraun theorem.)

(ii) Another obstruction which can bring even a small V' far from any development V of V_0 is the phenomenon of collapse, where V' is locally close to V_0 in the sense of (C) above, with multiballs and without having ε-balls close to $B_{v_0}(\varepsilon) \subset V_0$ as required by (A). We call this phenomenon *collapse*, since it corresponds in many cases to V' being Hausdorff close to a space of dimension strictly less than $\dim V'$. Notice that such a V' can be covered by relatively few ε-balls, but these are no good any more since they are not close to $B_{v_0}(\varepsilon) \subset V_0$. Of course, one could try smaller balls of radii $\varepsilon' \subset \varepsilon$ in V' which are δ-isometric to $B_{v_0}(\varepsilon') \subset V_0$, but then we have to overcome two problems:

(a) The number of these balls needed to cover V' is large.

(b) Being "δ-almost isometric" is a cheaper condition for smaller balls than for the large ones. For example, all sufficiently small balls in a compact Riemannian manifold are almost isometric (i.e. $(1+\delta)$-biLipschitz) to the Euclidean one, without making this manifold globally almost Euclidean in any sense. Thus, we need $\delta = \delta(\varepsilon) \to 0$ for $\varepsilon \to 0$, but even this cannot offset the possible disruption of stability due to collapse. We have seen this in 3.11, where we could have, for example, a sequence of homogeneous spaces $V_i = G/H_i$ where the groups H_i locally converge to some closed subgroup H in the ambient Lie group G without converging to G globally (i.e. Hausdorff converging to a subgroup $\overline{H} \subset G$ with $\dim \overline{H} > \dim H = \dim H_i$). Thus, one can have V_i' with local geometries converging to that of $S^2 \times S^3$ (in the sense of multiballs from (C)) without being remotely close to $S^2 \times S^3$ in the global sense.

We shall see later on that such collapse cannot happen if V' is 2 connected. But prior to this, we want to study collapse for general Riemannian manifolds V with bounded sectional curvature, say $|K(V)| \leq 1$, where "collapse" signifies some deviation of the local geometry of V from the Euclidean. First of all, we shall clarify basic families of noncollapsed manifolds with $|K(V)| \leq \text{const}$, which we start by briefly explaining the meaning of the sectional curvature $K(V)$.

B. Sign and the meaning of curvature

8.7. The Rauch theorem and equidistant deformations. The simplest manifestation of curvature is expressed by the *Rauch comparison theorem*, which estimates how much the *exponential map* $\exp_v : T_v(V) \to V$ deviates from being isometric in terms of $K(V)$. Recall that $\exp_v$ sends each vector x in the tangent space $T_v(V) = \mathbb{R}^n$, $n = \dim V$, to the end of the geodesic segment $[v, v']$ in V issuing from v in the direction x and having length$[v,v'] = \|x\|$. For example, *if* $|K(V)| \leq 1$, *then* $\exp_v$ *is locally* $(1+\delta)$*-bi-Lipschitz on the ball* $B_0(\rho) \subset T_v(V) = \mathbb{R}^n$ *for all* δ *and* ρ *satisfying* $\rho \leq 1$ *and* $\delta \leq \rho^r$, where "$(1+\delta)$-bi-Lipschitz" amounts here to the bounds $\|D\| \leq 1+\delta$ and $\|D^{-1}\| \leq 1+\delta$ for the norms of the differential of the map $\exp_v$ and its inverse. The above is a rather rough bound, but it coveys the essence. In fact, one knows that the condition $|K(V)| \leq 1$ makes

$$\|D_x\| \leq \sinh \|x\| \tag{$*$}$$

and

$$\|D_x^{-1}\| \leq 1/\sin \|x\| \tag{$**$}$$

for $\|x\| \leq \rho = \pi$ (see [Che–Ebin] and below). In other words, $(*)$ tells us that $\exp_v$ *expands no more* than the exponential map in hyperbolic space, while $(**)$ says that it *contracts no more* than exp in the unit sphere. Thus, the condition $|K(V)| \leq 1$ makes small multiballs in V almost Euclidean.

Exercises. (a) Show that $K(V) \geq 0$ if and only if $\|D\| = \|D(\exp_v(x))\| \leq 1$ for $\|x\| \leq \varepsilon = \varepsilon(v) > 0$ and all $v \in V$, while the opposite inequality $K(V) \leq 0$ is equivalent to $\|D^{-1}\| \leq 1$ everywhere. (This will become clear with our definition of $K(V)$ given below.)

(b) Show that the inequality $K(V) \geq 0$ ensures the inequality $\|D\exp_v(x)\| \leq 1$ insofar as the geodesic segment $[v, v' = \exp_v(x)]$ *remains locally distance-minimizing*, i.e., such that every curve between v and v' lying close to $[v, v']$ is longer than length$[v, v'] = \|x\|$. This minimizing condition is customarily expressed by saying that "*there are no conjugate points on the segment* $[v, v']$."

(b') Similarly, prove that $K(V) \geq 1$ makes $\|D\exp_v(x)\| \leq \sin \|x\|$ in the absence of conjugate points on $[v, v']$, while the condition $K(V) \geq -1$ makes $\|D\exp_v(x)\| \leq \sinh(x)$ before the first conjugate point appears (along $[v, v']$).

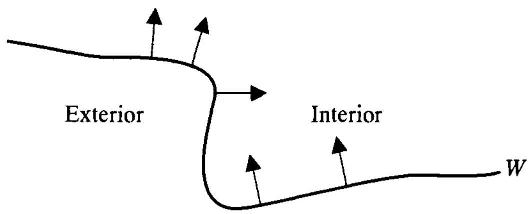
(c) Show that $K(V) \leq 0$ makes $D^{-1} \geq 1$ everywhere, that the condition $K(V) \leq 1$ implies $D^{-1} \leq (\sin \|x\|)^{-1}$ for $\|x\| \leq \pi$, and that $K(V) \leq -1$ yields $D^{-1} \leq (\sinh \|x\|)^{-1}$ everywhere. (Again, all of this will become perfectly clear by the end of this section.)

(d) The above geometric implications of the sign (and magnitude) of the curvature extend trivially to manifolds V for which $K(V)$ is bounded (from above and/or from below) by a constant $c \neq 1$ via the scaling $V \mapsto \sqrt{c}V$, since $K(\sqrt{c}V) = cK(V)$.

Although the Rauch theorem adequately reflects the bounds on $K(V)$, there are further geometric implications of these bounds which are hard to derive directly from the Rauch theorem. In fact, if we enlarge the category under consideration from Riemannian to Finslerian, then the properties of

$K(V)$ expressed by the Rauch theorem become inequivalent to stronger geometric manifestations of $K(V)$ discussed below.

Definition of $K(V)$ via moving hypersurfaces. Take a smooth, co-oriented hypersurface $W \subset V$ and let W_t, $t \in [-\varepsilon, \varepsilon]$ be *equidistant* hypersurfaces, which are the levels of the signed distance function $v \mapsto \pm \operatorname{dist}(v, W)$ for which the $+$ sign corresponds to what we call the *exterior* of W, and the $-$ sign refers to the *interior.*



We denote by g_t the Riemannian metric on W_t induced by the metric g on V, and we bring this g_t to $W_0 = W$ via the *normal projection* $W_t \to W$. Notice that for small ε, our function $\pm \operatorname{dist}(v, W)$ is smooth (away from the boundary of W), thus the levels W_t are also smooth, since $\| \operatorname{grad} \pm \operatorname{dist} \| = 1$ and the *normal projection* (or the gradient flow map) from W_t to W, for $v \mapsto$(nearest point in W), is a diffeomorphism. We denote by g_t^* the metric on W coming from g_t by the normal projection, and we make the following

Definition of the second fundamental form of W. This is a quadratic differential form on W, denoted II and defined via the t-derivative of g_t^* on W by

$$II = II^W = \frac{1}{2}\frac{d}{dt} g^*_{t=0}. \tag{+}$$

If the reader is used to another definition, we invite him/her to check that it is equivalent to our (+). Note that one should only bother doing this in $\mathbb{R}^n$, since every Riemannian manifold is infinitesimally Euclidean at each point v up to first order. *Thus, all formal properties of curvatures of submanifolds in V, as well as those of (first and) second derivatives of functions on V, are the same for general Riemannian V as for $V = \mathbb{R}^n$.*

Exercises: (a) Recall that W is called *strictly convex* at $w \in W$ if $II_W > 0$. Show that this is equivalent to the exponential pullback of W in $\mathbb{R}^n = T_w(V)$ being strictly convex (in the Euclidean sense) at zero. Thus show that sufficiently small, closed, locally strictly convex *immersed* hypersurfaces in V are in fact *embedded* (i.e. have no double points) for $n \geq 3$, provided that "strictly" is uniform as "small" $\to 0$. (This may look trivial, but a similar problem in Finsler geometry appears to be unsolved.)

(*b*) Show that the hessian of $\pm\operatorname{dist}(v, W)$ has rank $n-1$ with $\ker \mathrm{Hess} = \operatorname{grad}(\pm\operatorname{dist})$, and with the eigenvalues at $v \in W_t$ equal to those of $II_v^{W_t}$ with respect to g_t.

Now we want to define the curvature of V by comparing the second derivative of g_t^* to that of some W in Euclidean space. To do this, we introduce the *shape operator* A of W, which is the symmetric operator associated to II via g_0, i.e., defined by

$$II(\tau_1, \tau_2) = g_0(A\tau_1, \tau_2).$$

Thus, $A\colon T(W) \to T(W)$, and the eigenvalues of A are called the *principal curvatures* of W. A classical (going back to Gauss? Euler?) formula for $W \subset \mathbb{R}^n$ reads

$$\frac{dA_t^*}{dt} = -(A_t^*)^2,$$

at $t = 0$, where $A_t^* : T(W) \to T(W)$ is obtained from the shape operator A_t on W_t by the normal projection $W_t \to W$.

Proposition-Definition. *There exists a unique quadratic form $B = B_S$ on each hyperplane $S \subset T_v(V)$ such that every hypersurface $W \subset V$ passing through v and having $T_v(W) = S$ satisfies the following*

Tube formula at $v \in W$ with $T_v(W) = S$,

$$\frac{d}{dt} A_{t=0}^* = -(A_0^*)^2 + B_S. \tag{$\star$}$$

This B is related to the sectional curvature K_v by

$$K(\tau_1 \wedge \tau_2) = -g(B_S(\tau_1), \tau_2)$$

for all $S \subset T_v(V)$ and all orthonormal vectors τ_1 and τ_2, where τ_1 is normal to S (and hence $\tau_2 \in S$).

The reader who prefers another definition of K may amuse him/herself by identifying his/her definition with ours. This will be a simple (albeit boring) linear algebraic computation for whatever alternative definition one wants to substitute for ($\star$).

The efficiency of ($\star$) lies in its applicability to every hypersurface W_t in an equidistant family, insofar as they remain smooth. Thus, we can control the variation of II^{W_t} and g_t with t.

The traced tube formula and Ricci curvature. The Ricci curvature is responsible for the second derivative of the $(n-1)$-volume of W_t in the same way that the sectional curvature controls the induced metric. We start again with the first derivative of the *volume density* on W_t (thought of as an $(n-1)$-form on W_t stripped of its sign) and observe that

$$\frac{d\,\mathrm{vol}_t^*}{dt} = M(W_t)\,\mathrm{vol}_t^*\,.$$

Here, vol_t^* denotes the volume density of W_t brought to W_0, and $M(W) = \mathrm{Tr}_r II^W$. This is elementary for $V = \mathbb{R}^n$, and the case of a general V follows by the preceding discussion. Then we trace $(\star)$ and obtain

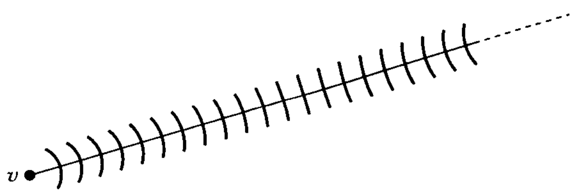
$$\frac{d}{dt}M(W_t) \doteq -\mathrm{Tr}A_t^2 - \mathrm{Ricci}(\nu,\nu), \qquad (\star)_{\mathrm{Tr}}$$

where ν is the unit normal (inward or outward makes no difference) field on W_t, and Ricci is the quadratic form defined by

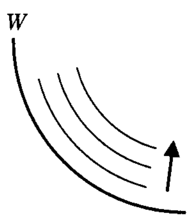
$$\mathrm{Ricci}(\nu,\nu) = \mathrm{Tr}B_S,$$

where ν is normal to $S \subset T(V)$ for all hyperplanes S.

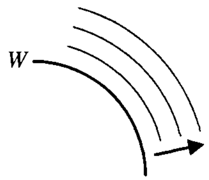
Exercises: (a) Recapture the Rauch theorem by applying $(*)$ to the concentric spheres around each point $v \in V$ or rather to germs W_t of these spheres normal to each geodesic segment issuing from v before the first conjugate point (which is equivalent to smoothness of these germs).



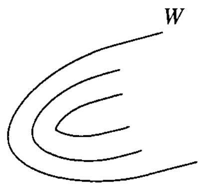
(b) Show that $K \geq 0$ is equivalent to the preservation of (strict) convexity under the inward deformations of W,



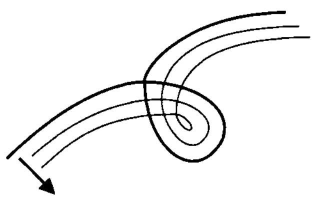
while $K \leq 0$ corresponds to preservation of convexity for outward deformations.



Show furthermore that if $K \geq 0$, then the convexity is preserved for inward deformations even after W_t loses the smoothness, where the (strict) convexity of a nonsmooth hypersurface W can be defined with the exponential pullbacks of W to $\mathbb{R}^n = T_v(V)$ as before. Then prove *Gromoll's contraction principle*, which claims that a compact manifold V with nonempty convex boundary W and $K(V) > 0$ contracts to a point by a family of such submanifolds (V_t, W_t). In particular, V is homeomorphic to the n-ball.



(c) Study inward equidistant deformations of closed, strictly convex *immersed* hypersurfaces W in complete manifolds V with $K(V) \geq 0$, and show that they remain immersed strictly convex (although they may become singular) for $\dim W \geq 2$. (The idea of the proof, suggested to me by W. Meyer more than 20 years ago, consists in excluding bad local singularities like those appearing for $\dim W = 1$ by analyzing what happens in $\mathbb{R}^n$, where everything is rather transparent.)



(d) Prove the Cartan-Hadamard theorem, which claims that *a complete, simply connected manifold V with $K(V) \leq 0$ is diffeomorphic to $\mathbb{R}^n$*, $n = \dim V$, by appealing to the convexity of the concentric spheres around some point.

(e) Assume V is complete and noncompact, and show that V admits a *horofunction* (generalizing the idea of *Busemann functions*), which is a limit

$$h(v) = \lim_{v_i \to \infty} (\mathrm{dist}(v, v_i), \mathrm{dist}(v_0, v_i))$$

for some points $v_i \to \infty$. Show that every horofunction is smooth and *convex* (i.e., $\mathrm{Hess}(h) \geq 0$) if $\pi_1(V) = 0$ and $K(V) \leq 0$. (On the other hand, it is concave and possibly nonsmooth when $K \geq 0$.) Furthermore, if $K(V) \leq -1$, then $\mathrm{div}(\mathrm{grad}(h)) \geq n - 1$, which combined with Stokes' formula yields the isoperimetric inequality in such V, as was mentioned earlier.

We refer the reader to the survey $[\mathrm{Gro}]_{\mathrm{sign}}$ for a more comprehensive account of the geometric effect of the sign of the curvature, while here we turn to the problem of collapse.

C. Elementary geometry of collapse

8.8. If we deal with manifolds V with bounded curvature, say $|K(V)| \leq 1$, then there is no ambiguity about what collapse means: either at some point $v \in V$ the exponential map $\exp_v : T_v(V) = \mathbb{R}^n \to V$ is injective on a large, or rather, not on a very small ball $B_0(\rho)$, i.e. for ρ comparable to 1, or the exponential map loses injectivity on a small ρ-ball in $T_v(V) = \mathbb{R}^n$. In the first case, going under the heading of "injectivity radius of V at v is $\geq \rho$ for $\rho \approx 1$," we say that V is noncollapsed at v, and if the *injectivity radius*, denoted $\mathrm{InjRad}_v(V)$ (defined as the maximal ρ where exp is injective) is small, much smaller than one, we say V is collapsed at v. To be precise, we must specify what is small and what is large, which can often be ambiguous. A simple way out is to deal with sequences of manifolds (V_i, v_i) with $|K(V_i)| \leq 1$ where noncollapse signifies

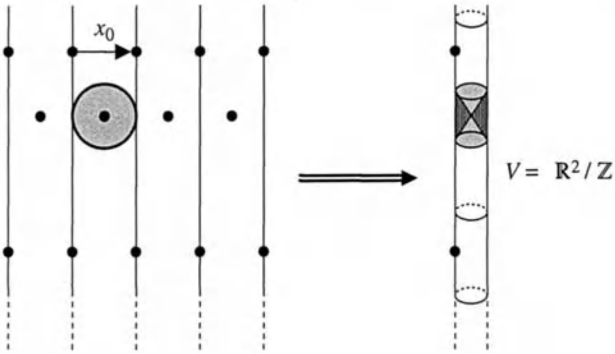
$$\mathrm{InjRad}_{v_i}(V_i) \geq \rho > 0 \tag{No-Co}$$

for $i = 1, 2, \ldots$, while collapse refers to

$$\mathrm{InjRad}_{v_i}(V_i) \to 0 \tag{Co}$$

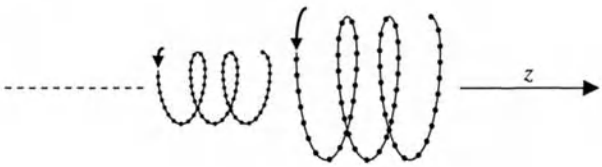
as $i \to \infty$. By passing to a subsequence, one always falls into one of these two possibilities, either (No-Co) or (Co).

8.9. Basic flat example. Let $V = \mathbb{R}^n/\Gamma$ for some isometry group Γ freely acting on $\mathbb{R}^n$, and look at V at the point v_0 corresponding to $0 \in \mathbb{R}^n$. The easiest case is where Γ acts by parallel translation, say Γ, generated by a single γ moving $x \mapsto x + x_0$.



Here, collapse progresses as $\|x_0\| \to 0$, where the cylinder $V = \mathbb{R}^n/\mathbb{Z}$ Hausdorff converges to $\mathbb{R}^{n-1}$. And the injectivity radius of V equals $\|x_0\|/2$ at all $v \in V$.

A more interesting picture arises with a single isometry γ on $\mathbb{R}^3$ which acts on (x, y, z)-space $\mathbb{R}^3$ by translating the z-axis along itself by 2ρ and by rotating the (xy)-plane by α. The orbit of different points here have different geometries. Points in the z-axis remain there to form just a copy of straight $\mathbb{Z}$, while the points away from the z-axis spiral about it. The specifics depend on the relative sizes of ρ and α, as well as on the distance from the z-axis.



The injectivity radius of V at the points v on the closed geodesic $S \subset V$ corresponding to the z-axis equal ρ, and it increases as $v \in V$ goes away from S. In fact, the displacement by γ of a point $\tilde{v} \in \mathbb{R}^3$ within distance d from the z-axis is approximately $d\alpha + \rho$. This may convey the wrong impression that the injectivity radius there is of the order at least $d\alpha$ no matter how small ρ is. However, as we go far from the z-axis, the minimal displacement of $\tilde{v}$ is caused not by γ but rather by some γ^p, $p > 1$, which turns around the z-axis and comes back closer to $\tilde{v}$ than to $\gamma(v)$. For example, in the limit case where $\rho = 0$, every point $\tilde{v} \in \mathbb{R}^3$ rotates by an angle α remaining in a fixed circle around the z-axis. Thus, some power $\gamma^p(\tilde{v})$ comes arbitrarily close to $\tilde{v}$. In general, to see what happens, we evaluate the volume of a ball $B_v(R) \subset V$. We denote by $N_v(R)$ the number of points of the orbit of $\tilde{v} \in \mathbb{R}^3$ over v in the Euclidean R-ball around $\tilde{v}$ and observe that $\mathrm{vol}\, B_v(R) \approx R^3/N_v(R)$, provided that R is significantly larger than $r = \mathrm{InjRad}_v(V)$, say $R > 100r$. On the other hand, the number $N_v(R)$ is pinched roughly between $N_{min} = R/2r$ and $N_{max} = R^3/r^3$. Indeed, the minimal distance between the points in the orbit of $\tilde{v}$ equals 2r, and so we must have at least $R/2r$ points in the orbit before it leaves the ball $\tilde{B}_{\tilde{v}}(R) \subset \mathbb{R}^3$ starting from the center. (Notice that every orbit eventually goes to infinity in $\mathbb{R}^3$ as the action has no fixed point.)

On the other hand, $N(R)$ cannot exceed the number of disjoint r-balls in $\tilde{B}(R)$, and the r-balls around the orbit points are disjoint, which makes $N(R) \leq R^3/r^3$ by the obvious volume count. It follows that $\mathrm{vol}\, B_v(R) \geq r^3$, which is obvious anyway, since the r-ball $B_{\tilde{v}}(r)$ with $r \leq \mathrm{InjRad}_v(V)$ *injects* into V under the map $\mathbb{R}^3 \to V$ by the very definition of the injectivity radius. (In fact, a little thought shows that $\mathrm{vol}_v(B_v(R)) \geq r^2 R$ in our case, since the diameter of $B_v(R)$ is at least R.) Now, let $d = \mathrm{dist}(v, S)$, so that $B_v(R) \subset B_{v_0}(d+R)$ for $v_0 \in S$. The volume of $B_{v_0}(d+R)$ is approximately $\rho(d+R)^2$ (where $\rho = \mathrm{InjRad}_{v_0} V$, $v_0 \in S$), and so

$$r^3 = (\mathrm{InjRad}_v(V))^3 \leq \mathrm{vol}\, B_v(R) \leq \rho(d+R)^2.$$

We choose $R = d$ and conclude that the upper bound on r in terms of $d = \mathrm{dist}(v, S)$ is

$$r \leq (\rho d^2)^{1/3}.$$

In particular, if the injectivity radius ρ is small at v_0, it remains small at all $v \in V$ not too far from v_0. We shall see below that the dependence of $\mathrm{InjRad}_v V$ on $v \in V$ persists for all V under the condition $|K(V)| \leq 1$.

8.10. Lemma: *Let V be a complete Riemannian manifold with $|K(V)| \leq 1$ and $\mathrm{InjRad}_v(V) = \rho \leq 1$. Consider the exponential map $\exp_v$ from the R-ball $\tilde{B}_0(R) \subset \mathbb{R}^n = T_v(V)$ to V for $R \in [\rho, 1]$. Then*

(a) $\mathrm{card}(\exp_v^{-1}(v)) \geq R/\rho$.

(b) $\mathrm{card}(\exp_v^{-1}(v')) \geq (R-d)/\rho$ *for* $d = \mathrm{dist}(v, v')$ *and all* $v' \in V$.

(c) $\mathrm{vol}\, B_v(R) \leq \mathrm{const}_n\, \rho R^{n-1}$.

Proof. We equip the ball $\tilde{B} = \tilde{B}_0(R)$ with the Riemannian metric induced by $\exp_v$ from V, so that $\exp_v \colon \tilde{B} \to V$ becomes locally isometric. We claim that this $\exp_v$ behaves like the restriction of a Galois covering map "$\tilde{V}$"$\to V$ to an R-ball in this ghost covering "$\tilde{V}$." Namely, we claim that there is a *pseudogroup* Γ of fixed-point free isometries operating on $\tilde{B}$ such that $\tilde{B}/\Gamma$ corresponds to the ball $B_v(R) \subset V$. This means that each $\gamma \in \Gamma$ acts by a *partial* isometry on $\tilde{B}$, such that whenever the domain of definition $D_\gamma \subset \tilde{B}$ of γ contains a point $\tilde{v}$, it also contains the r-ball $B_{\tilde{V}}(r) \subset \tilde{B}$

for $r = \min(\mathrm{dist}(\tilde{v}, \partial\tilde{B}), \mathrm{dist}(\gamma(\tilde{v}), \partial\tilde{B}))$, where we note that $\mathrm{dist}(v, \partial\tilde{B}) = R - \mathrm{dist}(\tilde{v}, \tilde{v}_0)$ for the center $\tilde{v}_0$ of $\tilde{B}$ and all $\tilde{v} \in \tilde{B}$. Then we have the usual rules: if γ_1 is defined at some $\tilde{v}_1 \in \tilde{B}$, i.e. $\tilde{v}_1 \in D_{\gamma_1}$ and γ_2 is defined at $\gamma_1(v_1)$, then $\gamma_2\gamma_1$ is defined at v_1 and sends it to $\gamma_2\gamma_1(v_1) = \gamma_2(\gamma_1(v_1))$. All this is shown as in the corresponding standard proof of these properties for coverings with two points kept in mind:

I. The path covering property holds true for our "partial covering" map $\tilde{B} \to V$ insofar as the paths in question do not reach the boundary $\partial\tilde{B}$ of $\tilde{B}$. Thus we have covering paths in $\tilde{B}$ issuing from $\tilde{v} \in \tilde{B}$ as long as the lengths of these paths do not exceed $R - \mathrm{dist}(\tilde{v}_0, \tilde{v})$.

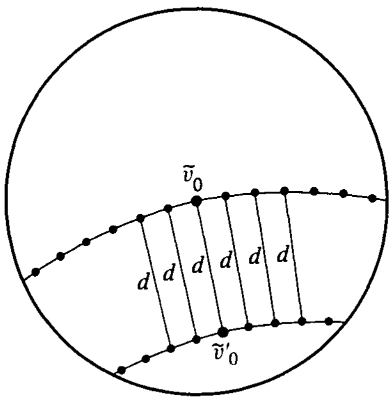
II. The ball $\tilde{B}$ with the metric induced from V is *geometrically simply connected.* In fact, there is a canonical homotopy between any two paths in $\tilde{B}$, since *every two points in* $\tilde{B}$ *can be joined by a unique geodesic segment* (see (5) below).

Granted I and II, we do have our pseudogroup Γ acting on $\tilde{B}$, where the Γ-orbits in $\tilde{B}$ identify with the $\exp_v$ pullbacks of points in V. (Here the reader should recall the elementary theory of covering spaces and replay it in the framework given by I and II.)

Now, to prove (a), we must show that the Γ-orbit of $\tilde{v}_0$ reaches the boundary of $\tilde{B}$. Indeed, we could otherwise take a minimal (radius) ball $B_{\tilde{v}'}(R') \subset \tilde{B}$ containing this orbit. Such a ball is unique (see (5) below), and so its center $\tilde{v}'$ would be fixed under Γ contrary to the freedom of the action of Γ. Thus, we have

$$\mathrm{card}\exp_v^{-1}(v) = \mathrm{card}\,\Gamma)(\tilde{v}_0) \geq R/r,$$

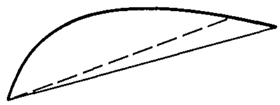
and the inequality (b) for $\Gamma(\tilde{v}')$ is clear from the picture.



Finally, we have (c), since $\operatorname{vol}\tilde{B} \simeq \mathbb{R}^n$ and $\tilde{B}$ covers the ball $B_v(R/2) \subset V$ with multiplicity at least $R/2\rho$. This bounds $\operatorname{vol}(B_v(R/2))$, and then the bound on $\operatorname{vol} B_v(R)$ follows by Bishop's inequality.

8.11. *Let V be a complete manifold (possibly) with convex boundary where no geodesic segment of length $\leq \ell$ contains a conjugate point. Then every geodesic 2-gon consisting of two segments of lengths $\leq \ell$ can be homotoped to a closed geodesic in V.*

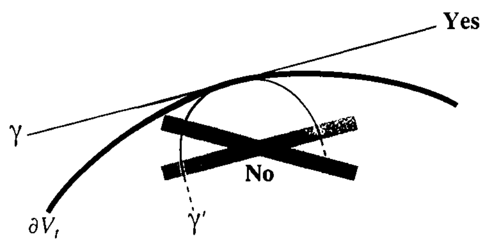
Proof. There is an obvious homotopy diminishing the lengths of the sides of the 2-gon,



which keeps diminishing the length unless we arrive at a closed geodesic. (This fails in the presence of conjugate points, as is seen for a pair of geodesic segments joining two opposite points on S^n.)

8.12. *Let V be a Riemannian manifold exhausted by a continuously increasing family of compact submanifolds V_t, $t \in [0,1]$, with strictly convex boundaries, $V = \bigcup_{t=0}^{1} V_t$, where $V_0 = \{v_0\}$. Then V contains no closed geodesic γ.*

Proof. Take the minimal V_t containing γ and observe that the boundary of this V_t would meet γ, which is clearly impossible.



8.13. Corollary: *Let $K(V) \leq 1$ and consider the unit ball $\tilde{B} = \tilde{B}(1) \subset \mathbb{R}^n = T_v(V)$ with the Riemannian metric induced by $\exp_v$ from V. Then every two points in $\tilde{V}$ are joined by a unique geodesic segment in B.*

Proof. A pair of segments between $\tilde{v}$ and $\tilde{v}'$ in $\tilde{B}$ could be deformed to a closed geodesic in $\tilde{B}$, which contradicts (4).

(6) *Let V be a complete manifold (possibly) with convex boundary where* *every ball of radius $\leq R$ has strictly convex boundary. Then every subset $V_0 \subset V$ of diameter $\leq R$ is constrained in a unique ball of minimal radius denoted $B_{v_0}(R_0) \supset V_0$.*

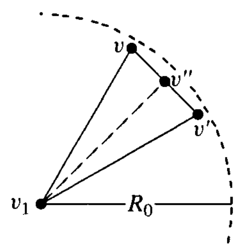
Proof. If V_0 is contained in the intersection $B_v(R_0) \cap B_{v'}(R_0)$, then it lies in a smaller ball with center v'' between v and v', since the inequalities

$$\operatorname{dist}(v, v_1) \leq R_0 \qquad \text{and} \qquad \operatorname{dist}(v', v_1) \leq R_0$$

imply

$$\operatorname{dist}(v'', v_1) < R_0$$

for the center v'' of the minimal segment $[v, v'] \subset V$ (where the uniqueness of the minimizing segments follows from the convexity of the balls).



8.14. Remarks. (*a*) The properties (*a*) and (*b*) in 8.11 need only $K(V) \leq 1$ (but not $K(V) \geq 1$), as follows from our argument which relies at this point only on the convexity of the balls in $\tilde{B}$.

(*b*) Property (*c*) in 8.11 can be restated and proved under the assumption $K(V) \leq 1$ with a suitably modified notion of the injectivity radius.

(*c*) One could prove (*a*) and (*b*) in 8.11 using only the Rauch theorem without any appeal to the convexity considerations. In fact, all one needs is the existence of a multiball $\tilde{B} \to V$ of radius R around v whose metric (induced from V) is (bi-)Lipschitz close to the Euclidean one. (We suggest that the reader work out the precise statement and proof.)

(*d*) Lemma 8.11 shows that there is an abrupt change in the behavior of the function $r \mapsto \operatorname{vol}_v(B(r))$ in the presence of collapse: it switches from the growth $\simeq r^n$ to something slower than r^{n-1} for r crossing the threshold of $\operatorname{InjRad}_v(V)$. For example, one cannot have $\operatorname{vol} B_v(r) \simeq \varepsilon r^n$ for a very small fixed ε and $r \to 0$ if $|K(V)| \leq 1$. On the other hand, this is possible for $K(V) \leq 1$, where the borderline between collapse and noncollapse is fuzzier than for $|K(V)| \leq 1$.

8.15. Variation of $\mathrm{InjRad}_v(V)$ **with** $v \in V$. *If* $|K(V)| \leq 1$, *then the injectivity radius* r *at* $v' \in V$ *can be bounded in terms of* $\rho = \mathrm{InjRad}_v(V)$ *and* $d = \mathrm{dist}(v, v')$ *as follows*

$$r \leq \mathrm{const}'_n (\rho(d+1)^{n-1} \exp(n(d+1)))^{1/n} \qquad (**)$$

in the range $0 \leq \rho \leq 1$ *and* $0 \leq r \leq 1$ *(where the latter means that* $(**)$ *is not applicable if its right hand side is* ≥ 1*).*

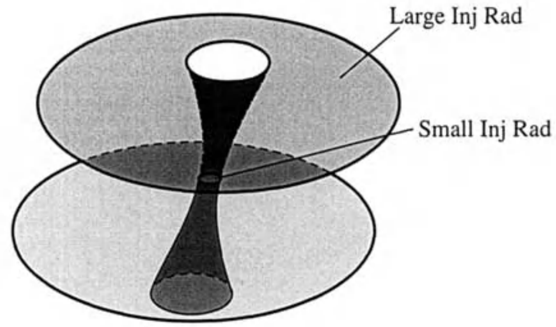
Proof. The $(d+r)$-ball around v satisfies

$$\mathrm{vol}_v B(d+r) \leq \mathrm{const}_n \, \rho(d+r)^{n-1} \exp(n(d+r)),$$

as follows from (c) in (2) and Bishop's inequality. On the other hand, the r-ball at v' has volume $\geq r^n$. Thus, $(**)$ follows from the inequality $\mathrm{vol}\, B_{v'}(r) \leq \mathrm{vol}\, B_v(d+1)$ for $r \leq 1$.

The meaning of $(**)$ is that collapse of V at a single point implies it everywhere in V.

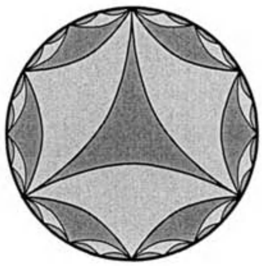
8.16. Remarks. (a) One could *define* collapse at v by requiring that the balls $B_v(R) \subset V$ have volumes significantly smaller than R^n, and then the above would hold under the assumption Ricci $\geq -(n-1)$ (or just the doubling property for balls, see below). On the other hand, this is not so if we allow Ricci $\to -\infty$, as the following picture shows.



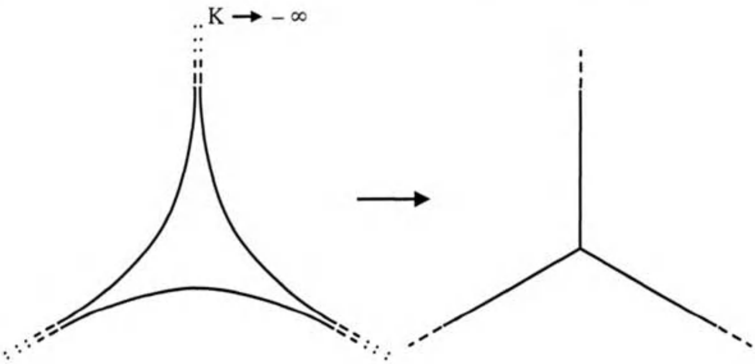
(b) Actually, our definition of collapse via InjRad $\to 0$ was solely adjusted to the bound $|K(V)| \leq -1$. In general, the collapse can (and should) be defined by a more appropriate condition which must be stable under small perturbations in a reasonable moduli space of manifolds, e.g., in the

Lipschitz metric (where InjRad is highly discontinuous), and which must express some closeness of V to a space of dimension $< \dim V$.

Example: A good intuitive measure of such collapse is the Uryson width or $\operatorname{diam}_{n-1} V$. On the other hand, complete, simply connected manifolds V with $K(V) \leq -1$ which are as far from Uryson collapse as anything can be, still converge to something 1-dimensional if $K(V) \to -\infty$. For example, if $\dim V = 2$, then one can subdivide V with $K(V) \leq \varepsilon < 0$ into ideal triangles,



and every such triangle converges to an infinite tripod for $K \to \infty$.



Thus, V's converge to a tree made out of a continuum of such tripods glued together.

8.17. Collapse for Ricci ≥ -1. Let us look more closely at the *volume collapse on the scale R* defined by

$$\operatorname{vol} B(R) \leq \delta R^n$$

for a given small δ. This condition propagates from R to larger scale by Bishop's theorem (with δ growing somewhat with R) and, on the other hand, this *volume collapse* implies *Uryson collapse* by the following

Theorem: *There is a critical value $\delta(n) > 0$ such that the inequality* $\operatorname{vol} B_v(R) \leq \delta(n) R^n$ *at all points $v \in V$ implies that the $(n-1)$-th Uryson width (or diameter measuring the distance of our n-dimensional V from an $(n-1)$-dimensional polyhedron) satisfies*

$$\operatorname{diam}_{n-1}(V) \leq 7R,$$

provided that V is a complete manifold with Ricci ≥ -1.

Proof. Use our old map $\Phi \colon V \to Q$ to the nerve Q of a covering of V by R-balls (see 5.33). Then, as was explained in 5.33, the local volume bound allows one to push this map to the $(n-1)$ skeleton $Q_{n-1} \subset Q$ and thus obtain a map $V \to Q_{n-1}$ with the diameters of the pull-back of all points bounded by 7R (where 7 is for safety).

Notice that the bound on $\operatorname{diam}_{n-1} V$ (trivially) gives us a similar bound on the filling radius of V, and this is known to be large for some V. For example, it is infinite for the universal covering of compact aspherical manifolds and more generally of essential manifolds.

Corollary: *Let V be a compact essential (e.g., aspherical) manifold with* Ricci ≥ -1 *and with fundamental group in the class $\mathcal{G}_C$ of 5.23 with $c \geq c_0 \geq 0$. Then there exists a point $v \in V$ with*

$$\operatorname{vol} B_v(1) \geq \text{const} = \text{const}(n, \operatorname{diam} V, c_0) > 0.$$

Proof. Take the point $\tilde{v} \in \tilde{V}$, where a ball $B_{\tilde{v}}(r) \subset \tilde{V}$ injects to V with $r > c$ (see 5.24) and observe that the volumes of all balls in $\tilde{V}$ are bounded in terms of a single one, and the diameter of V by Bishop's theorem.

Questions. (a) Is it essential to have Ricci ≥ -1 in the above theorem? Probably there are (by necessity rather complicated) examples where the volume collapse is *not* accompanied by Uryson (or filling radius) collapse, but still some condition much weaker than Ricci $\geq$ const may ensure the validity of the theorem.

(b) Can one make the const above independent of $\operatorname{diam} V$ and/or of c_0? One knows, for example, that if all balls in V are small, then the simplicial volume $\|[V]_D\|$ vanishes (see [Gro]$_{\mathrm{VBC}}$), which provides the positive answer in many cases.

D. Convergence without collapse

8.18. Consider the class of Riemannian manifolds V of dimension n distinguished by the conditions

$$|K(V)| \leq C < \infty$$

and

$$\operatorname{InjRad}_v(V) \geq \rho > 0.$$

We claim that the Hausdorff and the Lipschitz metrics d_H, d_L on the moduli space of such V's are essentially equivalent (compare [Che]$_{\text{Fin}}$). Namely,

8.19. *If $d_H(V, V') \leq \delta$ in our class, then the Lipschitz distance is also small $d_L(V, V') \leq \Delta$ for $\Delta = \Delta(\delta, C, \rho) \to 0$ as $\delta \to 0$.*

Proof. The idea is to embed V and V' into some Hilbert space $\mathbb{R}^N$ (which is infinite-dimensional if V and V' are noncompact) in a rather canonical way, so that their images will be C^1-close whenever V and V' are Hausdorff close. Then we take the normal projection of V to V' for the required bi-Lipschitz map $V \to V'$.

Let us explain our construction of the embedding $\Phi : V \to \mathbb{R}^N$. We choose an ε-separated ε-net in V, denoted $\{v_i\} \subset V$, $i = 1, 2, \ldots$ and take the space $\mathbb{R}^N$ of ℓ_2-functions on the set $\{v_i\}$ so that $N = \operatorname{card}\{v_i\}$. Then every $v \in V$ is sent to the vector $v \mapsto \{\varphi(\varepsilon^{-1} \operatorname{dist}(v, v_i))\}$, where φ is some standard smooth function with the following properties

$$\varphi(d) = d^2 \qquad \text{for} \qquad d \leq 5$$

$$\varphi(d) = 0 \qquad \text{for} \qquad d \geq 6.$$

Observe that the image of an R-ball $B(R) \subset V$ under Φ lands in a subspace $\mathbb{R}^{N_0} \subset \mathbb{R}^N$ with $N_0 = N_0(n, R)$. Furthermore, the map $\Phi|_{B(R)}$ is entirely determined by what happens on the concentric ball $B(R + 6\varepsilon)$ and so the study of Φ is essentially a local matter. We shall be interested in the picture where ε is small, $\varepsilon << 1$, as well as $\varepsilon << \operatorname{InjRad} V$, and we shall be looking at balls of radius, say 8ε, where the geometry is essentially Euclidean . So, we look first at our Φ in the case of $V = \mathbb{R}^n$ and observe that this $\Phi = \Phi^\circ : \mathbb{R}^n \to \mathbb{R}^\infty$ enjoys the following properties

(A) Φ° is a smooth embedding, i.e., the function $\varphi_i(v) = \varphi(\operatorname{dist}(v, v_i))$ separates the points in $\mathbb{R}^n$, and the differential $D\Phi^\circ : T(\mathbb{R}^n) \to \mathbb{R}^\infty$ is everywhere injective.

(B) The curvature of the embedded $\Phi^\circ(\mathbb{R}^n) \subset \mathbb{R}^\infty$, i.e. the norm of the second fundamental form, is bounded by some universal constant const $=$ const_n.

(C) Φ° is invariant under the scaling: if we pass from an ε-net $\{v_i\} \subset \mathbb{R}^n$ to the $\lambda\varepsilon$-net $\lambda\{v_i\} \subset \lambda\mathbb{R}^n = \mathbb{R}^n$, we get the same image in $\mathbb{R}^\infty$.

The proof of (A), (B), and (C) is trivial. In fact, one can say more than that. Namely, there is the following (equally obvious) quantitative version of (A):

(A′) The metric in $\mathbb{R}^n$ induced from $\mathbb{R}^\infty$, say dist* is comparable to the Euclidean dist,

$$C_1 \operatorname{dist}(v_1, v_2) \leq \operatorname{dist}^*(v_1, v_2) \leq C_2 \operatorname{dist}(v_1, v_2)$$

for all v_1 and v_2 in V satisfying $\operatorname{dist}(v_1, v_2) \leq 10\varepsilon$ and for some universal constants C_1, C_2 depending only on n.

Now, let us compare Φ° on $\mathbb{R}^n$ with Φ on an arbitrary V where ε is small compared to $|K(V)|^{-1}$ and $\operatorname{InjRad}(V)$. Since everything is scale invariant, we may assume $\varepsilon = 1$ while $|K(V)|^{-1}$ and $\operatorname{InjRad} V$ are very large, so that the balls, say $B(7) \subset V$ on which we study Φ are close to the Euclidean $B(7)$ in the Lipschitz metric.

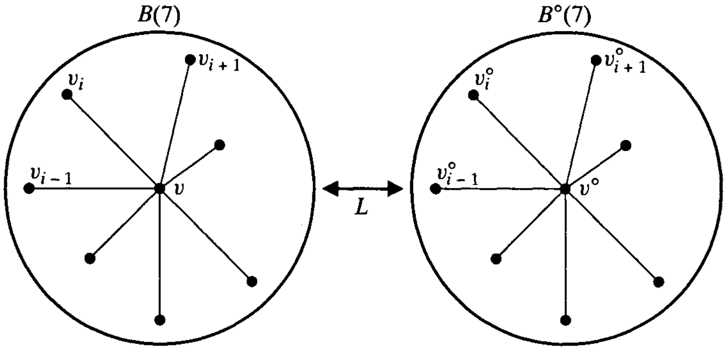
So, we fix such a $(1+\Delta_0)$-bi-Lipschitz diffeomorphism

$$V \supset B(7) \stackrel{L}{\longleftrightarrow} B^\circ(7) \subset \mathbb{R}^n$$

and compare Φ on $B(7) \subset V$ to Φ° on $B^\circ(7) \subset \mathbb{R}^n$, where the net in $\mathbb{R}^n$ corresponds to a given net $\{v_i\} \subset V$ for $v_i \stackrel{L}{\longleftrightarrow} v_i^\circ$. We claim that:

If $|K(V)| \leq \kappa = \kappa(n) > 0$ and $\operatorname{InjRad}(V) \geq 20$, then the map $\Phi\colon V \to \mathbb{R}^n$ satisfies the same properties (A), (B), (C), (A′) possibly with different, but still universal, implied constants. Furthermore, the Φ-image of $B(7)$ in $\mathbb{R}^N$ lies C^1-close to the Φ°-image of $B^\circ(7)$, where "close" means Δ-close with $\Delta = \Delta_n(\kappa) \xrightarrow{x \to 0} 0$ for $\kappa \to 0$.

Proof. Properties (A), (C), and (A′) follow from the Rauch theorem, while (B) is immediate with our bounds on the curvatures of spheres in V by the curvature. The most important feature to grasp is the differential $D\Phi(v)$ as it compares to $D\Phi^\circ(v^\circ)$ at the corresponding point $v^\circ \in \mathbb{R}^n$. These differentials are determined by the gradients of the corresponding functions $\varphi(\operatorname{dist}_V(v, v_i))$ and $\varphi(\operatorname{dist}_{\mathbb{R}^n}(v^\circ, v_i^\circ))$, and the Rauch theorem applied to $\exp_v : T_v(V) \to V$ shows that the metrics in V and $\mathbb{R}^n$ induced from $\mathbb{R}^N$ are mutually close. In fact, Φ and Φ° distort the original metrics in V and in $\mathbb{R}^n$ in the same way, since the gradients of the functions $\varphi(\operatorname{dist}(v, v_i))$ at v are directed by the geodesic segments $[v, v_i]$ with the norm equal to $|\varphi'(0)|$.



It follows that the normal projection of $\Phi(B(7))$ to $\Phi^\circ(\mathbb{R}^n)$ (as well as of $\Phi(B^\circ(7))$ to $\Phi(V)$) is an almost isometric map, i.e. $(1+\Delta)$-bi-Lipschitz with $\Delta \to 0$ as $\kappa \to 0$.

Now everything is prepared to study two Riemannian manifolds V, V' which are Hausdorff-close. Here, by the definition of d_H, we have ε-nets $\{v_i\} \subset V$ and $\{v_i'\} \subset V'$, where the correspondence $v_i \overset{H}{\longleftrightarrow} v_i'$ is almost isometric, and where the nets can be assumed to be ε-separated. So, we have two embeddings $\Phi : V \to \mathbb{R}^N$ and $\Phi' : V' \to \mathbb{R}^N$ which are both almost parallel to $\Phi^\circ : \mathbb{R}^n \to \mathbb{R}^N$ on the balls of radii 7ε in V and V'. It follows that the normal projection P of $\Phi(V')$ to $\Phi(V)$ gives us the desired $(1+\Delta)$-bi-Lipschitz map

$$V' \xrightarrow{\Phi'} \Phi'(V') \xrightarrow{P} \Phi(V) \xrightarrow{\Phi^{-1}} V.$$

As an additional bonus, our (B) gives us a bound on the derivatives of our diffeomorphism $V' \to V$ which will come in handy later on.

Encouraging remark. All this may appear rather elaborate at first glance with many details needing to be verified. However, once one gets a clear picture of Φ for $V = \mathbb{R}^n$ and accepts that all V's are locally very close to $\mathbb{R}^n$, as far as the distance functions and their first and second derivatives are concerned (provided, of course, that $|K(V)|$ is small and $\operatorname{InjRad} V$ is not very small), then everything becomes obvious up to a point where one cannot even remember what the problem was to begin with.

8.20. $C^{1,1}$-convergence. The class of manifolds with $|K(V)| \leq C$ form a Hausdorff precompact family, and the extra condition $\operatorname{InjRad}_{v_0}(V) \geq \rho_0 > 0$ makes it *Lipschitz precompact*, since the lower bound on the injectivity radius propagates from point to point, as we saw earlier, and then (1) applies. This shows that the Hausdorff limits V_∞ of manifolds V_i from our class are Riemannian manifolds and that the convergence $V_i \to V_\infty$ is Lipschitz.

Moreover, this limit manifold V_∞ admits a natural $C^{1,1}$ structure, where $C^{1,1}$ refers to C^1 smooth functions with Lipschitz derivatives. Indeed, our V_i's were embedded into $\mathbb{R}^N$ with a *uniform control on the curvature* of the embeddings and uniformly C^1 functions become Lipschitz in the limit.

Thus, we arrive at the class of *Riemannian* $C^{1,1}$ *manifolds* V, which means the following:

I. V has a $C^{1,1}$ atlas with the metric tensor g of V being Lipschitz in these $C^{1,1}$-coordinates.

II. The squared distance function dist_g^2 is *locally* $C^{1,1}$.

III. Every two nearby points in V are joined by a unique minimal geodesic.

IV. The exponential map $\exp_v\ T_v(V) \to V$ is globally defined (if V is complete) and locally Lipschitz. Moreover, it satisfies Rauch bounds on $D\exp_v$ corresponding to the inequality $|K| < C$.

V. There exists a bounded, measureable quadratic form B_S satisfying the definition of the curvature for equidistant deformations of hypersurfaces in V given in 8.7.

Summing up all of the above, we conclude with the following

$C^{1,1}$-compactness theorem: *The class of the above manifolds V with the additional requirement* $\mathrm{InjRad}_{v_0}(V) \geq \rho_0 > 0$ *is compact in the Lipschitz (as well as Hausdorff) metric for pointed manifolds. Moreover, the implied bi-Lipschitz maps $L(R)$ from R-balls $B_{v_0}(R) \subset V_i$ to V_∞ can be chosen $C^{1,1}$ with the bounds on the Lipschitz constants of the differentials $DL_i(R)$ depending only on R, for all $R > 0$. In particular, if the V_i are compact with diameters bounded independently of i, then the V_i are diffeomorphic to V_∞ for all sufficiently large i.*

Remarks. (a) This result for *flat tori* is equivalent to the Mahler compactness theorem and the case of locally homogeneous manifolds is due to Chabauty (see [Chab], [Cass], p. 134, and $[\mathrm{Che}]_{\mathrm{Fin}}$).

(b) Since g has a (generalized) bounded curvature and since curvature carries the essential (Diff-invariant) information on the second derivatives of g, one can expect that g is $C^{1,1}$ in suitable coordinates. In fact, it receives a regularity boost if one uses the harmonic coordinates in V, i.e. those given by harmonic functions (see [Greene] and the references therein).

8.21. Geometry of V with $|K(V)| \leq 1$ on the scale $\rho_0 = \mathrm{InjRad}_{v_0}(V)$. Suppose that we have our V with small ρ_0 and then scale it by ρ_0^{-1}. Now the family $\{\rho_0^{-1}V_0, v_0\}$ is amenable to the compactness theorem above for $\rho_0^{-1} \to \infty$. Here, if the $\rho_0^{-1}V$ converge to some V_∞ for $\rho_0 \to 0$, then the limit manifold V_∞ is necessarily (and obviously, since $K(\rho^{-1}(V_i)) = \rho_0^2 K(V) \to 0$) *flat.* Thus, the geometry of each ball $B_{v_0}(\lambda\rho_0) \subset V$ is close to that of a $\lambda\rho_0$-ball in a flat manifold. This applies to all $\lambda \geq 1$ with the implied closeness depending on λ and becoming "infinitely close" for every fixed λ as $\rho_0 \to 0$. In other words, everything we know about flat manifolds essentially extends to all V if we are concerned with balls $B_{v_0}(R)$ with R comparable to ρ_0.

8.22. Pinching without collapse. We look again at our stability problem (often referred to as "pinching" in the geometric literature). If the manifolds V' we study, which are infinitesimally (or locally) close to the model V_0, all have $|K(V')| \leq C$ and $\mathrm{InjRad}(V') \geq \rho$, then, assuming that they also have uniformly bounded diameter, we conclude that they must be close to some development V of V_0. Indeed, since we have a sequence of such manifolds V_i getting locally closer and closer to V_0, we can take a convergent subsequence and pass to the limit V'_∞. This manifold, on the one hand, is locally *isometric* to V_0 as "close" between V'_i and V_0 becomes "infinitely close" in the limit. On the other hand, the V'_i are globally close to V'_∞ by the preceding discussion.

8.23. Criteria for noncollapse. It is psychologically easy to include the bound $|K(V')| \leq C$ into the local closeness of V' to V_0 condition, since V_0, being locally homogeneous, has bounded curvature. But the lower bound on $\mathrm{InjRad}(V')$ is of a global nature and cannot be accepted without some resistance. Fortunately, there are some criteria for this encoded into the local geometry of V_0, such as nonvanishing of some Chern-Weil form on V_0. In fact, the most interesting of these is the (generalized) Cheeger criterion, which requires the 2-connectedness of the simply connected development of V_0 (see $[\mathrm{Gro}]_{\mathrm{Stab}}$ and the references therein). Notice, however, that even the innocuous looking condition $|K(V')| \leq$ const somewhat limits the logical structure of the general stability problem and, in fact, most of the results established (in the present-day literature) for $|K| \leq C$ extends to the general setting of 8.5. (Some of this is briefly indicated in 8.25 below.)

Besides local criteria, there are global topological ones for noncollapse of metrics on V. For example, let V admit a proper map of nonzero degree onto a closed manifold V_0 with negative sectional curvature. Then it cannot collapse if we keep $|K(V)| \leq \mathrm{const}_1$ and $\mathrm{diam}(V) \leq \mathrm{const}_2$, as follows trivially from 5.24, in the case where $\pi_1(V) = \pi_1(V_0)$. The general case (where one needs a trivial generalization of 5.24 to nonuniversal coverings $\tilde{V} \to V$) is left to the reader.

8.24. Remarks and references. (*a*) Most ideas of this section can be traced to Cheeger's thesis. Actually, our compactness theorem is a restatement (actually known to Cheeger) of *Cheeger's finiteness theorem*, which claims that *there are at most finitely many diffeomorphism classes of Riemannian manifolds V satisfying*

1. $\dim V = n$,
2. $|K(V)| \leq \text{const}_1$,
3. $\text{diam}(V) \leq \text{const}_2$.

Also the essential ideas of pinching and flat approximation on the scale of $\text{InjRad}\, V$ are due to Cheeger.

In his original proof, Cheeger (see $[\text{Che}]_{\text{Fin}}$) was looking at the matching between exponential charts where the Lipschitz constants (of the coordinate transformations) are controlled by the Rauch theorem. In fact, one could forget at this stage about the curvature and just *postulate* bi-Lipschitz closeness of the unit balls in V to the Euclidean ones. This would still lead to the finiteness and Lipschitz compactness conclusion following Shikata (compare [Shik]). Also, one can handle in this manner the general pinching problem with V' being Lipschitz-close to V_0 without an extra curvature assumption on V'.

(*b*) An efficient use of the second derivatives and/or covariant parallel fields appears in various proofs (Ruh, Grove, Karcher, see [G–K–R]) of the differential sphere theorem, some of which even apply to Hilbert manifolds and show, in particular, that *a sufficiently strongly pinched simply connected Hilbert manifold V is bi-Lipschitz equivalent to the sphere S^∞ with the Lipschitz distance between V and S^∞ controlled by the pinching.* The same applies to other infinite dimensional model spaces covered by *finitely many* standard ball-shaped domains, e.g., products $S^\infty \times S^\infty \times S^\infty \times V_0^n$, $n < \infty$. (See the expository article $[\text{Gro}]_{\text{Stab}}$ for other aspects of pinching.)

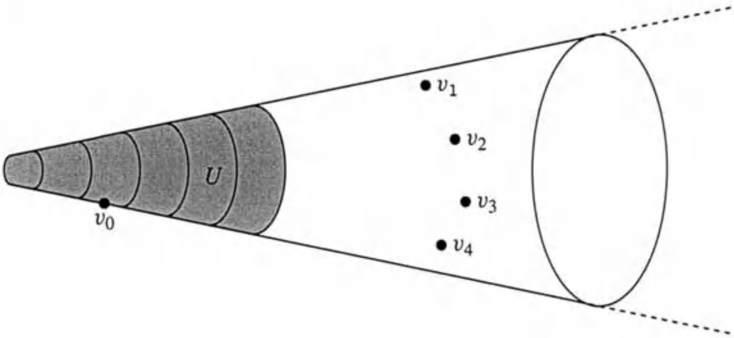
(*c*) The study of noncollapsed manifolds with $K(V) \geq -1$ was initiated by Grove and Peterson, and then an essential local geometric property was proven by Perelman.

If $K(V) \geq -1$ and $\text{vol}\, B_{v_0}(1) \geq \varepsilon$, then the point v_0 admits a convex neighborhood $U_0 \subset V$ containing the δ-ball $B_{v_0}(\delta)$ for $\delta \geq \delta_n(\varepsilon) > 0$, $n = \dim V$.

Idea of the proof. Take points $v_i \in V$, $i = 1, 2, \dots$, roughly within unit distance from v_0 and consider the functions

$$d_i(v) = (2 - \mathrm{dist}(v_i, v))^N$$

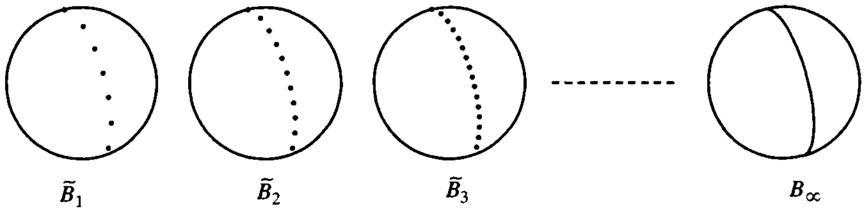
for large N. Such a function is "strongly convex in many directions" near v_0, i.e., its second derivative is large and positive along many geodesics. And it is never too concave. (All this follows from the discussion in 8.7.) Then, one can show, (using the lower volume bound $B_{v_0}(1) \geq \varepsilon$ that the sum $\sum_i d_i(v)$ for suitably distributed points $v_i \in V$ is convex and its levels provide enough convex subsets in V to accomplish what we want (see [Per]$_{\mathrm{SCBB}}$ for further results and references).



(d) The Lipschitz regularity of the Riemannian metric tensor g on our $C^{1,1}$ manifolds V can be upgraded to $C^{1,\alpha}$, $\alpha > 0$, by use of *harmonic coordinates* on V, built of harmonic (instead of distance) functions (see [Sak], [Pugh], [Pet], [Peters], [Katsu], [Gr–Wu], [Greene] and the references therein).

E. Basic features of collapse

8.25. Let us look at the Hausdorff limit $V_\infty = \lim_{i\to\infty} V_i$ of a sequence of manifolds V_i with $|K(V)| \leq 1$ and $\mathrm{InjRad}\, V_i \to 0$ near some point $v_\infty \in V_\infty$. To see what it looks like, we consider the corresponding points $v_i \in V_i$ converging to v_∞ and take the exponential unit (multi-)balls $\tilde{B}_i \to V_i$ at v_i with the metrics induced from V_i. Each $\tilde{B}_i$ is isometrically acted upon by a discrete pseudogroup Γ_i, and these define in the limit some isometry pseudogroup Γ_∞ on the $C^{1,1}$-limit $\tilde{B}_\infty$ of $\tilde{B}_i$ (see 8.20). Since the minimal displacement of $\tilde{v}_i \in \tilde{B}$ by Γ_i, i.e. $2\,\mathrm{InjRad}_{v_i}(V_i)$ goes to zero with i, this limit pseudogroup Γ_∞ is not discrete anymore: it contains a connected normal subpseudogroup $\Delta_\infty \subset \Gamma_\infty$ acting isometrically on $\tilde{B}_\infty$.



It is not hard to see that Δ_∞ makes a local Lie group action on $\tilde{B}_\infty$ (the $C^{1,1}$-regularity is enough for this, believe me), and then this local action extends to a global action of a Lie group, say $\hat{\Delta}_\infty \supset \Delta_\infty$ on some manifold $\hat{B}_\infty \supset \tilde{B}_\infty$, such that $\hat{B}/\Delta_\infty = \tilde{B}_\infty/\Delta_\infty$.

We add to $\hat{\Delta}_\infty$ those elements γ in Γ which fix $\tilde{v}_\infty$ and thus send the orbit $\tilde{\Delta}_\infty(\tilde{v}_\infty)$ into itself and denote the resulting group (not just pseudogroup!) by $\hat{\Gamma}_\infty \supset \hat{\Delta}_\infty$. The rest of Γ_∞ acts freely on $\tilde{B}_\infty$, and so we see that V_∞ *is locally isometric at* v_∞ to $\hat{B}_\infty/\hat{\Gamma}_\infty$. This gives a fair idea of the local geometry of V_∞.

8.26. The description above does not tell much about the geometry of the V_i themselves and/or about Γ_i acting on $\tilde{B}_i$. For example, it remains unclear at this stage if the action of the pseudogroup Γ_i comes from a group $\hat{\Gamma}_i$ acting on some $\hat{B}_i \supset \tilde{B}$. The first case one may try to look at is where $\operatorname{diam}(V_i) \to 0$, and so the V_i converge to a single point where the description above is totally vacuous. The manifolds V where $|K(V)| \leq 1$ and the diameters are very small are called *almost flat.* Obviously *flat* Riemannian manifolds are almost flat. Furthermore, if $V' \to V$ is a circle bundle over an almost flat manifold V, then V' is also almost flat by a Berger-like consideration (see 3.11), and one can even allow more general fibers which are flat Riemannian manifolds insofar as the structure group of the fibration is isometric on the fibers. Actually, one can prove that this covers all almost flat possibilities, but one gains in perspective by switching the point of view and identifying the above manifolds with *infranil manifolds.* (The idea of relating almost flatness with nilmanifolds was suggested to me by G. Margulis about 25 years ago.) Actually, a suitable version of the *Margulis lemma* says that the pseudogroup Γ_i acting on $\tilde{B}_i$ is virtually nilpotent (compare [Zassen]). Then one can show that it does extend to an action of a virtually nilpotent *group* $\hat{\Gamma}_i \supset \Gamma_i$ on some $\hat{B}_i$ (see [Bus–Kar]). Eventually, one can go deeper into the geometry of this action and then analyze how these fit together on different (multi-)balls in V_i, thus getting a pretty good picture of the local geometry of the collapsed manifolds V in terms of the so-called *N-structure* (see [Che–Fu–Gr] and the references therein). As an upshot of this, one can prove the existence of the *critical radius*, i.e., a number $\rho = \rho_n > 0$ such that the metric g on a manifold (V, g) with $|K| \leq 1$ and $\mathrm{InjRad}_v \leq \rho$ for all $v \in V$ admits a deformation g_t so that $\mathrm{InjRad}_v(V, g_t) \overset{t\to\infty}{\longrightarrow} 0$ with the curvature remaining bounded by 1. On the other hand, one does not know if there is a similar critical value for the volume (see [Che–Rong], [Fang–Rong], [Pet–Tus], and [Pet–Rong–Tus] in this regard).

We conclude by noting that the structure of collapse has also been studied for manifolds with $K \geq -1$ and with Ricci ≥ -1. For example, Perelman (see $[\mathrm{Per}]_{\mathrm{Wid}}$, $[\mathrm{Col}]_{\mathrm{ARC}}$, [Grom] and the references therein) proved that collapsed manifolds with $K \geq -1$ look (very) roughly like convex subsets in $\mathbb{R}^n$ or like solids $[0, \varepsilon_1] \times [0, \varepsilon_2] \times \cdots \times [0, \varepsilon_n]$ for some $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_n$. Namely, he has shown that, say for $K \geq 0$, the volume of V is commensurable to the product of Uryson diameters (widths)

$$C_1 \prod_{i=0}^{n-1} \mathrm{diam}_i(V) \leq \mathrm{vol}\, V \leq C_2 \prod_{i=1}^{n1} \mathrm{diam}_i\, V$$

for some positive constants C_1 and C_2 depending on n.

The most challenging open problem is to work out the concept of collapse for manifolds with scalar curvature bounded from below by a constant C, at least in the case $C > 0$.

Appendix A

"Quasiconvex" Domains in $\mathbb{R}^n$

by P. Pansu

A.1. Definition: A subset X in $\mathbb{R}^n$ is called *C-quasiconvex* if, for all points $x, y \in X$, there is an arc joining x to y in X having length at most $C\|x - y\|$.

A.2. Remark: In 1.14, we saw that 1-quasiconvex subsets are actually convex, and that a $\pi/2$-quasiconvex set is necessarily simply connected.

On the regular polyhedron with $n + 1$ vertices, inscribed in the unit sphere in $\mathbb{R}^n$, consider two contiguous vertices a, b. Let α denote their distance apart with respect to the metric on the sphere: α is the angle between the vectors a and b, and is given by the condition $\cos(\alpha) = -1/n$. Let τ denote the ratio of the angle α and the length of the edge ab, i.e., $\tau = \alpha/2\sin(\alpha/2)$.

A.3. Proposition: *Let X be a compact, C-quasiconvex subset of $\mathbb{R}^n$ with $C < \tau$. Consider $\mathbb{R}^n$ as the complement of a point in S^n. If X is precisely the support of a positive finite measure, then $S^n \setminus X$ is diffeomorphic to a ball.*

The proof consists of an application of Morse theory to a smooth version of the distance function from X. Given a finite measure μ whose support equals X, we construct a smooth, convex, and decreasing function $f \colon \mathbb{R}_+^* \to$

$\mathbb{R}_+^*$ such that the integral $D(x) = \int f(\|x - z\|)\,d\mu(z)$ is finite for precisely those points x not belonging to X. The condition of quasiconvexity ensures that the smooth function D on $\mathbb{R}^n \setminus X$ has no critical points.

A.4. Lemma: *For each positive, finite measure μ on $\mathbb{R}^n$ whose support equals X, and for each $r > 0$, the number $s(r) = \sup\{\mu(X)/\mu(B(x,r)) : x \in X\}$ is finite.*

Recall that the support of a positive measure is defined by the condition that $\mathbb{R}^n \setminus \mathrm{Supp}(\mu)$ is the union of all open subsets of $\mathbb{R}^n$ having zero measure.

Fix a positive measure μ whose support equals X. Then for each $r > 0$ and $x \in X$, we have $\mu(B(x,r)) > 0$. Now observe that, in fact, $\mu(B(x,r)) \geq \varepsilon(r) > 0$ for $x \in X$, $r > 0$. Otherwise, there would exist a sequence $x_i \in X$ such that $\mu B(x_i, r) \to 0$. By passing to a subsequence, we may assume that the x_i converge to $x \in X$ and that $\mu(B(x_i, r)) \leq 2^{-i}$ for each i. Then

$$B(x,r) \subset \bigcap_{m \geq 0} \bigcup_{i=m}^{\infty} B(x_i, r),$$

and so for each m,

$$\mu(B(x,r)) \leq \sum_{i=m}^{\infty} \mu(B(x_i, r)) \leq 2^{-m+1},$$

and finally, $\mu(B(x,r)) = 0$, which is impossible.

A.5. Lemma: *For each $a > 0$ and $\lambda > 1$, there exists a smooth, strictly decreasing function $f : (0, a] \to \mathbb{R}_+$ which bounds s from above and whose derivative f' is increasing and satisfies $-f'(r) \geq 2\sqrt{n}s((\lambda - 1)r)(-f'(\lambda r))$ for $r \leq a/\lambda$.*

Proof. We first construct a continuous, piecewise affine function s_1 having the desired properties: s_1 is affine on each interval $[\lambda^{-k-1}a, \lambda^{-k}a]$, of slope p_k. We set $s_1(a) = s(a/\lambda)$ and $p_1 = (\lambda/(\lambda - 1)a)(s(\lambda^{-1}a) - s(\lambda^{-2}a))$. Then we inductively define p_k by setting

$$p_k = \sup \left\{ \begin{array}{c} -p_{k-1}, -p_{k-1} 2\sqrt{n}\lambda s((\lambda - 1)\lambda^{-k}a) \\ -(\lambda^{k+1}/(\lambda - 1)a)(s(\lambda^{-k-1}a) - s(\lambda^{-k}a)) \end{array} \right\}.$$

By construction, $p_k < 0$, so s_1 is decreasing $p_k \leq p_{k-1}$, s_1' is increasing, and $s_1(\lambda^{-k}a) \geq s(\lambda^{-k-1}a)$; thus for each $r \in [\lambda^{-k-1}a, \lambda^{-k}a]$,

$$s_1(r) \geq s_1(\lambda^{-k}a) \geq s(\lambda^{-k-1}a) \geq s(r);$$

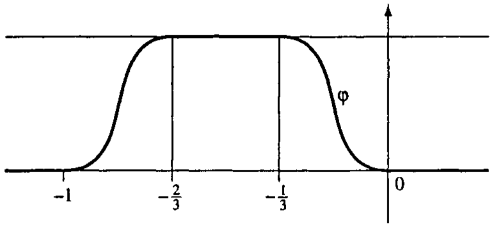
finally, for $r \in (\lambda^{-k-1}a, \lambda^{-k}a]$ we have

$$\begin{aligned} -s_1'(r) &= -p_k \geq -p_{k-1}2\sqrt{n}\lambda s((\lambda-1)\lambda^{-k}a) \\ &\geq (-s_1'(\lambda r))2\sqrt{n}\lambda s((\lambda-1)r). \end{aligned}$$

It remains only to perform a convolution

$$f(r) = 3\int s_1(r+t)\varphi(2t/r)\,dt,$$

where φ is smooth, equals 1 on $[-2/3, -1/3]$, and vanishes outside of $(-1, 0)$.



Then f is smooth, and

$$f(r) \geq s_1(r) \geq s(r);$$

like s_1', the derivative f' of f is negative and increasing, and we have

$$\begin{aligned} -f'(r) &= 3\int s_1'(r+t)\varphi(2t/r)\,dt \\ &\geq 6\sqrt{n}\lambda \int -s_1'(\lambda r + \lambda t)s((\lambda-1)(r+t))\varphi(2t/r)\,dt \\ &= 6\sqrt{n}\int_{-\lambda r/2}^{0} -s_1'(\lambda r + t)s((\lambda-1)(r+t/\lambda))\varphi(2t/\lambda r)\,dt \\ &\geq 6\sqrt{n}s(\lambda-1)r)\int -s_1'(\lambda r+t)\varphi(2t/\lambda r)\,dt \\ &= 2\sqrt{n}s((\lambda-1)r)(-f'(\lambda r)) \end{aligned}$$

since $t \leq 0 \Rightarrow s((\lambda-1)(r+t/\lambda)) \geq s((\lambda-1)r)$.

In the following, we set $a = \operatorname{diam}(X)(\lambda/\lambda-1)$, and we extend f to a smooth, strictly decreasing function on $\mathbb{R}_+^*$.

A.6. Lemma: *The function* $D(x) = \int f(\|z-x\|)\,d\mu(Z)$ *is finite precisely on* $\mathbb{R}^n \setminus X$.

Proof. By hypothesis, μ is finite, and if $x \notin X$, the function $z \mapsto f(\|z-x\|)$ is bounded above by $f(d(x,X)) < \infty$; thus, $D(x)$ is finite. Conversely, if

$x \in X$, then for each r such that $0 < r < a$,

$$\begin{aligned} D(x) &\geq \int_{B(x,r)} s(\|z-x\|)\,d\mu(z) \\ &\geq \int_{B(x,r)} s(r)\,d\mu(z) \\ &= s(r)\mu(B(x,r)) \\ &\geq \mu(X), \end{aligned}$$

and so the integral cannot be finite.

A.7. Lemma: *Suppose, moreover, that X is C-quasiconvex, with $C < \tau$. There exists a $\lambda > 1$ such that for each $x \in \mathbb{R}^n \setminus X$, the distance from x to the convex hull of $X \cap B(x, \lambda d(x,X))$ is at least*

$$\left(\frac{1}{n} + \lambda^2 \left(\frac{1}{n^2} - \frac{1}{n}\right)\right)^{1/2} d(x,X).$$

Proof. Fix $x \in \mathbb{R}^n \setminus X$. We will first show that if λ is sufficiently close to 1, then x is not in the convex hull of $X \cap B(x, \lambda d)$, where $d = d(x,X)$. If it were, then we could write

$$\sum_{i=0}^{n} t_i(z_i - x) = 0,$$

where $t_i \geq 0$, $\sum_i t_i = 1$, and $z_i \in X$, $\|z_i - x\| \in [d, \lambda d]$. Then, for each j, $\sum_{i=0}^{n} t_i\langle z_i - x, z_j - x\rangle = 0$, or in other words, $\sum_{i\neq j} t_i\langle z_i - x, z_j - x\rangle/t_j\|z_j - x\|^2 = -1$. Let j be the index for which the number $t_j\|z_j - x\|$ is largest. Then there exists an $i \neq j$ such that $t_i\langle z_i - x, z_j - x\rangle/t_j\|z_j - x\|^2 \leq (1/n)$, and so the cosine of the angle β between $z_i - x$ and $z_j - x$ is at most equal to $-(1/n)$. Every path connecting z_i to z_j in X has length at most βd, so that $\|z_j - z_i\|$ is at most $2\lambda d \sin(\beta/2)$. The condition of C-quasiconvexity implies that $\beta/2\lambda\sin(\beta/2) \leq C < \alpha/2\sin(\alpha/2)$ so that, since $\beta \geq \alpha$, $\lambda \geq C2\sin(\alpha/2)/\alpha$. If we fix $\lambda < C2\sin(\alpha/2)/\alpha$ so that, for each $x \in X$, x is distinct from its orthogonal projection x' onto the convex hull of $X \cap B(X, \lambda d)$. In particular, x' lies on the boundary of the convex hull, so that we can write

$$x' = \sum_{i=1}^{n} t_i z_i \qquad \text{with} \qquad \sum_{i=1}^{n} t_i = 1.$$

Knowing that for each i, j we have $\langle z_i - x, z_j - x\rangle/\|z_i - x\|\|z_j - x\| \geq -(1/n)$, we can compute

$$\begin{aligned}
\|x' - x\| &= \sum_{i=1}^{n} t_i^2 \|z_i - x\|^2 + \sum_{i \neq j} t_i t_j \langle z_i - x, z_j - x\rangle \\
&\geq \sum_{i=1}^{n} t_i^2 d^2 + \sum_{i\neq j} t_i t_j \left(\frac{1}{n}\lambda^2 d^2\right) \\
&= d^2 \left[\left(\sum_{i=1}^{n} t_i^2\right)\left(1 + \frac{1}{n}\lambda^2\right) - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)^2 \lambda^2\right] \\
&\geq \frac{1}{n} + \lambda^2 \left(\frac{1}{n^2} - \frac{1}{n}\right) d^2,
\end{aligned}$$

since the function $(t_1, \dots, t_n) \mapsto t_1^2 + \cdots + t_n^2$ attains its minimum at the point $t_1 = \cdots = t_n = 1/n$.

A.8. Lemma: *If X is C-quasiconvex, with $C < \tau$, then the function D of Lemma A.6 is smooth and has no critical points on $\mathbb{R}^n \setminus X$.*

Proof. Since the measure μ is finite and since the functions $z \mapsto f(\|z - x\|)$ and $z \mapsto (d^p/dz^p) f(\|z - x\|)$ are uniformly bounded in a neighborhood of each point of $\mathbb{R}^n \setminus X$, the function D is of class C^∞. The gradient of D at a point $x \in X$ is $\operatorname{grad}_x D = \int (f'(\|z - x\|)/\|z - x\|)(x - z)\, d\mu(z)$. Set $d = d(x, X)$ and fix a λ satisfying the condition of Lemma A.7. The point

$$g = x + \frac{1}{I} \int_{B(x,\lambda d)} \frac{f'(\|z - x\|)}{\|z - x\|}(x - z)\, d\mu(z),$$

where

$$I = \int_{B(x,\lambda d)} -\frac{f'(\|z - x\|)}{\|z - x\|}\, d\mu(z)$$

is the center of mass of $X \cap B(x, \lambda d)$ equipped with the density $(-f'(\|z - x\|)/\|z - x\|)\mu$, and is therefore contained in the convex hull of $X \cap B(x, \lambda d)$, so that by Lemma A.7

$$\|x - g\| \geq \left(\frac{1}{n} + \lambda^2\left(\frac{1}{n^2} - \frac{1}{n}\right)\right)^{1/2} d.$$

On the other hand, $(1/I)\operatorname{grad}_x D - (x - g)$ equals the remainder of the integral, so that

$$\begin{aligned}
\left\| \int_{\mathbb{R}^n \setminus B(x,\lambda d)} \frac{f'(\|z - x\|)}{\|z - x\|}(x - z)\, d\mu(z) \right\| &\leq \int_{\mathbb{R}^n \setminus B} -f'(\lambda d)\, d\mu(z) \\
&\leq -\mu(X) f'(\lambda d),
\end{aligned}$$

so that

$$I = \int_{B(x,\lambda d)} -\frac{f'(\|z-x\|)}{\|z-x\|}\, d\mu(z) \geq -\frac{f'(d)}{d}\mu(B(x,\lambda d)),$$

and we obtain

$$\left\|\frac{1}{I}\operatorname{grad}_x D - (x-g)\right\| \leq d\mu(X)\frac{f'(\lambda d)}{f'(d)}\mu(B(x,\lambda d)).$$

However, there exists a $z \in X$ such that $\|z-x\| = d$. Thus $B(z,(\lambda-1)d) \subset B(x,\lambda d)$, so that $\mu(X)/\mu(B(x,\lambda d)) \leq \mu(X)/\mu(B(z,(\lambda-1)d)) \leq s((\lambda-1)d)$, which gives the inequality $\|(1/I)\operatorname{grad}_x D-(x-g)\| < \|x-g\|$ for $d \leq a/\lambda = \operatorname{diam}(X)/\lambda - 1$, and λ sufficiently close to 1. For $d > \operatorname{diam}(X)/\lambda - 1$, $X \subset B(x,\lambda d)$, so $(1/I)\operatorname{grad}_x D = x - g$. In both cases, we conclude that $\operatorname{grad}_x D \neq 0$.

A.8. Proof of the Proposition: The function f was constructed explicitly on $[0,a]$, but until now we have only used the fact that on $[a,\infty)$, f is smooth and strictly decreasing. We can therefore assume that for sufficiently large r, $f(r) = 1/r^2$. Then D tends to 0 at infinity. To verify that the natural extension of D to S^n is smooth, it suffices to compose D with the inversion $x \mapsto x/\|x\|^2$, which is a chart from $\mathbb{R}^n$ onto $S^n \setminus \{0\}$. We may assume that 0 is the center of mass of X with respect to the measure μ, in which case it is easy to check that if D_0 is the function D associated with the set $\{0\}$ equipped with the Dirac measure $\mu(X)$, then

$$D\left(\frac{X}{\|X\|^2}\right) = D_0\left(\frac{X}{\|X\|^2}\right) + 0(\|X\|^3),$$

which proves that D is thrice-differentiable at 0, and has the same derivative and hessian as D_0. But $D_0(X/\|X\|) = \|X\|^2$, and so D_0 (and D) admit a nondegenerate critical point at ∞ having index n. Morse theory (cf. [Milnor]$_{\mathrm{MT}}$) then enables us to conclude that $S^n \setminus X$ is diffeomorphic to a ball.

Remark: We can construct the diffeomorphism $\varphi : B^n \to S^n \setminus x$ in an equivariant way, i.e., there exists a homomorphism $h : \operatorname{Isom}(X) \to \operatorname{Isom}(B^n)$ such that for each isometry i fixing X, we have $i \circ \varphi = \varphi \circ h(i)$. Indeed, the compact Lie group $\operatorname{Isom}(X)$ admits a fixed point μ in the compact, convex set of measures μ of support X, mass 1, such that

$$\mu = \int_{\operatorname{Isom}(X)} g^*\mu_0\, dg$$

for a Haar measure dg on $\operatorname{Isom}(X)$.

A.9. Corollary: *Under the hypotheses of the proposition, and if* X *is also an absolute neighborhood retract having the homotopy type of a CW complex, then* X *is contractible.*

Indeed, it follows from Alexander duality (see [Span], Ch. 7) that X has trivial cohomology and is thus contractible.

Appendix B

Metric Spaces and Mappings Seen at Many Scales

by Stephen Semmes

This appendix describes some geometric ideas from classical analysis and quasiconformal mappings. We shall study spaces and mappings at all locations and scales at once. Every ball in a metric space provides a snapshot, of the space itself or a mapping on it. We shall often consider situations in which these snapshots are "bounded". We shall encounter a variety of forms of boundedness of snapshots. A bound on a single snapshot is not so exciting, but uniform bounds on all snapshots can lead to interesting structure.

An important point is to use measure theory. With measure theory we can see interesting phenomena which happen "most of the time". We shall sometimes see hidden rigidities, in which there is special structure that does not occur in all contexts and which is not explicit in the definitions.

In Part I we review some definitions, making precise the idea of bounded snapshots in various ways. We shall also review examples of metric spaces and a few basic facts to show how the concepts fit together. In Part II we go through some machinery for working with spaces and functions which have bounds on their snapshots. The results discussed there work in great generality and are not very mysterious. In Part III we concentrate on more mysterious phenomena that involve special structure and hidden rigidities. Part IV provides an introduction to real variable methods that support much of the material in Part III and which deal with the interplay between measure theory and the geometry of snapshots.

I. Basic concepts and examples

B.1. Euclidean spaces, hyperbolic spaces, and ideas from analysis.

Euclidean spaces and hyperbolic spaces are basic examples of metric spaces that one thinks of as being very different. They *are* very different, but they can also be seen as different ways to look at the same thing.

Euclidean space means $\mathbb{R}^n$ with the usual metric $|x-y|$. For hyperbolic space we use the upper half-space model. Set $\mathbb{R}^{n+1}_+ = \{x = (x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : x_{n+1} > 0\}$. The hyperbolic Riemannian metric on $\mathbb{R}^{n+1}_+$ is defined by

$$\frac{dx_1^2 + \cdots + dx_{n+1}^2}{x_{n+1}^2}. \tag{B.1.1}$$

Let $h(x,y)$ be the geodesic distance function on $\mathbb{R}^{n+1}_+$ associated to this Riemannian metric, so that $h(x,y)$ is the length with respect to (1.1) of the shortest curve in $\mathbb{R}^{n+1}_+$ which connects x to y. One can write down an explicit formula for $h(x,y)$, but I think that it is easier to understand some cruder considerations. Fix a point $x \in \mathbb{R}^{n+1}_+$. If $y \in \mathbb{R}^{n+1}_+$ satisfies

$$|x-y| \le \frac{x_{n+1}}{2}, \tag{B.1.2}$$

so that y is substantially closer to x than to the boundary hyperplane $\mathbb{R}^n \times \{0\}$, then $h(x,y)$ is bounded by a universal constant. Thus if you start at x and move a distance at most $x_{n+1}/2$ in the Euclidean metric, then you move only a bounded distance in the hyperbolic metric. More generally, if

$$|x-y| \le k\, x_{n+1} \quad \text{and} \quad k^{-1} x_{n+1} \le y_{n+1} \le k\, x_{n+1} \tag{B.1.3}$$

for some constant $k > 0$, then $h(x,y) \le C(k)$ for some constant $C(k)$ that does not depend on x or y. Conversely, if $h(x,y) \le A$ for some constant A, then (1.3) holds for some constant k that depends only on A and not on x or y.

So how is it that hyperbolic spaces and Euclidean spaces describe the same thing?

Think of $\mathbb{R}^{n+1}_+$ as parameterizing all the balls in $\mathbb{R}^n$. That is, if $x \in \mathbb{R}^{n+1}_+$, then we set $x' = (x_1, \ldots, x_n)$, we get a ball $B(x) = B(x', x_{n+1})$ in $\mathbb{R}^n$, and all balls arise this way. If we view $\mathbb{R}^{n+1}_+$ as the space of balls in $\mathbb{R}^n$, then the coarse geometry described above is very natural. To say that $x, y \in \mathbb{R}^{n+1}_+$ have bounded hyperbolic distance $h(x, y)$ means that the balls $B(x)$ and $B(y)$ are pretty similar, i.e., they have approximately the same radius (to within a bounded factor), and the distances between the centers is bounded by a constant times the radius.

We can think of this relationship between Euclidean space $\mathbb{R}^n$ and hyperbolic space $\mathbb{R}^{n+1}_+$ in the following way. Think of $\mathbb{R}^n$ as being the place where we live, like the surface of the Earth, and think of $\mathbb{R}^{n+1}_+$ as being the sky above. We might be able to fly around in $\mathbb{R}^{n+1}_+$ like an insect or a bird or a helicopter and look down on $\mathbb{R}^n$ below. If we are at the point $x \in \mathbb{R}^{n+1}_+$, then the ball $B(x)$ is roughly the part of $\mathbb{R}^n$ that we can see best. We cannot see things in $B(x)$ too clearly, we cannot make distinctions that exist at distances much smaller than x_{n+1}. Our vision is limited both by location and scale. If we need to see something more precisely we have to fly down closer to the ground. If we want a broad perspective, then we have to fly high.

This idea plays a very important role in analysis. Suppose that $n = 1$, so that we can think of $\mathbb{R}^2_+$ as being the upper half-plane in the complex plane. Suppose that $f(z)$ is a bounded holomorphic function on $\mathbb{R}^2_+$. It turns out that the gradient of f is automatically bounded in the hyperbolic metric. In other words f has to be almost constant on small hyperbolic balls. This can be checked using the Cauchy integral formula. If we think of f as being represented as the Cauchy integral of its boundary values on the real axis – i.e.,

$$f(z) = \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{f(t)}{t - z}\, dz \tag{B.1.4}$$

for $z \in \mathbb{R}^2_+$ – then the value $f(z)$ is most sensitive to the boundary values $f(t)$ for $t \in B(z)$, because the Cauchy kernel $\frac{1}{t-z}$ is largest for $t \in B(z)$ and decays as t moves away from $B(z)$. (This works better if one uses the Poisson integral formula.)

These comments are not special to holomorphic functions of one complex variable. They apply to harmonic functions, and in higher dimensions. They should be seen in the context of elliptic partial differential equations.

There are versions for holomorphic functions of several complex variables, although the correct geometry is different.

Similar ideas arise in analysis in other ways. One often studies functions on $\mathbb{R}^n$ by looking at their average behavior on balls, looking at different centers and radii. This includes the average oscillations of a function on balls. In this way one often starts with a function on $\mathbb{R}^n$ and then passes to an associated function on $\mathbb{R}^{n+1}_+$ which captures some aspect of the behavior of the original function on average, at different scales and locations. For instance, given a locally integrable function $g(u)$ on $\mathbb{R}^n$, one might consider the functions $G(x)$, $H(x)$ on $\mathbb{R}^{n+1}_+$ defined by

$$G(x) = \frac{1}{|B(x)|} \int_{B(x)} g(u)\, du, \tag{B.1.5}$$

$$H(x) = G(x) - G(\widehat{x}), \quad \widehat{x} = (x_1, \ldots, x_n, 2\, x_{n+1}). \tag{B.1.6}$$

Here $|B(x)|$ the volume of the ball $B(x)$, so that $G(x)$ is just the average of g over $B(x)$, and $H(x)$ is the difference between the average of g over $B(x)$ and its double (the ball with the same center but twice the radius). One can think of $H(x)$ as measuring the oscillation of g near $B(x)$, so that $H(x)$ is small when g is almost constant on the double of $B(x)$. (This will happen when g is continuous and $B(x)$ is small, i.e., x_{n+1} is small.) The behavior of both $G(x)$ and $H(x)$ are approximately compatible with hyperbolic geometry on $\mathbb{R}^{n+1}_+$, e.g., we do not expect them to oscillate very much on small hyperbolic balls. If $g(u)$ is bounded, for instance, then $G(x)$ and $H(x)$ are Lipschitz functions with respect to the hyperbolic metric, so that the amount by which their values change when we move x is bounded by a constant times the hyperbolic distance between x and the new point.

B.2. Quasimetrics, the doubling condition, and examples of metric spaces.

For the record, a metric on a (nonempty) set M is a nonnegative function $d(x, y)$ on $M \times M$ which is symmetric, vanishes exactly when $x = y$, and satisfies the triangle inequality. We call $d(x, y)$ a quasimetric if it satisfies the same properties except that the triangle inequality is replaced by

$$d(x, z) \leq K(d(x, y) + d(y, z)) \tag{B.2.1}$$

for some constant $K > 0$ and all $x, y, z \in M$.

The diameter of a set $E \subseteq M$ with respect to $d(x, y)$ is given by $\operatorname{diam} E = \sup\{d(x, y) : x, y \in E\}$. We write $B(x, r)$ and $\overline{B}(x, r)$ for the open and closed balls with center x and radius r.

If $d(x, y)$ is a quasimetric on M, then there exists a metric $\rho(x, y)$ on M and constants $C > 0$ and $s \geq 1$ so that

$$C^{-1} \rho(x, y)^s \leq d(x, y) \leq C \rho(x, y)^s \tag{B.2.2}$$

for all $x, y \in M$. This is essentially the content of the proof of Theorem 2 in [Mac–Sego].

B.2.3. Definition: *A metric space M is said to be* doubling *if there is a constant C so that any ball B in M can be covered with at most C balls whose radius is half the radius of B.*

This is a kind of uniform finite-dimensionality condition. It is satisfied by Euclidean spaces and it is not satisfied by hyperbolic spaces. For hyperbolic spaces one has the doubling property for balls of bounded radius but not for large balls. If we think of hyperbolic spaces and Euclidean spaces as corresponding in the manner of Section 1, then the doubling property for all balls for Euclidean spaces is connected to the doubling property for balls of bounded radius for hyperbolic spaces.

The doubling property provides a way to say that all snapshots of a metric space are bounded.

B.2.4. Lemma: *A metric space M is doubling if and only if every ball B in M can be covered by a bounded number of subsets of M with diameter at most half the radius of B.*

This is easy to check.

B.2.5. Corollary: *If the metric space M is doubling, then so is any subspace of M.*

In particular any subset of a Euclidean space is doubling.

B.2.6. Definition: *Let $(M, d(x, y))$ and $(N, \rho(u, v))$ be metric spaces and let $f : M \to N$ be a mapping between them. We say that f is C-*Lipschitz *if*

$$\rho(f(x), f(y)) \leq C\, d(x, y)$$

*for all $x, y \in M$, and we say that f is C-*bi-Lipschitz *if*

$$C^{-1} d(x, y) \leq \rho(f(x), f(y)) \leq C\, d(x, y) \tag{B.2.7}$$

for all $x, y \in M$. We shall just say Lipschitz or bi-Lipschitz if we do not need to specify the constant. We say that two metric spaces are bi-Lipschitz equivalent *if there is a bi-Lipschitz mapping from one onto the other.*

Note that compositions of Lipschitz mappings are Lipschitz, and similarly for bi-Lipschitz mappings.

The Lipschitz and bi-Lipschitz conditions can be seen as ways to bound the snapshots of a given mapping.

We need to start collecting some examples of metric spaces. Let us begin with a general operation called the *snowflake functor.* If $(M, d(x,y))$ is a metric space and $0 < s < 1$, then we can get a new metric space by replacing $d(x,y)$ with $d(x,y)^s$. This is indeed a metric, because of the well-known fact that

$$(a+b)^s \leq a^s + b^s \tag{B.2.8}$$

when a, b are nonnegative real numbers and $0 < s < 1$. When $s > 1$ we do not get a metric but we do get a quasimetric.

This operation is called the snowflake functor because it is an abstract version of the familiar constructions of snowflake curves in the plane. Indeed the standard snowflake constructions give curves which are bi-Lipschitz equivalent to the snowflake functor applied to a circle or line segment (with a suitable choice of s). [The most famous snowflake curve is the Von Koch snowflake, which is obtained in the following manner. One starts with an equilateral triangle T_0 in the plane, viewed as a simple closed curve. One then replaces the middle third of each of the three sides of T_0 with a pair of segments of the same length which point outside T_0, so as to make a simple closed curve T_1 which is the union of twelve segments of equal length. One then repeats the construction to produce Jordan curves $T_2, T_3, \ldots$, with each T_j consisting of $3 \cdot 4^j$ segments of length 3^{-j} times the length of the sides of T_0. The Von Koch snowflake is obtained as the limit of the T_j's (in the Hausdorff topology for compact sets).]

Note that $(M, d(x,y))$ is doubling if and only if $(M, d(x,y)^s)$ is doubling.

As a variation on the theme of snowflakes consider $\mathbb{R}^{n+1}$ with the metric defined by

$$d(x,y) = |x' - y'| + |x_{n+1} - y_{n+1}|^{\frac{1}{2}}, \tag{B.2.9}$$

where x' denotes the point $(x_1, \ldots, x_n)$ of x in $\mathbb{R}^n$. This space is doubling. It defines the natural geometry on $\mathbb{R}^{n+1}$ for studying the heat equation, in the same way that Euclidean geometry is natural for the standard Laplacian.

Now let F be a finite set with at least two elements, and let F^∞ denote the set of sequences of elements of F. Given two elements $x = \{x_i\}_{i=1}^\infty$ and $y = \{y_i\}_{i=1}^\infty$ of F^∞ and a number $a \in (0,1]$ set

$$\begin{aligned} d_a(x,y) &= 0 \qquad \text{when } x = y \\ &= a^j \qquad \text{if } x_i = y_i \text{ for } i < j, x_j \neq y_j. \end{aligned} \tag{B.2.10}$$

This is a metric for each $a \in (0, 1]$. In fact these are *ultrametrics*, which means that

$$d_a(x, z) \le \max\{d_a(x, y), d_a(y, z)\} \qquad \text{(B.2.11)}$$

for all $x, y, z \in F^\infty$, which is much stronger than the triangle inequality. When $a = 1$ (2.10) simply gives the discrete metric on F^∞, so that distinct points in F^∞ have distance 1 from each other. For $a < 1$ these metrics all determine the same topology, for which F^∞ is homeomorphic to the Cantor set. If F has exactly two elements and $a = 1/3$, then $(F^\infty, d_a(x, y))$ is bi-Lipschitz equivalent to the standard Cantor set. For any $a \in (0, 1)$ we can make a Cantor set construction in some $\mathbb{R}^n$ which is bi-Lipschitz equivalent to $(F^\infty, d_a(x, y))$ (with $n \to \infty$ as $a \to 1$).

Note that $(F^\infty, d_a(x, y))$ is doubling when $a < 1$ but not when $a = 1$. When $0 < a < 1$ the metrics $d_a(x, y)$ are all related to each other under the snowflake functor.

Cantor sets are totally disconnected. There are some amusing sets of intermediate connectivity. The first is often called the *Sierpinski carpet.* Let Q denote the unit square $[0, 1] \times [0, 1]$ in the plane. We can decompose Q into 9 closed subsquares Q_i in the obvious manner, so that they each have sidelength $1/3$. Let us throw away the square Q_5 in the center of Q and keep the remaining eight squares Q_i, and let K denote the subset of Q which is the union of these eight squares. Now apply this same process $Q \to K$ to the eight squares in K to get a new compact subset K_2 of Q which is the union of 64 squares of sidelength $1/9$. Repeating this process indefinitely we get compact sets K_l of Q which consist of 8^l subsquares of Q with disjoint interiors and with sidelength 3^{-l}. The Sierpinski carpet is the intersection of all these sets K_l. This is very similar to the construction of the Cantor set, but it gives a set which is pathwise connected.

Another amusing set in this vein is the *Sierpinski gasket.* This is obtained by taking an equilateral triangle T (instead of a square), decomposing it into 4 equilateral triangles of the same size, throwing away the middle triangle and keeping the remaining three, and then repeating the process indefinitely. The Sierpinski gasket itself is the intersection of the sequence of compact sets obtained in this manner, and it is also pathwise connected. However there are not as many paths connecting pairs of points in the Sierpinski gasket as in the Sierpinski carpet.

One can also make fractal versions of trees using the same construction. Let Q and Q_i be as before, and this time consider the set E which is the union of the five squares Q_i which are either the central square or one of the four corners of Q. Repeat the process $Q \to E$ to get compact subsets

E_j of Q, where each E_j is the union of 5^j squares of sidelength 3^{-j}. The intersection of all the E_j's gives a fractal version of a tree.

Next we consider the Heisenberg group H_n and a natural metric on it. As a set H_n is $\mathbb{C}^n \times \mathbb{R}$, and the group operation $(z_1, \dots, z_n, t)\,(z_1', \dots, z_n', t')$ is defined as

$$\left(z_1 + z_1', \dots, z_n + z_n', +t' + 2\operatorname{Im} \sum_{j=1}^{n} z_j\, \overline{z}_j'\right). \tag{B.2.12}$$

This defines a noncommutative nilpotent group. There is a natural family of dilations on H_n, namely the mappings $\delta_r : H_n \to H_n$ defined by $\delta_r(z_1, \dots, z_n, t) = (r\, z_1, \dots, r\, z_n, r^2\, t)$, $r > 0$. These dilations preserve the group structure (2.12). In this way they are analogous to ordinary dilations on Euclidean spaces.

This structure of group and dilations automatically determines a quasimetric on H_n. To understand this we start with the "homogeneous norm"

$$\|(z_1, \dots, z_n, t)\| = \left(\sum_{j=1}^{n} |z_j|^2 + |t|\right)^{\frac{1}{2}}. \tag{B.2.13}$$

This quantity has the property that

$$\|\delta_r(x)\| = r\,\|x\| \tag{B.2.14}$$

for all $x \in H_n$ and $r > 0$, and also that

$$\|x^{-1}\| = \|x\| \tag{B.2.15}$$

for all $x \in H_n$, where x^{-1} denotes the inverse of x in H_n. (If $x = (z_1, \dots, z_n, t)$, then $x^{-1} = (-z_1, \dots, -z_n, -t)$.) For our purposes the specific form of $\|\cdot\|$ does not matter, what matters is that it satisfies (2.14) and (2.15), and that it is continuous and positive away from the origin. Any function on H_n with these properties will be bounded above and below by constant multiples of $\|\cdot\|$. We use this homogeneous norm to define a quasimetric on H_n by

$$d(x, y) = \|x^{-1}\, y\|. \tag{B.2.16}$$

This is symmetric because of (2.15), it is clearly nonnegative and vanishes exactly when $x = y$, and it is not hard to check that it satisfies the approximate triangle inequality (2.1).

This quasimetric has important symmetry properties. If $a \in H_n$ let $\lambda_a : H_n \to H_n$ denote the mapping of left translation by a (with respect to the group structure), so that $\lambda_a(x) = a\,x$. Our quasimetric $d(x, y)$ is invariant under left translations, $d(\lambda_a(x), \lambda_a(y)) = d(x, y)$ for all $x, y, a \in H_n$. It also behaves well under dilations,

$$d(\delta_r(x), \delta_r(y)) = r\, d(x, y) \tag{B.2.17}$$

for all $x, y \in H_n$ and all $r > 0$.

Notice that $d(x, y)$ is compatible with the Euclidean topology on H_n. Any other quasimetric on H_n which is compatible with the Euclidean topology and which has the same behavior under translations and dilations is practically the same as $d(x, y)$, i.e., is bounded above and below by constant multiples of $d(x, y)$.

The Heisenberg geometry is somewhat similar to that of the metric in (2.9). They both respect the parabolic dilations δ_r for instance. At the origin the Heisenberg geometry looks very much like the geometry of the metric in (2.9). This is true at other points in H_n too, except that the "axes" have to be tilted. This tilting is performed by the Heisenberg group action, and there is substantial twisting of the axes as one moves from point to point, because of the nontrivial nature of the group structure.

Is the quasimetric $d(x, y)$ actually a metric, i.e., does it satisfy the triangle inequality? I am not certain, but it does not really matter. What does matter is that there is a metric $d'(x, y)$ on H_n which is bounded above and below by constant multiples of $d(x, y)$. For instance we would get a metric if we replaced (2.13) with

$$\left(\sum_{j=1}^{n} |z_j|^4 + t^2\right)^{\frac{1}{4}}.$$

See [Kor–Rei]. Do not get lost in the formulas here, there is a simple geometric reason for there to exist such a metric $d'(x, y)$. To understand this we should specify first a certain distribution of hyperplanes in the tangent spaces of points in H_n. At the origin we identify the tangent space of H_n with $\mathbb{C}^n \times \mathbb{R}$ in the obvious manner, and we take the hyperplane $\mathbb{C}^n \times \{0\}$. At a general point $x \in H_n$ we choose the hyperplane in the tangent space to H_n at x which is the image under the differential of λ_x of the hyperplane at the origin. This defines a smooth distribution of hyperplanes in the tangent bundle of H_n, and it turns out that this distribution is completely nonintegrable, so that every pair of points can be connected by a smooth curve which remains tangent to this distribution of hyperplanes. We can get an interesting metric on H_n by taking the infimum of the lengths of the curves that connect a given pair of points and which are always tangent to our distribution of hyperplanes. This is our metric $d'(x, y)$. It is clearly larger than the Euclidean metric, but it turns out that it is compatible with the Euclidean topology too, which amounts to saying that it is bounded on compact subsets of H_n. Once we know this we can conclude that it is bounded above and below by constant multiples of $d(x, y)$ because it has the correct invariance properties. These invariance properties follow from corresponding symmetries of our hyperplane distribution. The invariance of the hyperplane distribution under left translations follows from the definition, but for the dilations it is a fact to be checked.

It is easy to see that the Heisenberg group is doubling. For the unit ball centered at the origin this is clear, just a matter of compactness, and the general case can be reduced to that of the unit ball using translations and dilations.

Let us now look back at the preceding examples of metric spaces as a whole. A basic point to observe is that these spaces have a lot of symmetry to them, symmetry in terms of both location and scale. "Self-similarity" might be a better description. On Euclidean spaces we have the Euclidean metric which is invariant under translations and which scales properly under dilations. The parabolic metric (2.9) enjoys similar features, as does the Heisenberg group. What about our symbolic Cantor set F^∞, the infinite product of a given finite set F? Let Σ denote the group of permutations on F^∞ which come from permuting the coordinates separately, so that Σ is an infinite product of copies of the group of permutations on F. This group acts on F in such a way as to preserve the metric given by (2.10) (no matter the choice of a). More generally a mapping σ from F^∞ onto itself will preserve the metric when σ has the property that if $x \in F^\infty$ and $y = \sigma(x)$, then y_j depends only on $x_1, \dots, x_j$ and not on x_i for $i > j$. Thus F^∞ has a large group of isometries, which we can think of as being like translations, and indeed this group of isometries is transitive. One also has analogues of dilations, namely shift mappings. This would work better on the "unbounded" version of F^∞, in which we take doubly infinite sequences of elements of F, for then the left and right shift mappings would be bijections. As it is the left shift is not one-to-one and to define a right shift we have to make a choice for the first coordinate. In any case these shifts behave fairly well with respect to the metric in (2.10), and if we were to use doubly infinite sequences then the shifts would enjoy a scaling property like (2.17). Thus F^∞ has almost as much symmetry as the other examples, except that the symmetries are discrete.

The fractal examples, the Sierpinski carpet and gasket and the fractal tree, have also a lot of symmetry but not quite as much. Instead of thinking in terms of groups for these examples it is easier to think about a more limited kind of self-similarity. Let us start with the Sierpinski carpet. Fix a point x in it and a positive integer j. At the j^{th} stage of the construction the Sierpinski carpet is a union of 8^j squares of sidelength 3^{-j}. Fix one of these squares that contains x, call it P. The construction of the Sierpinski carpet is made so that the intersection of P with the Sierpinski carpet is an exact replica of the whole Sierpinski carpet, but scaled by a factor of 3^{-j}. The same phenomena occurs with the Sierpinski gasket and the fractal tree described above.

So let us try to have less symmetry, starting with F^∞. Instead of taking an infinite product of copies of the same finite set, let us use a sequence of finite sets $\{F_j\}$, each with at least two elements. We can still define the metrics $d_a(x, y)$ as in (2.11). We can still get plenty of isometries with respect to these metrics by taking infinite products of permutations on the F_j's, for instance, but if the F_j's are not all the same and do not follow any periodic pattern then we cannot build shift mappings in the same manner as before. The metrics will be different at different scales.

We could define more complicated metrics than the $d_a(x, y)$'s in order to break the symmetry that remains (coming from the permutations).

What about the doubling property? For the metrics $d_a(x, y)$ in (2.9), say, with $0 < a < 1$. It is not hard to see that the new version of F^∞ obtained from the sequence $\{F_j\}$ will be doubling if and only if the number of elements of the F_j's is uniformly bounded.

The doubling condition can be seen as a much weaker cousin of self-similarity. Instead of saying that the space looks the same at all scales and locations, the doubling condition says that geometries that occur at different scales and locations in the space lie in a bounded family. This boundedness is not very strong. In the preceding example we could construct an interesting space from any bounded sequence of positive integers > 1 (the cardinalities of the F_j's). Self-similarity would correspond in this construction to periodic sequences.

Similarly we could modify the constructions of the Sierpinski carpet and gasket or the fractal tree so they are not so self-similar.

Note that the snowflake functor does not disturb the self-similarity properties that we have been discussing. In the next section we describe another way to deform the geometry of metric spaces.

B.3. Doubling measures and regular metric spaces, deformations of geometry, Riesz products and Riemann surfaces.

The next definition provides a notion of compatibility between measures and metrics.

B.3.1. Definition: *Let $(M, d(x, y))$ be a metric space. A positive Borel* *measure μ on M is said to be* doubling *if there exists a constant $C > 0$ so that*

$$\mu(2\,B) \leq C\,\mu(B) \tag{B.3.2}$$

for all balls B in M (and if μ is not identically zero). Here $2\,B$ denotes the ball with the same center as B but twice the radius.

A basic example of a doubling measure is Lebesgue measure on a Euclidean space. We shall give more examples soon.

There is nothing magical about the number 2 in (3.2), any number larger than 1 would work just as well. For instance, we can always iterate (3.2) to obtain

$$\mu(2^k\,B) \leq C^k\,\mu(B) \tag{B.3.3}$$

for all positive integers k.

We shall use doubling measures to deform geometries. Before we get to that let us record a nice fact.

B.3.4. Lemma: *If $(M, d(x,y))$ is a metric space which admits a doubling measure μ, then $(M, d(x,y))$ is doubling as a metric space.*

To see this fix a ball B in M with radius r say. Let A be a finite set of points in B such that $d(x,y) \geq r/2$ for all $x, y \in A$, $x \neq y$. Then A can have at most K elements, where K depends on the doubling constant but not on anything else. To see this consider the balls $B(x) = B(x, r/4)$, $x \in A$. Our assumption on A implies that these balls are pairwise disjoint, and they are all contained in $2\,B$. Thus

$$\sum_{x \in A} \mu(B(x)) \leq \mu(2\,B). \tag{B.3.5}$$

However the doubling condition (iterated as in (3.3)) implies that $\mu(B(x))$ is bounded from below by a constant multiple of $\mu(2\,B)$ for each $x \in A$. Thus (3.5) implies a bound on the number of elements of A.

Since we have a bound on the number of elements of any such set A we can take one which has a maximal number of elements. Maximality implies that if $z \in B$ then $d(x,z) < r/2$ for some $x \in A$. Thus the balls $B(x, r/2)$, $x \in A$, cover B. This implies the doubling condition, and Lemma 3.4 follows.

Now let us use doubling measures to deform geometry. Let $(M, d(x,y))$ be a metric space, let μ be a doubling measure on M, and let $\alpha > 0$ be given. Set

$$D(x,y) = \{\mu(\overline{B}(x, d(x,y))) + \mu(\overline{B}(y, d(x,y)))\}^{\alpha} \tag{B.3.6}$$

If $M = \mathbb{R}^n$ and $\alpha = 1/n$, then $D(x, y)$ gives back a constant multiple of the Euclidean metric. In general we want to think of $D(x, y)$ as defining a new geometry on M.

The definition of $D(x, y)$ may look a little strange. On a Euclidean space we could simply take the μ-measure of the smallest closed ball that contains x and y instead of the μ-measure of the balls $B(x, d(x, y))$, $B(y, d(x, y))$, but on a general metric space it is more convenient to do the latter. All variations on this theme are practically equivalent, because of the doubling condition.

B.3.7. Lemma: *Under the assumptions above, $D(x, y)$ defines a quasimetric on M.*

This is not hard to check, using the doubling condition on μ. Typically we do not get a metric, but remember that for each quasimetric we can find a metric as in (2.2).

To understand better this idea of making quasimetrics from doubling measures let us consider the case when the new geometry is essentially the same as the old one.

B.3.8. Definition: *A metric space $(M, d(x, y))$ is said to be (Ahlfors) regular of dimension s if it is complete and if there is a positive Borel measure μ on M and a constant $C > 0$ so that*

$$C^{-1}\, r^s \leq \mu(B(x, r)) \leq C\, r^s \tag{B.3.9}$$

for all $x \in M$ and all $0 < r \leq \operatorname{diam} M$.

In this case the quasimetric $D(x, y)$ is bounded above and below by constant multiples of $d(x, y)$ if we take $\alpha = 1/s$. The requirement of completeness in Definition 3.8 is not needed for this fact, but it is convenient in practice.

It turns out that the measure μ in Definition 3.8 does not play a sensitive role, in the following sense. If a metric space M is regular with dimension s, then any two measures which satisfy conditions like (3.9) are each bounded by constant multiples of each other, and s-dimensional Hausdorff measure on M will also satisfy a condition like (3.9).

To get examples of regular metric spaces we can follow our discussion of examples of metric spaces in the preceding section. Euclidean spaces are regular with s equal to their ordinary dimension and μ taken to be Lebesgue measure. If $(M, d(x, y))$ is a regular metric space of dimension s, and if we apply the snowflake functor to get the metric space $(M, d(x, y)^t)$, then this new metric space is also regular, with dimension s/t.

If we take the parabolic metric (2.9) on $\mathbb{R}^{n+1}$, then we get a regular metric space of dimension $n+2$, with μ taken to be Lebesgue measure.

The Heisenberg group H_n with the geometry discussed in Section 2 is regular with dimension $2n+2$, again taking μ to be Lebesgue measure. This uses the easy fact that Lebesgue measure is invariant under translations with respect to the group structure (2.12).

Let F be a finite set with $n \geq 2$ elements and let F^∞ be the space of sequences with values in F, as in Section 2. Let ν denote the uniform probability measure on F, which assigns to each element of F the measure $1/n$. Let μ denote the probability measure on F^∞ obtained by taking the product of infinitely many copies of ν. Then $(F^\infty, d_a(x,y))$ is a regular metric space when $0 < a < 1$, with this choice of μ, and with the dimension s determined by the formula $a^s = 1/n$. Note that s runs through all of $(0,\infty)$ as a runs through $(0,1)$.

Suppose that Z is the Sierpinski carpet, as described in Section 2. Then Z is regular with dimension s, where s satisfies $3^{-s} = 1/8$. This is because there is a natural probability measure μ on Z with the property that if P is one of the 8^j squares of sidelength 3^{-j} obtained at the j^{th} stage of the construction of Z, then $\mu(P) = 8^{-j}$.

Similarly, the Sierpinski gasket and the fractal tree described in Section 2 are both regular, and they admit measures μ analogous to the one just described.

Now let us consider some examples of doubling measures with less symmetry.

On $\mathbb{R}^n$ take the measure $\mu = |x|^a\,dx$, where $a > -n$ and dx denotes Lebesgue measure. This is doubling. We can also take $\mu = |x_1|^a\,dx$ on $\mathbb{R}^n$, where $a > -1$.

Let K denote the usual middle-thirds Cantor set on $\mathbb{R}$ and set $\mu = \operatorname{dist}(x,K)^a\,dx$. This defines a doubling measure on $\mathbb{R}$ for positive values of a and also some negative values (in a range which can be computed explicitly).

These examples are all absolutely continuous with respect to Lebesgue measure. Can a doubling measure on a Euclidean space be singular? The answer turns out to be yes. To see this we use *Riesz products*, a kind of multiplicative (lacunary) Fourier series. Let $\{a_j\}_{j=1}^\infty$ be a sequence of real numbers such that $|a_j| \leq 1$ for each j. Define a measure μ on $\mathbb{R}$ by

$$\mu = \prod_{j=1}^{\infty} (1 + a_j \cos 3^j x). \tag{B.3.10}$$

This definition requires some explanation. This infinite product should be interpreted as the limit of the finite products

$$\prod_{j=1}^{n} (1 + a_j \cos 3^j x) \tag{B.3.11}$$

as $n \to \infty$, with the limit taken in the sense of distributions. It is not hard to check that this limit exists, by multiplying out the product and then writing it as a Fourier series. (See Theorem 7.5 on p. 209 of [Zyg].) The main point is that each term in the Fourier series arises only once as a term in the expanded product. The distributional limit of (3.11) need not be a function in general, but it always is a measure, and this comes down to the fact that each of the functions in (3.11) is nonnegative. One can also show that the limiting measure μ is doubling if $\sup_j |a_j| < 1$. This is approximately what is shown in the last section of [Beur-Ahl], but unfortunately [Beur-Ahl] does not quite correspond to our present requirements. Let us sketch the main points. Suppose that I is an interval in $\mathbb{R}$, and we want to show that $\mu(2I) \leq C\,\mu(I)$. This is trivial if I has length $\geq 2\pi/3$, because μ is periodic with period $2\pi/3$. Suppose that $|I| < 2\pi/3$, and let k be the smallest positive integer such that $|I| > 2\pi/3^k$. Write

$$\begin{aligned} \mu &= \prod_{j=1}^{k-1} (1 + a_j \cos 3^j x) \cdot \mu_k, \qquad \text{where} \\ \mu_k &= \prod_{j=k}^{\infty} (1 + a_j \cos 3^j x), \end{aligned} \tag{B.3.12}$$

and where this last infinite product is interpreted in the same manner as before. We have that μ_k is periodic with period $2\pi/3^k < |I|$. One can also check that

$$\sup_{2I} \prod_{j=1}^{k-1} (1 + a_j \cos 3^j x) \leq C \inf_{2I} \prod_{j=1}^{k-1} (1 + a_j \cos 3^j x) \tag{B.3.13}$$

with a constant C that does not depend on I. Let us sketch the proof of this. Each factor $1 + a_j \cos 3^j x$ is uniformly bounded and bounded away from 0 on $\mathbb{R}$, because of our assumption that $\sup_j |a_j| < 1$. Thus no finite number of these factors matter for (3.13), but the problem for (3.13) is that there are infinitely many of them. On the other hand,

$$|\cos 3^j x - \cos 3^j y| \leq 3^j\, 2\,|I| \leq 4\pi\, 3^{j-k+1} \tag{B.3.14}$$

whenever $x, y \in 2I$. The first inequality follows from the mean value theorem, the second from our choice of k (to be as small as possible). For

$j \leq k-1$ the left side behaves like a convergent geometric series. Using this estimate and the fact that each factor $1 + a_j \cos 3^j x$ is bounded and bounded away from 0 it is not difficult to verify (3.13). Once we have (3.13) it is easy to check that $\mu(2I) \leq C\,\mu(I)$ using (3.12) and the fact that μ_k is periodic with period $< |I|$.

Thus the Riesz product (3.10) defines a doubling measure on $\mathbb{R}$. See also 8.8 (a) on p.40 of $[\text{Stein}]_{HA}$.

It turns out that the Riesz product (3.10) is completely singular with respect to Lebesgue measure if $\sum_j a_j^2 = \infty$. (See Theorem 7.6 on p.209 of [Zyg].) Thus there are doubling measures on $\mathbb{R}$ which are completely singular, i.e., have no absolutely continuous part with respect to Lebesgue measure. We can get these measures to be self-similar by taking all the a_j's to be the same. On the other hand we can get singular doubling measures that are a little bit smoother by taking a sequence $\{a_j\}$ which tends to 0.

We can also obtain singular doubling measures on any $\mathbb{R}^n$ by taking products. (In general Cartesian products of doubling measures are doubling.)

There are doubling measures on $\mathbb{R}$ (and hence on $\mathbb{R}^n$) of the form $\mu = \omega(x)\,dx$ where $\omega(x)$ is the characteristic function of a set whose complement has positive measure. Such a μ is absolutely continuous with respect to Lebesgue measure but not the other way around. See 8.8 (b) on p.40 of $[\text{Stein}]_{HA}$.

Now let us describe a class of singular doubling measures on $\mathbb{R}$ which arise from Riemann surfaces. Let S_1 and S_2 be two compact Riemann surfaces without boundary and with common genus $g > 1$. Let $f : S_1 \to S_2$ be an orientation-preserving diffeomorphism which is not homotopic to a conformal mapping. By the uniformization theorem the universal covers of S_1 and S_2 are isomorphic as Riemann surfaces to the upper half-plane $U = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$. Let $F : U \to U$ be obtained by lifting $f : S_1 \to S_2$ to a mapping between the universal covers of S_1 and S_2 and then using the uniformizing isomorphisms to get to U. We can choose these uniformizing isomorphisms so that $F(z) \to \infty$ when $z \to \infty$ in U (this is just a normalization). It is well-known that F is a "quasiconformal" mapping, and that with this normalization F extends to a homeomorphism from $\overline{U}$ to itself. In particular F induces a homeomorphism $h : \mathbb{R} \to \mathbb{R}$. The distributional derivative of this homeomorphism defines a locally finite measure μ on $\mathbb{R}$ (sometimes called the Lebesgue-Stieltjes measure associated to h), and this measure μ is doubling. (See [Beur-Ahl].) A version of the Mostow rigidity theorem implies that this measure μ must be completely singular with respect to Lebesgue measure. (See [Agard].)

Thus we get another class of singular doubling measures on the line.

These doubling measures have a lot of self-similarity, coming from the groups of Möbius transformations associated to the uniformization of S_1 and S_2. The nature of this self-similarity is amusingly different from that of the Riesz products (in the case where the a_j's are all the same, say). There is no such group for the Riesz products.

See also [Bish–Steg] concerning the doubling measures that come from Riemann surfaces, and [Tukia] for another construction of singular doubling measures.

Next let us consider our symbolic Cantor sets again. Let F be a finite set with at least two elements, and let F^∞ denote the space of sequences which take values in F. Let $\{\nu_j\}_{j=1}^\infty$ be a sequence of probability measures on F, and let μ denote the probability measure on F^∞ which is their product. Let $\epsilon > 0$ be given, and let us assume that each ν_j has the property that it gives mass at least ϵ to each point in F. Then μ is a doubling measure on F^∞ with respect to any of the metrics $d_a(x, y)$ defined in (2.10) with $0 < a < 1$. This is not hard to check. These measures are analogous to Riesz products. They will typically be singular with respect to the "uniform" measure that we defined on F^∞.

Notice that we can obtain new doubling measures from old ones, by adding them, or taking products of doubling measures on products of metric spaces.

Keep in mind that we can use doubling measures to deform geometry on a space, as in (3.6), so that our examples should be seen also as examples of geometry. To put the relationship between geometry and measure into perspective let us consider a simple situation. Suppose that we are given a doubling measure μ on $\mathbb{R}^n$, and assume that it is mutually absolutely continuous with respect to Lebesgue measure. This means that we can write μ as $\omega(x)\,dx$ for some positive locally integrable function $\omega(x)$ on $\mathbb{R}^n$. The associated quasidistance $D(x, y)$ defined in (3.6) with $\alpha = 1/n$ is then approximately Euclidean most of the time. For instance, if z is a Lebesgue point of ω, so that

$$\lim_{r \to 0} r^{-n} \int_{B(z,r)} |\omega(u) - \omega(z)|\,du = 0, \tag{B.3.15}$$

then $D(x, y)$ looks asymptotically like a multiple of the Euclidean metric near z. (Recall that almost every $z \in \mathbb{R}^n$ is a Lebesgue point of ω.) Thus $(\mathbb{R}^n, D(x, y))$ is approximately Euclidean almost everywhere in this case. We shall encounter this idea again in Part III.

B.4. Quasisymmetric mappings and deformations of geometry from doubling measures.

The following definition provides a useful concept for comparing the geometry of two different metric spaces.

B.4.1. Definition: *Let $(M, d(x, y))$ and $(N, \rho(x, y))$ be metric spaces. A mapping $f : M \to N$ is said to be* quasisymmetric *if it is not constant and if there exists a homeomorphism $\eta : [0, \infty) \to [0, \infty)$ such that for any points $x, y, z \in M$ and $t > 0$ we have that*

$$d(x, y) \leq t\, d(x, z) \quad \textit{implies that} \quad \rho(f(x), f(y)) \leq \eta(t)\, \rho(f(x), f(z)). \tag{B.4.2}$$

We may sometimes say that f is η-quasisymmetric to be more precise. A basic reference for quasisymmetric mappings is [Tuk–Vais].

What does (4.2) mean? If y is much closer to x than z is, then we can take t to be very small, and we conclude that $f(y)$ is much closer to $f(x)$ than $f(z)$ is (since $\eta(t)$ will also be very small). In other words quasisymmetric mappings do not distort the *relative* size of distances very much, even though the distances themselves may be distorted tremendously. This should be compared with the bi-Lipschitz condition (2.7), which requires that the actual distances not be distorted very much. Bi-Lipschitz mappings are automatically quasisymmetric (with a linear choice of η), but the converse is not true. A basic example of a quasisymmetric mapping is given by $f : \mathbb{R}^n \to \mathbb{R}^n$, $f(x) = |x|^a\, x$, $a > -1$. When $a \neq 0$ this mapping distorts distances strongly near the origin and near infinity.

The quasisymmetry condition gives another way to say that the snapshots of a mapping are bounded, but with a different normalization than for bi-Lipschitz mappings. Namely we permit ourselves to rescale both domain and range when we take the snapshot.

It is not difficult to show that f is quasisymmetric if it satisfies (4.2) with $\eta : [0, \infty) \to [0, \infty)$ not necessarily a homeomorphism but just a mapping which satisfies $\eta(t) \to 0$ when $t \to 0$ and which is bounded on bounded intervals. That is, one can always find a homeomorphism which is larger than such a given η. In practice this observation makes it easier to check that a mapping is quasisymmetric.

Roughly speaking, quasisymmetric mappings are mappings which may change the size of balls very much but not their shapes. This idea is illustrated by the following example. Let $(M, d(x, y))$ be a metric space and let $0 < s < 1$ be given. The the identity mapping on M is quasisymmetric as a mapping from $(M, d(x, y))$ to $(M, d(x, y)^s)$ or vice-versa. The spaces $(M, d(x, y))$ and $(M, d(x, y)^s)$ have exactly the same classes of balls, they differ only in the radii of these balls.

Let us say that two metric spaces are *quasisymmetrically equivalent* if there is a quasisymmetric mapping from one onto the other. Note that the inverse of a quasisymmetric mapping from one metric space onto another is also quasisymmetric. This is not difficult to verify.

Another approximate statement is that quasisymmetric mappings understand the concept of expanding a ball by a bounded factor even if they do not understand the size of the radius of the ball, even to within a bounded factor. For instance, if two metric spaces are quasisymmetrically equivalent, then one is doubling if and only if the other one is. (Exercise.) Similarly, given two metric spaces and a quasisymmetric map from one onto the other, any measure on one can be pushed over to the other using the mapping, and this correspondence takes doubling measures to doubling measures.

In the previous section we saw how doubling measures can be used to deform the geometry of a space. We want to take a closer look at this deformation now. It will be convenient to make a mild nondegeneracy assumption about our metric spaces though.

B.4.3. Definition: *A metric space $(M, d(x,y))$ is said to be* uniformly perfect *if there is a constant $C > 0$ so that for each $x \in M$ and $0 < r \leq \operatorname{diam} M$ there is a point $y \in M$ such that*

$$C^{-1} r \leq d(x,y) \leq r. \tag{B.4.4}$$

This condition prevents the occurrence of isolated points or even islands that are too isolated. All of the metric spaces described in Section 2 satisfy this condition.

Isolated points are not necessarily bad. The set of integers $\mathbb{Z}$ equipped with the Euclidean metric is a quite reasonable space even though it consists only of isolated points. Isolated points can cause some nuisances for us though. For simplicity we shall require the uniformly perfect condition in order to avoid technical irritations, even though it is stronger than we really need.

Another way to state the uniformly perfect condition is to require that

$$\operatorname{diam} B(x,r) \geq C^{-1} r \tag{B.4.5}$$

whenever $x \in M$ and $r \leq \operatorname{diam} M$. (This constant need not be quite the same as the one above.)

It is not hard to imagine how the uniform perfectness condition can be useful in the context of doubling measures and quasisymmetric mappings. To use the hypothesis of quasisymmetry, for instance, one needs to have three points x, y, z in hand, and the uniformly perfect condition can help to provide these points.

There is a nice relationship between doubling measures and quasisymmetric geometry on uniformly perfect spaces.

B.4.6. Proposition: *If $(M, d(x,y))$ is a uniformly perfect metric space and μ is a doubling measure on M, and if $D(x,y)$ is defined as in (3.6) (for any choice of $\alpha > 0$), then the identity mapping on M is quasisymmetric as a mapping from $(M, d(x,y))$ to $(M, D(x,y))$.*

Thus when we deform the geometry using a doubling measure, the new space remains quasisymmetrically equivalent to the original one.

The statement of the theorem is slightly illegal, since $D(x,y)$ is only a quasimetric and quasisymmetric mappings were defined officially only for metric spaces, but this is not a serious issue.

The following is the main point.

B.4.7. Lemma: *Let $(M, d(x,y))$ be a uniformly perfect metric space and let μ be a doubling measure on M. Then there is a (small) constant $a \in (0,1)$ such that*

$$\mu(B(x, a\,r)) \leq (1-a)\,\mu(B(x,r)) \tag{B.4.8}$$

whenever $x \in M$ and $0 < r \leq \operatorname{diam} M$.

Compare this with (3.3), which says that measures of balls do not grow too fast as the radius is increased. This says that measures of balls decrease at a definite rate as the radius decreases. Indeed one can iterate (4.8) to get

$$\mu(B(x, a^k\, r)) \leq (1-a)^k\,\mu(B(x,r)) \tag{B.4.9}$$

for all $x \in M$, $0 < r \leq \operatorname{diam} M$, and positive integers k. This inequality implies in particular that μ has no atoms, that sets with only one element always have measure 0. In general a doubling measure can have an atom only at an isolated point. (Note that counting measure on $\mathbb{Z}$ is doubling.)

Let us prove the lemma. Let $x \in M$ and $0 < r \leq \operatorname{diam} M$ be given. The requirement that M be uniformly perfect implies that there is a point $y \in M$ such that $2\,b\,r \leq d(x,y) \leq r/2$ for some constant b that does not depend on x or r. Thus $B(x, b\,r)$ and $B(y, b\,r)$ are disjoint subsets of $B(x,r)$, and so

$$\begin{aligned} \mu(B(x, b\,r)) &\leq \mu(B(x,r)) - \mu(B(y, b\,r)) \\ &\leq \mu(B(x,r)) - C^{-1}\,\mu(B(x,r)) \end{aligned} \tag{B.4.10}$$

for a suitable constant C, because of the doubling property for μ (and the fact that $B(x,r) \subseteq B(y, 2\,r)$). This implies (4.8), and the lemma follows.

Now let us see why Proposition 4.6 is true. It suffices to show that there exists a mapping $\eta : [0, \infty) \to [0, \infty)$ such that $\lim_{t \to 0} \eta(t) = 0$, η is bounded on bounded intervals in $[0, \infty)$, and such that

$$d(x,y) \le t\, d(x,z) \quad \text{implies that} \quad D(x,y) \le \eta(t)\, D(x,z). \tag{B.4.11}$$

for any points $x, y, z \in M$ and any $t > 0$. For $t \ge 1$ one can get this kind of estimate using (3.3), the doubling condition for μ, and the definition of $D(x,y)$. For $t < 1$ one can derive the necessary bound from (4.9). In each case one can choose $\eta(t)$ so that it is a constant multiple of a power of t, but one needs to use different powers for the different cases (a large power for large t, a small power for small t). This proves Proposition 4.6, modulo details which we leave as an exercise.

Let us now continue with the same assumptions as in Proposition 4.6 and take a closer look at the $D(x,y)$ balls. Our goal is to check that $(M, D(x,y), \mu)$ automatically satisfies the Ahlfors regularity condition (3.9) with $s = 1/\alpha$. Fix $x \in M$ and $\rho > 0$ and consider the set

$$\beta(x, \rho) = \{y \in M : D(x,y) \le \rho\}. \tag{B.4.12}$$

(We shall continue to write $B(x,r)$ for the balls with respect to $d(x,y)$.) It may be that $\beta(x,\rho) = M$, and if this is the case we require that ρ be chosen as small as possible with this property. The smallest such ρ will be comparable in size to the $D(\cdot,\cdot)$-diameter of M.

Choose $r > 0$ as small as possible so that $\beta(x,\rho) \subseteq \overline{B}(x,r)$. Thus there is a point $z \in \beta(x,\rho)$ such that $d(x,z) > r/2$. Using (4.11) we get that

$$B(x, C^{-1} r) \subseteq \beta(x,\rho) \subseteq \overline{B}(x,r) \tag{B.4.13}$$

for a suitable constant C which does not depend on x, z, or ρ.

Let us check that

$$C^{-1} \rho \le \mu(B(x,r))^{\alpha} \le C \rho \tag{B.4.14}$$

for a suitable constant $C > 0$, where $\alpha > 0$ was used in the definition of $D(x,y)$, as in the statement of Proposition 4.6. The upper bound follows from the fact that

$$\rho \ge D(x,z) \ge \mu(B(x, d(x,z)))^{\alpha} \ge \mu(B(x, r/2))^{\alpha} \tag{B.4.15}$$

and the doubling condition for μ. To get the lower bound we use the requirement that M be uniformly perfect to obtain a point $w \in M$ such that $r < d(x, w) \leq C\,r$ for some constant C. (Small lie, but wait a moment.) Our choice of r ensures that $D(x, w) \geq \rho$. Going back to the definition (3.6) we obtain the lower bound in (4.14) using also the doubling condition for μ. This argument does not work in the case where $C\,r \geq \operatorname{diam} M$, because then the uniformly perfect condition does not provide us with such a point w. In this case $\mu(B(x, r))$ is approximately the same as $\mu(M)$, because of the doubling condition, and then the lower bound in (4.15) follows because ρ is never larger than the $D(\cdot, \cdot)$-diameter of M.

From (4.14), (4.13), and the doubling condition on μ we obtain that

$$C^{-1}\,\rho \leq \mu(\beta(x, \rho))^{\alpha} \leq C\,\rho. \tag{B.4.16}$$

There is a minor technical point of measurability of $\beta(x, \rho)$, but this is not serious. Thus we see that $(M, D(x, y))$ satisfies the regularity condition (3.9) with dimension $1/\alpha$, modulo the fact that $D(x, y)$ is not necessarily a metric but only a quasimetric, and the fact that $D(x, y)$ may not be complete. The first issue can be avoided using (2.2). As to the second, one can check that $(M, D(x, y))$ is complete if and only if $(M, d(x, y))$ is.

Notice, by the way, that if a metric space is doubling and complete, then all of its closed and bounded subsets are compact. This follows from the well-known result that a metric space is compact if and only if it is complete and totally bounded. (The latter means that for each $\epsilon > 0$ the space can be covered by finitely many balls of radius ϵ.)

B.5. Rest and recapitulation.

The doubling condition for metric spaces provides a way to say that the space has bounded snapshots. The Lipschitz, bi-Lipschitz, and quasisymmetry conditions are boundedness conditions for snapshots of mappings. The doubling condition for a measure says that its snapshots are bounded.

Using doubling measures we can deform the geometry of metric spaces. Under mild technical conditions this deformation is quasisymmetrically equivalent to the original space, so that the class of balls is approximately preserved, just the radii are changed. This is a variant of the idea of conformal deformations in Riemannian geometry.

In Section III. we shall consider the extent to which Lipschitz, bi-Lipschitz, and quasisymmetric mappings have hidden rigidity, or conditions under which spaces have hidden rigidity. These rigidity properties will have a measure-theoretic flavor to them. Something special happens, not uniformly at all points, only for 90% of them, something like that.

These rigidity phenomena do not happen all the time, just under certain circumstances. It is not understood exactly when they should occur or what form they should take.

In the Section II. we look at general facts about spaces and mappings which enjoy bounds on their snapshots, and another concept of bounded snapshots.

II. Analysis on general spaces

B.6. Hölder continuous functions on metric spaces.

B.6.1. Definition: *Let $(M, d(x,y))$ and $(N, \rho(u,v))$ be metric spaces and let $\alpha > 0$ be given. We say that $f : M \to N$ is* Hölder continuous of order α *if there is a constant $K > 0$ so that*

$$\rho(f(x), f(y)) \leq K\, d(x,y)^\alpha \tag{B.6.2}$$

for all $x, y \in M$. In the case where N is the real line with the standard metric we call f a Hölder continuous function of order α and we denote the space of all of these by Λ_α.

We shall sometimes say that "f is Hölder continuous of order α with constant K" to make the bound explicit. If we say "Hölder continuous function" (instead of mapping) we shall mean a mapping into the real numbers.

When $\alpha = 1$ we have already christened this the Lipschitz condition, but the latter is special and deserves a special name. Traditionally one does not consider $\alpha > 1$ because on many spaces Hölder continuous functions of order $\alpha > 1$ are all constant. This happens on Euclidean spaces, for instance, because Hölder continuity of order $\alpha > 1$ implies that the first derivatives of the function exist and vanish everywhere.

Notice that if we replace $(M, d(x,y))$ with the snowflake $(M, d(x,y)^\alpha)$ then (6.2) becomes simply the Lipschitz condition with respect to the new metric. So all theorems about Lipschitz functions on metric spaces apply to Hölder continuous functions. The reverse fails in a strong way, as we shall see.

Typically when there are plenty of functions which satisfy a condition like (6.2) with $\alpha > 1$ it means that the space that we are working on is already the result of applying the snowflake functor to a metric space. One can formulate precise versions of this statement which are easy to prove.

Can we build plenty of Hölder continuous functions on a metric space? Can we build plenty of Lipschitz functions for that matter? The simple answer to the second question is yes, and we can build them using the fact that the metric $d(x,y)$ is Lipschitz in each variable. (This uses the triangle inequality and does not work for quasimetrics.) By the snowflake remark above we can build Hölder continuous functions of order $\alpha < 1$ at least as easily as we can build Lipschitz functions, but it turns out that there are more flexible constructions available too. The basic information of this type is contained in the following.

B.6.3. Theorem: *Let $(M, d(x,y))$ be a metric space and let $0 < \alpha < 1$ be given. Let $k > 1$ be some constant, and suppose that we have a family of functions $\{f_j\}_{j\in\mathbb{Z}}$, $f_j : M \to \mathbb{R}$, such that*

$$\sup_M |f_j| \leq k\, 2^{j\alpha}, \tag{B.6.4}$$

$$f_j \text{ is } k\, 2^{j(\alpha-1)}\text{-Lipschitz}, \tag{B.6.5}$$

$$\sum_{j\in\mathbb{Z}} f_j(x_0) \quad \text{converges for some } x_0 \in M. \tag{B.6.6}$$

Then $\sum_{j\in\mathbb{Z}} f_j(x)$ converges uniformly on bounded subsets of M to a function f which is Hölder continuous of order α. Conversely, every Hölder continuous function of order α on M admits such a representation.

What on earth does this mean? Let us check first the following fact.

B.6.7. Lemma: *Suppose that $0 < \alpha < 1$ and that $f : M \to \mathbb{R}$ satisfies*

$$\sup_M |f| \leq k\, t^{\alpha}, \tag{B.6.8}$$

$$f \text{ is } k\, t^{(\alpha-1)}\text{-Lipschitz}, \tag{B.6.9}$$

for some $k, t > 0$. Then f is Hölder continuous of order α, with constant $2\,k$.

The point here is that the constant does not depend on t, only on k. Before we pursue this matter let us quickly check the lemma. Let $x, y \in M$ be given. We need to estimate $|f(x) - f(y)|$. If x and y are close together then we should use the Lipschitz condition (6.9). When they are far apart we should use the bound (6.8). Specifically, if $d(x,y) \leq t$, then (6.9) implies that

$$|f(x) - f(y)| \leq k\, t^{(\alpha-1)}\, d(x,y) \leq k\, d(x,y)^{\alpha}. \tag{B.6.10}$$

If $d(x,y) \geq t$, then we get

$$|f(x) - f(y)| \leq 2 \sup_M |f| \leq 2\, k\, t^{\alpha} \leq 2\, k\, d(x,y)^{\alpha}. \tag{B.6.11}$$

Thus we get the Hölder continuity condition (6.2) in either case. This proves Lemma 6.7.

The precise choices of parameters in (6.4), (6.5), (6.8), and (6.9) may have seemed odd, but hopefully the point of these choices makes better sense now. We are making the right normalizations to get functions which are Hölder continuous of order α with a uniform bound. Think of the t in Lemma 6.7 as defining the scale at which the real action takes place. At larger scales the function is sort of boring because (6.8) implies that it is too small to be interesting. At smaller scales the function is rather smooth, because of (6.9).

For a function on the real line, think of the Lipschitz condition as meaning that the derivative of the function is bounded. Think of Hölder continuity of order α as meaning something like the "derivative of order α" is bounded. If you want to switch from a derivative of order α to a derivative of order 1, you have to pay for it. If the function is smooth at the scale of t, then you should pay by having a norm which looks like $t^{(\alpha-1)}$.

If we were working with functions on the real line, then we would think of functions f as in Lemma 6.7 as being Hölder continuous functions of order α whose Fourier transform is concentrated in the range of frequencies of size t^{-1}. On a general metric space we can still think of a function f as in Lemma 6.7 as being a function whose "frequencies" are concentrated in the range of t^{-1}, but now this is more of an idea than a precise statement.

In Theorem 6.3 the functions f_j should be viewed as functions which live at the scale of 2^j, whose frequencies are mostly of the size of 2^{-j}.

These words reflect a well-developed area of Harmonic Analysis called Littlewood-Paley theory.

Theorem 6.3 says that every Hölder continuous function on M can be decomposed into pieces living at the scale of 2^j, $j \in \mathbb{Z}$, and that such pieces can be combined in an arbitrary manner to get a Hölder continuous function of order α. Let us pause a moment to appreciate this. Lemma 6.7 implies that each of the f_j's in Theorem 6.3 is Hölder continuous of order α with a constant that does not depend on j, but these Hölder constants do not have to be anything more than bounded. We cannot control the Hölder constant of $\sum_j f_j$ by summing the Hölder constants of the individual pieces, because these individual Hölder constants need not be any more than bounded. Thus we are getting a kind of orthogonality which implies that an infinite sum of bounded objects is bounded even when the sum of the norms is infinite. The point is that the various scales are more-or-less independent of each other, so that the f_j's do not really interact much.

The condition (6.6) is just a normalization and should not be taken too seriously.

Let us now proceed to the proof of Theorem 6.3. Assume first that $\{f_j\}$ is given as in the theorem, and we want to prove that $\sum_j f_j$ is Hölder continuous. The following is the main point.

B.6.12. Lemma: *If* $0 < \alpha < 1$ *and* $\{f_j\}$ *satisfies (6.4) and (6.5), then*

$$\sum_{j\in\mathbb{Z}} |f_j(x) - f_j(y)| \leq C(\alpha)\, k\, d(x,y)^\alpha \tag{B.6.13}$$

for all $x, y \in M$.

This constant $C(\alpha)$ depends only on α and it blows up as α approaches 0 or 1.

Let $x, y \in M$ be given, and choose $p \in \mathbb{Z}$ so that $2^p \leq d(x,y) < 2^{p+1}$. (We may as well assume that $x \neq y$.) In order to estimate the sum in (6.13) we consider separately the pieces where $j \leq p$ or $j > p$, and we use (6.4) or (6.5) accordingly. For the first we use (6.4) to obtain that

$$\begin{aligned}\sum_{j\leq p} |f_j(x) - f_j(y)| &\leq \sum_{j\leq p} 2 \sup_M |f_j| \\ &\leq \sum_{j\leq p} k\, 2^{j\alpha} \\ &\leq C(\alpha)\, k\, 2^{p\alpha} \\ &\leq C(\alpha)\, k\, d(x,y)^\alpha,\end{aligned} \tag{B.6.14}$$

because of our choice of p. For the second piece we use (6.5) to get

$$\begin{aligned}\sum_{j>p} |f_j(x) - f_j(y)| &\leq \sum_{j>p} k\, 2^{j(\alpha-1)}\, d(x,y) \\ &\leq C(\alpha)\, k\, 2^{p(\alpha-1)}\, d(x,y) \\ &\leq C(\alpha)\, k\, d(x,y)^\alpha,\end{aligned} \tag{B.6.15}$$

again using our choice of p. Lemma 6.12 follows by combining (6.14) and (6.15).

If $\{f_j\}$ satisfies (6.4), (6.5), and (6.6), then from Lemma 6.12 and (6.6) we can conclude that $\sum_j f_j(x)$ converges for all $x \in M$. To show that the convergence is uniform on compact sets one can make the same kind of analysis as in the proof of Lemma 6.12, but simply taking $p = 0$. The point is that the series converge uniformly because they are dominated by geometric series.

This proves the first half of Theorem 6.3. To prove the second half, to represent any Hölder continuous function in this manner, it will be more convenient to establish first another characterization of Hölder continuous functions.

B.6.16. Theorem: *Let $(M, d(x, y))$ be a metric space, let $0 < \alpha < 1$ be given, and let $g : M \to \mathbb{R}$ be a real-valued function on M. Suppose that there exists $k > 0$ so that for each $t > 0$ we can find $g_t : M \to \mathbb{R}$ such that*

$$\sup_M |g - g_t| \leq k\, t^\alpha \qquad \text{and} \tag{B.6.17}$$

$$g_t \text{ is } k\, t^{\alpha-1}\text{-Lipschitz.} \tag{B.6.18}$$

Then g is Hölder continuous of order α. Conversely, if g is Hölder continuous of order α, then for each $t > 0$ there is a function $g_t : M \to \mathbb{R}$ which satisfies (6.17) and (6.18), with k taken to be the Hölder constant of g.

Again the powers of t in (6.17) and (6.18) may look a little strange. It would be more standard to try to approximate g to within an error of t and then compute what is the right Lipschitz norm. For me it is easier to make the above normalizations because I can remember them better. Roughly speaking, g_t is the approximation to g that is smooth at the scale of t. This idea makes sense no matter what α is, and then we compute the bounds above in terms of α.

Let us prove Theorem 6.16. Suppose that g is given and that k and g_t exist as above. Let $x, y \in M$ be given, and let us estimate $|g(x) - g(y)|$. Set $t = d(x, y)$. We have that

$$\begin{aligned} |g(x) - g(y)| &\leq |g(x) - g_t(x)| + |g_t(x) - g_t(y)| \\ &\quad + |g_t(y) - g(y)| \\ &\leq k\, t^\alpha + k\, t^{\alpha-1}\, d(x, y) + k\, t^\alpha \\ &\leq 3\, k\, d(x, y)^\alpha. \end{aligned} \tag{B.6.19}$$

Thus g is Hölder continuous of order α with constant $3\, k$.

In order to prove the second half of Theorem 6.16 we need a preliminary fact.

B.6.20. Lemma: *Suppose that $(M, d(x, y))$ is a metric space and that $\mathcal{H}$ is a collection of L-Lipschitz real-valued functions on M for some number L. If $\inf\{h(x) : h \in \mathcal{H}\}$ is finite for some $x \in M$, then it is finite for all $x \in M$, and the resulting function on M is L-Lipschitz.*

This is a standard fact. One writes the Lipschitz condition on $h \in \mathcal{H}$ as

$$h(y) - L\, d(x, y) \leq h(x) \leq h(y) + L\, d(x, y) \tag{B.6.21}$$

for all $x, y \in M$, and one concludes that

$$\inf_{h \in \mathcal{H}} h(y) - L\, d(x, y) \leq \inf_{h \in \mathcal{H}} h(x) \leq \inf_{h \in \mathcal{H}} h(y) + L\, d(x, y). \tag{B.6.22}$$

Lemma 6.20 follows easily from this.

So let us prove now the second half of Theorem 6.16. Let $g : M \to \mathbb{R}$ be given and Hölder continuous of order α. We may as well assume that g is Hölder continuous with constant ≤ 1, since we can always achieve this normalization by dividing g by a constant. Let $t > 0$ be given, and set

$$g_t(x) = \inf\{g(y) + t^{\alpha-1}\, d(x,y) : y \in M\}. \tag{B.6.23}$$

This is a standard device for approximating functions on a metric space by Lipschitz functions. We view $g(y) + t^{\alpha-1}\, d(x,y)$ here as a function of x, with y treated as a parameter. As a function of x is it $t^{\alpha-1}$-Lipschitz, because $d(x,y)$ is 1-Lipschitz as a function of x, as one can check from the triangle inequality. Lemma 6.20 implies that g_t is also $t^{\alpha-1}$-Lipschitz, at least if we can show that it is finite.

Let us show that we can replace (6.23) by a more localized formula, namely

$$g_t(x) = \inf\{g(y) + t^{\alpha-1}\, d(x,y) : y \in M, d(x,y) \leq t\}. \tag{B.6.24}$$

Indeed, if $d(x,y) > t$, then

$$g(y) + t^{\alpha-1}\, d(x,y) > g(y) + d(x,y)^\alpha, \tag{B.6.25}$$

since $\alpha < 1$. Thus we get that $g(y) + t^{\alpha-1}\, d(x,y) > g(x)$ in this case, since we are assuming that g is Hölder continuous with constant ≤ 1. This means that this y cannot contribute to the infimum in (6.23), because it gives a larger answer than $y = x$ does. This proves (6.24).

From (6.24) and the Hölder continuity of g it follows easily that g_t is finite, and so $t^{\alpha-1}$-Lipschitz, by Lemma 6.20.

It remains to check that (6.17) holds. Notice first that

$$g_t(x) \leq g(x) \tag{B.6.26}$$

for all $x \in M$, by the definition (6.23) of g_t (just take $y = x$). To prove (6.17) it suffices to show that

$$g(x) \leq g_t(x) + t^\alpha \tag{B.6.27}$$

for all $x \in M$. This follows from (6.24): given $y \in M$ with $d(x,y) \leq t$, we can use the Hölder condition on g to get

$$g(x) \leq g(y) + d(x,y)^\alpha \leq g(y) + t^\alpha \leq g(y) + t^{\alpha-1}\, d(x,y) + t^\alpha. \tag{B.6.28}$$

Taking the infimum over y we get (6.27). Combining (6.26) with (6.27) we have (6.17), as desired.

This completes the proof of Theorem 6.16. Let us go back now and finish the proof of Theorem 6.3. Let f be a real-valued function on M which is Hölder continuous of order α. Let g_t be the approximation to f at the scale of t provided by the second half of Theorem 6.16 (with $g = f$). Set $\phi_j = g_{2^{j+1}} - g_{2^j}$. We would like to use these functions for f_j, but that does not quite work. They satisfy (6.4) and (6.5) (with k taken to be a multiple of the Hölder constant for f), because of (6.17) and (6.18), but there is a small problem with (6.6). So instead we set

$$f_j(x) = \phi_j(x) - \phi_j(x_0), \tag{B.6.29}$$

where $x_0 \in M$ is some arbitrarily chosen basepoint. Actually we do this for all but one j. For one choice j_0 of j we set

$$f_j(x) = \phi_j(x) - \phi_j(x_0) + f(x_0). \tag{A.6.29$'$}$$

We choose this j_0 so large so that $|f(x_0)| \leq 2^{j_0\,\alpha}$. This choice ensures that f_j satisfies (6.4) and (6.5) for all j, with k taken to be a multiple of the Hölder constant for f. (Keep in mind that adding a constant to a function does not affect its Lipschitz constant.)

So with this choice of $\{f_j\}$ we have (6.4) and (6.5), and we also have (6.6) because $f_j(x_0) = 0$ for all j except $j = j_0$, for which we have $f_{j_0}(x_0) = f(x_0)$. It remains to show that

$$\sum_{j=-\infty}^{\infty} f_j(x) = f(x) \tag{B.6.30}$$

for all $x \in M$. Let m and n be integers, with $m \leq j_0 \leq n$. We have that

$$\sum_{j=m}^{n} f_j(x) = g_{2^{n+1}}(x) - g_{2^m}(x) - \{g_{2^{n+1}}(x_0) - g_{2^m}(x_0)\} + f(x_0). \tag{B.6.31}$$

The definition of g_t ensures that

$$\lim_{m \to -\infty} g_{2^m}(x) - g_{2^m}(x_0) = f(x) - f(x_0). \tag{B.6.32}$$

Specifically, we are using (6.17) here (with $g = f$). Using (6.18) we get that

$$|g_{2^{n+1}}(x) - g_{2^{n+1}}(x_0)| \leq k\, 2^{(n+1)\,(\alpha-1)}\, d(x, x_0). \tag{B.6.33}$$

Since $\alpha < 1$ we conclude that this tends to 0 as $n \to \infty$. Combining this with (6.31) and (6.32) we obtain (6.30).

This completes the proof of Theorem 6.3.

B.7. Metric spaces which are doubling.

In the preceding section we saw how one can decompose a Hölder continuous function on a metric space into layers for which the most interesting behavior was concentrated at a given scale. At a fixed scale, however, there was not much to say. In this section we shall assume that the metric space is doubling and try to say more about what happens at fixed scales. (The discussion of hyperbolic geometry in Section 1 is relevant here.)

The first main point will be to construct a suitable partition of unity. To do that we find coverings of M by balls of a fixed size with controlled overlapping.

Let us assume from now on in this section that $(M, d(x,y))$ is a metric space which is doubling. Fix a $t > 0$.

B.7.1. Lemma: *There is a subset $A = A(t)$ of M such that*

$$d(x,y) \geq t \qquad \text{when } x, y \in A, x \neq y, \tag{B.7.2}$$

and such that $M \subseteq \bigcup_{x \in A} B(x,t)$.

We want to A to be a maximal subset of M which satisfies (7.2). In order to produce this maximal set we work in bounded subsets initially where any such set must be finite.

Specifically, suppose that B is a ball in M, and that E is a subset of B which satisfies the analogue of (7.2). Then E can have at most a bounded number of elements, where the bound depends on t and the radius of B but not on E. Indeed, the doubling condition implies that B can be covered by a bounded number of balls of radius $t/2$, and any such ball can contain at most one element of E. Therefore E can contain at most a bounded number of elements.

Fix a sequence of balls $\{B_j\}$ in M with $B_j \subseteq B_{j+1}$ for all j and $\bigcup_j B_j = M$, and let us try to produce a sequence of subsets A_j of the B_j's with the property that $A_j \supseteq A_{j-1}$ and A_j is a maximal subset of B_j which satisfies (7.2). We build these subsets recursively. For $j = 1$ we take any maximal subset of B_1 which satisfies (7.2). Such a maximal subset exists, since we know that any subset of B_1 which satisfies (7.2) can have only a bounded number of elements. If A_j has been constructed, then we choose A_{j+1} to be a maximal subset of B_{j+1} which contains A_j and which satisfies (7.2). Again this is possible because of our bound on the number of elements of such a subset. This A_{j+1} will then be maximal among subsets of B_{j+1} which satisfy (7.2). (Maximal in the sense that it cannot be made larger, not that it has the largest possible number of elements.)

Now we can take $A = \bigcup_j A_j$, and it is easy to see that A is a maximal subset of M which satisfies (7.2). This proves Lemma 7.1.

B.7.3. Lemma: *If $A \subseteq M$ satisfies (7.2) and $k > 0$ is given, then there is a constant $C(k)$ so that no point in M lies in $B(x, k\,t)$ for more than $C(k)$ points $x \in A$. Here $C(k)$ depends also on the doubling constant for M but not on A or t.*

Indeed, let $p \in M$ be given, and let F denote the set of $x \in A$ such that $p \in B(x, k\,t)$. Then $F \subseteq B(p, k\,t)$ and F satisfies (7.2). The doubling condition implies that $B(p, k\,t)$ can be covered by $\leq C(k)$ balls of radius $t/2$, and each of these balls cannot contain more than one element of F. Thus we conclude that F has at most $C(k)$ elements, and the lemma follows.

B.7.4. Lemma (Partition of unity): *Let A be as in Lemma 7.1. We can find a family of real-valued functions $\{\phi_a\}_{a \in A}$ on M with the following properties: $0 \leq \phi_a \leq 1$ for all a; each ϕ_a is $C\,t^{-1}$-Lipschitz, where the constant C depends on the doubling constant for M but nothing else; $\phi_a(x) = 0$ when $d(a, x) \geq 2\,t$; and*

$$\sum_{a \in A} \phi_a(x) = 1 \tag{B.7.5}$$

for all $x \in M$.

To prove this let us define first some auxiliary bump functions ψ_a on M. Let $f : [0, \infty) \to [0, 1]$ denote the function which satisfies $f \equiv 1$ on $[0, t]$, $f \equiv 0$ on $[2\,t, \infty)$, and f is linear on $[t, 2\,t]$. Thus f is t^{-1}-Lipschitz. Define $\psi_a : M \to [0, 1]$ by $\psi_a(x) = f(d(a, x))$. Then ϕ_a is t^{-1}-Lipschitz, $\psi_a(x) = 0$ when $d(a, x) \geq 2\,t$, and $\psi_a(x) = 1$ when $d(x, a) \leq t$.

Set

$$\phi_a(x) = \frac{\psi_a(x)}{S(x)}, \tag{B.7.6}$$

where $S(x) = \sum_{b \in A} \psi_b(x)$. Lemma 7.3 ensures that $\psi_b(x)$ is nonzero for only a bounded number of b's for any given x, and so this sum is always finite. We have that $S(x) \geq 1$ for all $x \in M$, because $\bigcup_{p \in A} B(p, t) \supseteq M$. It is not hard to check that $S(x)$ is $C\,t^{-1}$-Lipschitz, using the fact that the balls $B(p, 2\,t)$, $p \in A$, have bounded overlap (by Lemma 7.3). That is, given $x \in M$, $\psi_b(x) \neq 0$ for only a bounded number of $b \in A$, and we can use the fact that each ψ_b is t^{-1}-Lipschitz to get a similar bound for $S(x)$. Once we have this bound for $S(x)$ we can easily conclude that each ϕ_a is $C\,t^{-1}$-Lipschitz for a suitable constant C. That $0 \leq \phi_a \leq 1$ follows from the same property of ψ_a and the fact that $S \geq 1$ everywhere. $\phi_a(x) = 0$ when $d(a, x) \geq 2\,t$ also follows from the corresponding property of ψ_a. The identity (7.5) is an immediate consequence of (7.6).

This completes the proof of Lemma 7.4.

The partition of unity in Lemma 7.4 permits us to define a linear operator for approximating Hölder continuous functions by Lipschitz functions. (Compare with the nonlinear procedure in (6.23).) Let A, $\{\phi_a\}$ be as in Lemma 7.4, and define an operator E_t acting on functions on M by

$$E_t(f)(x) = \sum_{a \in A} f(a)\phi_a(x). \tag{B.7.7}$$

B.7.8. Lemma: *There is a constant C, depending only on the doubling constant of M, so that*

$$\sup_M |f - E_t(f)| \leq C\,k\,t^\alpha \tag{B.7.9}$$

whenever $f : M \longrightarrow \mathbb{R}$ is Hölder continuous of order α and with constant k.

Indeed,

$$f(x) - E_t(f)(x) = \sum_{a \in A} (f(x) - f(a))\phi_a(x), \tag{B.7.10}$$

because of (7.5) and (7.7). If $\phi_a(x) \neq 0$, then $x \in B(a, 2t)$, and this can happen for only a bounded number of a's, by Lemma 7.3. For each of these a's we have that $|f(x) - f(a)| \leq k\,d(x, a)^\alpha \leq k\,(2t)^\alpha$. From this we get that $|f(x) - f(a)| \leq C\,k\,t^\alpha$, and (7.9) follows since $x \in M$ is arbitrary. This proves Lemma 7.8.

B.7.11. Lemma: *There is a constant C, depending only on the doubling constant of M, so that $E_t(f)$ is $C\,k\,t^{\alpha-1}$-Lipschitz when $f : M \longrightarrow \mathbb{R}$ is Hölder continuous of order α and with constant k.*

Let $x, y \in M$ be given, and let us estimate $|E_t(f)(x) - E_t(f)(y)|$. Suppose first that $d(x, y) \geq t$. In this case we have that

$$\begin{aligned} |E_t(f)(x) - E_t(f)(y)| &\leq |E_t(f)(x) - f(x)| + |f(x) - f(y)| \\ &\quad + |f(y) - E_t(f(y)| \\ &\leq C\,k\,t^\alpha + k\,d(x, y)^\alpha + C\,k\,t^\alpha \qquad \text{(B.7.12)} \\ &\leq C\,k\,t^{\alpha-1}\,d(x, y), \end{aligned}$$

since $d(x,y) \geq t$. This is the estimate that we want. If $d(x,y) \leq t$, then we compute as follows:

$$\begin{aligned} E_t(f)(x) - E_t(f)(y) &= \sum_{a \in A} f(a)\,(\phi_a(x) - \phi_a(y)) \\ &= \sum_{a \in A} (f(a) - f(x))\,(\phi_a(x) - \phi_a(y)) \end{aligned} \tag{B.7.13}$$

For this last equality we have used the observation that $\sum_{a\in A}(\phi_a(x) - \phi_a(y)) = 0$, because of (7.5). If $\phi_a(x) \neq 0$, then $x \in B(a, 2\,t)$, and this can happen for only a bounded number of a's. If $\phi_a(y) \neq 0$, then $y \in B(a, 2\,t)$, and this can happen for only a bounded number of a's as well. In this case $x \in B(a, 3\,t)$, since we are assuming that $d(x,y) \leq t$. In either event we have that $x \in B(a, 3\,t)$ and so $|f(a) - f(x)| \leq k\,(3t)^\alpha$. We also have that $|\phi_a(x) - \phi_a(y)| \leq t^{-1}\,d(x,y)$. Altogether we get then that

$$|E_t(f)(x) - E_t(f)(y)| \leq C\,k\,t^\alpha\,t^{-1}\,d(x,y), \tag{B.7.14}$$

which is the estimate that we wanted. This completes the proof of Lemma 7.11.

Lemmas 7.8 and 7.11 provide a new proof of the second half of Theorem 6.16. This new proof has the disadvantage that it requires the additional assumption that M be doubling, but it has the advantage of providing a construction for the Lipschitz approximations that one can understand with one's eyes. This leads also to another proof of the second half of Theorem 6.3 in the case of doubling spaces in which we can control the form of the functions f_j much better.

This type of analysis has a consequence for the first half of Theorem 6.3 also. Let us write $A(2^j)$ for $A = A(t)$ with $t = 2^j$, and let us write $\phi_{j,a}$, $a \in A(2^j)$, for the functions provided by Lemma 7.4, with $t = 2^j$. Fix some basepoint $x_0 \in M$, and set

$$\theta_{j,a}(x) = \phi_{j,a}(x) - \phi_{j,a}(x_0). \tag{B.7.15}$$

B.7.16. Lemma: *Notations and assumptions as above. Let $\lambda_{j,a}$, $j \in \mathbb{Z}$, $a \in A(2^j)$, be a uniformly bounded collection of real numbers, and let $0 < \alpha < 1$ be given. Then*

$$\sum_{j\in\mathbb{Z}} \sum_{a\in A(2^j)} \lambda_{j,a}\,2^{j\alpha}\,\theta_{j,a}(x) \tag{B.7.17}$$

converges for every $x \in M$, and the limit defines a Hölder continuous function on M of order α.

The point here is that we get a Hölder continuous function for any choice of $\{\lambda_{j,a}\}$ whatsoever, so long as it is bounded. We shall discuss this further in a moment.

The double sum in (7.17) actually converges absolutely, but one can interpret it as a sum first in a, in which all but finitely many terms vanish, and then a sum in j.

To prove the lemma we set $f_j(x) = \sum_{a \in A(2^j)} \lambda_{j,a}\, 2^{j\alpha}\, \theta_{j,a}(x)$, and we try to apply Theorem 6.3. We have that $\theta_{j,a}(x_0) = 0$ for all j and a, and so $f_j(x_0) = 0$ for all j, and so (6.6) follows. Let us check (6.4). If $\theta_{j,a}(x) \neq 0$, then either $x \in B(a, 2\,t)$ or $x_0 \in B(a, 2\,t)$, and in either case this can happen for only a bounded number of a's. Since $\phi_{j,a}$ takes values in $[0, 1]$ we have that $\theta_{j,a}$ takes values in $[-1, 1]$, and we conclude that each $\lambda_{j,a}\, 2^{j\alpha}\, \theta_{j,a}(x)$ is bounded by a constant times $2^{j\alpha}$. Since the number of nonzero terms is bounded we get (6.4).

Now let us check (6.5). Let $x, y \in M$ be given, and let us estimate $|f_j(x) - f_j(y)|$. Notice that $\theta_{j,a}(x) - \theta_{j,a}(y) = \phi_{j,a}(x) - \phi_{j,a}(y)$, and that this quantity is nonzero for only a bounded number of a's again. When it is nonzero it is bounded by $C\, 2^{-j}\, d(x, y)$, by Lemma 7.4. Thus we conclude that $|f_j(x) - f_j(y)|$ is bounded by $C\, 2^{j\alpha}\, 2^{-j}\, d(x, y)$, and this implies the Lipschitz condition (6.5).

Theorem 6.3 implies now that $\sum_j f_j(x)$ converges for every $x \in M$ and defines a Hölder continuous function on M of order α. This proves Lemma 7.16.

Note that the precise form of the $\phi_{a,j}$'s does not matter for Lemma 7.16, just the basic properties of their size, support, and Lipschitz norms. We do not need the identity (7.5).

What does Lemma 7.16 mean? Do not take its precise form too seriously, the point is more the general type of construction that it permits us to make. It tells us that we can build Hölder continuous functions with approximately prescribed behavior at each individual scale and location. Think of $\theta_{j,a}$ as being some kind of bump function that lives near $a \in A(2^j)$ and which lives at the scale of 2^j. It is not quite fair to say that $\theta_{j,a}$ lives near a; although $\phi_{j,a}$ is supported in $B(a, 2\,t)$, this need not be true of $\theta_{j,a}$, because of its definition (7.15). However, the only disturbance to this statement is the additive constant $-\phi_{j,a}(x_0)$, which plays only a minor technical role. This bump function is not normalized to have bounded Hölder constant for the exponent α, but $2^{j\alpha}\, \theta_{j,a}$ is properly normalized. Thus Lemma 7.16 permits us to combine these normalized bumps in an arbitrary manner so long as we give them bounded coefficients. This is somewhat remarkable, even if it is not very difficult to prove. It says that the various scales and locations are approximately independent of each other for Hölder continuous functions. In the previous section we saw that the different scales were independent of each other, and now we are using the doubling condition to make both scales and locations independent of each other.

B.8. Spaces of homogeneous type.

A *space of homogeneous type* is a metric space $(M, d(x, y))$ equipped with a doubling measure μ. Actually, let us require only that $d(x, y)$ be a quasimetric, but to extend Definition 3.1 (doubling measures) to this case one has to be a little careful because the balls for quasimetrics need not be measurable. It is easy enough to avoid this technicality in formulating the notion of doubling measure ((3.3) is helpful in this regard), but instead we simply restrict ourselves to quasimetrics for which open balls are Borel sets. Quasimetrics which are powers of metrics, as in (2.2), always enjoy this property.

Actually, we should also ask that μ be Borel regular, so that one has the usual compatibility between topology and measure. It suffices to ask that M be complete, because then closed and bounded subsets of M are compact, open subsets are σ-compact, and we can apply Theorem 2.18 in [Rudin].

Spaces of homogeneous type are very useful in analysis, because they provide a general setting where certain types of arguments work. The basic references are [Coif–Weis]$_{AH}$, [Coif–Weis]$_{HS}$, see also [Stein]$_{HA}$. ([Coif–Weis]$_{HS}$ has a very nice discussion of examples.)

Let us discuss examples for a moment. We can easily get some by remembering the discussion of Ahlfors regular spaces in Section 3 and realizing that they are all examples of spaces of homogeneous type. To appreciate these examples one should think a little about their setting in analysis. Euclidean spaces equipped with Lebesgue measure are spaces of homogeneous type, and this corresponds to "standard" analysis. $\mathbb{R}^{n+1}$ equipped with the parabolic metric (2.9) and Lebesgue measure is also a space of homogeneous type, and it corresponds to analysis of the heat equation (instead of the Laplacian). The Heisenberg group H_n with the metric discussed in Section 2 and Lebesgue measure is another space of homogeneous type, and the natural analysis on it arises in several complex variables. Our symbolic Cantor sets F^∞ from Section 2 can also be made into spaces of homogeneous type, and analysis on them can be viewed in terms of martingales in probability theory.

So what kind of analysis can we do on a space of homogeneous type $(M, d(x, y), \mu)$? Here is the simplest example. Suppose that $f \in L^1_{loc}(M, \mu)$.

Then μ-almost every element of M is a Lebesgue point of f, i.e.,

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, d\mu(y) = 0 \tag{B.8.1}$$

for almost every x. The main point of the proof is that standard arguments for proving this result on Euclidean spaces (involving covering lemmas) extend to this setting. (See Section 34.)

This example is very nice because it illustrates a basic point: a lot of classical analysis on Euclidean spaces does not make sense simply for measure spaces, or simply for metric spaces, but in the presence of both structures one can at least formulate certain concepts, like the concept of a Lebesgue point. The notion of spaces of homogeneous type imposes a mild compatibility between measure and metric which is adequate to prove a lot of theorems. In particular it is sufficient to get a lot of mileage out of covering lemmas.

What other kinds of analysis work on spaces of homogeneous type? We shall give some precise examples in the next three sections, but for the moment let us take a brief look at singular integral operators. I do not want to get into a serious discussion of singular integral operators, but to illustrate the concept consider the following. Let Δ denote the Laplace operator on $\mathbb{R}^n$, which is given by

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}. \tag{B.8.2}$$

If $g(x)$ is a smooth function with compact support on $\mathbb{R}^n$, then all of the second partial derivatives of g can be recovered from Δg through singular integral operators. Indeed, as is well known, we have that

$$g(x) = c_n \int_{\mathbb{R}^n} \frac{1}{|x-y|^{n-2}} \, \Delta g(y) \, dy \tag{B.8.3}$$

when $n \neq 2$, where c_n is some constant, and there is an analogous formula using $\log|x-y|$ when $n = 2$. If $n \geq 2$, $1 \leq i, j \leq n$, and $i \neq j$ (for instance), then

$$\frac{\partial^2}{\partial x_i \partial x_j} g(x) = c_n' \text{ P.V.} \int_{\mathbb{R}^n} \frac{(x_i - y_i)\,(x_j - y_j)}{|x-y|^{n+2}} \, \Delta g(y) \, dy, \tag{B.8.4}$$

where the integral on the right side should be interpreted as a Cauchy principal value.

This is a basic example of a singular integral operator on $\mathbb{R}^n$, viewed here as an operator acting on Δg and giving $\frac{\partial^2}{\partial x_i \partial x_j} g$. It is not hard to show that such an operator is bounded on $L^2(\mathbb{R}^n)$, but it is in fact bounded on L^p when $1 < p < \infty$, and not when $p = 1, \infty$. (L^2 is always special because of the Hilbert space structure, the Fourier transform, etc. To go to L^p spaces in this context one has to come to terms with some geometry.) These operators are subtle because they cannot be analyzed successfully just in terms of the size of their kernels, one has to take into account the cancellations as well. See [Stein]$_{SI}$.

There are analogues of these operators on $\mathbb{R}^{n+1}$ associated to the heat operator instead of the Laplacian, for which the parabolic metric defines the right geometry. There are also analogous operators on the Heisenberg group using the natural "sub-Laplacian" there. On our symbolic Cantor set F^∞ there does not seem to be a nice version of $\frac{\partial}{\partial x}$, but there are still operators of this general type, martingale transforms.

These specific operators do not have an abstract form on spaces of homogeneous type, but there are natural classes of singular integral operators which include these examples and which can be defined on spaces of homogeneous type. This was one of the main motivations behind the concept, the study of singular integral operators in a general setting.

Note however that it is not so easy to define specific interesting operators on a general space of homogeneous type. In practice the interesting operators come from the particular space, and the general theory simply applies to the whole class of operators without distinguishing certain ones.

B.9. Hölder continuity and mean oscillation.

Let $(M, d(x,y), \mu)$ be a space of homogeneous type, and $g \in L^1_{loc}(M, \mu)$ be given. Set

$$\mathrm{mo}(x,r) = \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |g(y) - \mathrm{m}_{x,r}\, g| \, d\mu(y), \tag{B.9.1}$$

where

$$\mathrm{m}_{x,r}\, g = \frac{1}{\mu(B(x,r))} \int_{B(x,r)} g(z) \, d\mu(z). \tag{B.9.2}$$

$\mathrm{mo}(x,r)$ is the *mean oscillation* of g over the ball $B(x,r)$, the average amount by which g deviates from its average on that ball. One can think of it as a way to measure a snapshot of g. In this section and the next two we want to look at how the mean oscillation controls the behavior of g. (For notational simplicity we do not make the dependence of $\mathrm{mo}(x,r)$ on g explicit, but it should always be clear from the context.)

B.9.3. Proposition: *Notation and assumptions as above. Assume that $\alpha > 0$ and that there is a constant $k > 0$ so that*

$$\mathrm{mo}(x, r) \leq k\, r^{\alpha} \tag{B.9.4}$$

for all $x \in M$ and $r > 0$. Then g can be modified on a set of μ-measure zero to get a Hölder continuous function of order α.

Note that the converse is true, and easy to derive from the definitions.

Let us prove the proposition. As mentioned in the previous section almost every point in M is a Lebesgue point for g. Let us show that

$$|g(x) - g(y)| \leq C\, k\, d(x, y)^{\alpha} \tag{B.9.5}$$

when x and y are Lebesgue points of g, with a constant C that does not depend on x or y.

Notice that

$$|\mathrm{m}_{z,t}\, g - \mathrm{m}_{z,2t}\, g| \leq C \mathrm{mo}(z, 2t) \tag{B.9.6}$$

for all $z \in M$ and $t > 0$. This is not hard to check, using the doubling condition for μ. From (9.4) we get that

$$|\mathrm{m}_{z,t}\, g - \mathrm{m}_{z,2t}\, g| \leq C\, k\, t^{\alpha}. \tag{B.9.7}$$

Since x is a Lebesgue point for g we have that

$$\lim_{t \to 0} \mathrm{m}_{x,t}\, g = g(x), \tag{B.9.8}$$

and similarly for y. This is all that we really need to know about x and y. If $r > 0$, then we obtain that

$$\mathrm{m}_{x,r}\, g - g(x) = \sum_{j=0}^{\infty} (\mathrm{m}_{x,2^{-j}r}\, g - \mathrm{m}_{x,2^{-j-1}r}\, g). \tag{B.9.9}$$

Hence

$$|g(x) - \mathrm{m}_{x,r}\, g| \leq \sum_{j=0}^{\infty} C\, k\, (2^{-j}\, r)^{\alpha} \leq C\, k\, r^{\alpha}. \tag{B.9.10}$$

Notice that this last constant C depends on α (and blows up as $\alpha \to 0$).

Set $r = d(x, y)$. Since $d(x, y)$ is a quasimetric we have $B(x, r) \cup B(y, r) \subseteq B(x, L\, r)$ for some constant L which does not depend on x or y. (If $d(x, y)$ is a metric then we can take $L = 2$.) We have that

$$|\mathrm{m}_{x,r}\, g - \mathrm{m}_{y,r}\, g| \leq |\mathrm{m}_{x,r}\, g - \mathrm{m}_{x,L\,r}\, g| + |\mathrm{m}_{x,L\,r}\, g - \mathrm{m}_{y,r}\, g| \leq C\, \mathrm{mo}(x, L\, r). \tag{B.9.11}$$

This can be derived from the definitions, using the doubling condition, as in (9.6). We conclude that

$$|\mathrm{m}_{x,r}\, g - \mathrm{m}_{y,r}\, g| \le C\, k\, r^{\alpha}. \tag{B.9.12}$$

Altogether we obtain that

$$|g(x) - g(y)| \le |g(x) - \mathrm{m}_{x,r}\, g| + |\mathrm{m}_{x,r}\, g - \mathrm{m}_{y,r}\, g| + |\mathrm{m}_{y,r}\, g - g(y)| \le C\, k\, r^{\alpha}, \tag{B.9.13}$$

because of (9.10), (9.12), and the analogue of (9.10) for y instead of x. This proves (9.5), and Proposition 9.3 follows.

In short Proposition 9.3 implies that good bounds on the mean oscillation of g imply good bounds on the maximal oscillation of g. In geometry one can sometimes control the mean oscillation using isoperimetric inequalities.

B.10. Vanishing mean oscillation.

Let $(M, d(x, y), \mu)$ be a space of homogeneous type. Let V denote the space of locally integrable functions g on M such that

$$\lim_{r \to 0} \sup_{x \in M} \mathrm{mo}(x, r) = 0, \tag{B.10.1}$$

where $\mathrm{mo}(x, r)$ is as in (9.1). This is a trivial variation on the concept of *vanishing mean oscillation*, but the precise choice of definition is slightly different from the norm, and so we use the nonstandard notation V to denote the space of these functions.

The idea is that we measure an approximate continuity using mean oscillation instead of maximal oscillation. If we replaced the average in (9.1) with a supremum, then (10.1) would reduce to the requirement that g be uniformly continuous.

Are there nontrivial examples of functions in V? On the real line, with the usual metric and Lebesgue measure, the functions $\big|\log|x|\big|^s$ lie in V for any $0 < s < 1$. If one wants bounded examples, one can take the cosine of these functions. Thus Proposition 9.3 does not work in this context. One can see this in the proof, in the importance of α being positive in the summing of the geometric series in (9.10).

Thus functions in V do not have to be continuous, even after correcting their values on a set of measure zero. Still they do enjoy some residual continuity properties, some aspects of which we shall describe now. Note that functions which lie in the critical Sobolev spaces, the Sobolev spaces that just miss being contained in the space of continuous functions, are contained in V. (See also the next section.)

B.10.2. Lemma: *Suppose that $(M, d(x,y), \mu)$ is a space of homogeneous type, and let g be a function in V. Then for each $t > 0$ we can find a continuous function g_t on M such that*

$$\lim_{t\to 0} \sup\{|g_t(x) - g_t(y)| : x, y \in M, d(x,y) \le t\} = 0, \quad \text{(B.10.3)}$$

and

$$\lim_{t\to 0} \sup_{x\in M} \frac{1}{\mu(B(x,t))} \int_{B(x,t)} |g(y) - g_t(y)| \, d\mu(y) = 0. \quad \text{(B.10.4)}$$

The first condition says that g_t is almost constant at the scale of t when t is small, and the second condition says that g_t approximates g very well on average at the scale of t.

Before we prove this let us record a corresponding "uniqueness" result.

B.10.5. Lemma: *Notations and assumptions as in Lemma 10.3. If h_t is another family of functions on M that satisfy (10.3) and (10.4) with g_t replaced with h_t, then*

$$\lim_{t\to 0} \sup_M |g_t - h_t| = 0. \quad \text{(B.10.6)}$$

Thus a function in V has an almost canonical family of continuous approximations.

Lemma 10.5 is easy to verify and we leave the details to the reader. The main point is that $g_t(x)$ and $h_t(x)$ are both very well approximated by the average of g over $B(x,t)$ when t is small, because of (10.3) and (10.4).

Now let us prove Lemma 10.2. We first reduce to the case where $d(x,y)$ is a metric instead of a quasimetric. We can do this because of (2.2). It is a little easier to do this in two steps, to first reduce to the case where $d(x,y)$ is a power of a metric, and then to get rid of the power. The point is that if we change $d(x,y)$ by a power, or we change it to something of the same size, then the requirements on g_t are not really changed, it is just a question of relabelling the t's. Actually there is also a small matter of using the doubling property for μ so as not to disturb (10.4) too much when we replace $d(x,y)$ by something which is merely of the same size.

Thus we may assume that $d(x,y)$ is actually a metric.

Fix $t > 0$, and let $A \subseteq M$ and $\{\phi_a\}_{a\in A}$ be as in Lemmas 7.1 and 7.4. (Remember Lemma 3.4, which ensures that $(M, d(x,y))$ is doubling as a metric space.) Define g_t by

$$g_t(x) = \sum_{a\in A} (\mathrm{m}_{a,t} g) \, \phi_a(x), \quad \text{(B.10.7)}$$

where $\mathrm{m}_{a,t}g$ is defined as in (9.2). This is just the obvious way to try to build a continuous function using partitions of unity whose value at a point x is approximately the average of g over $B(x,t)$.

Let us check (10.3) and (10.4). Let $x, y \in M$ be given, with $d(x,y) \le t$. We have that

$$g_t(x) - g_t(y) = \sum_{a \in A} \mathrm{m}_{a,t}g\,(\phi_a(x) - \phi_a(y)). \tag{B.10.8}$$

Using (7.5) we can convert this into

$$g_t(x) - g_t(y) = \sum_{a \in A} (\mathrm{m}_{a,t}g - \mathrm{m}_{x,t}g)\,(\phi_a(x) - \phi_a(y)). \tag{B.10.9}$$

If $\phi_a(x) \neq 0$, then $x \in B(a, 2\,t)$, because of Lemma 7.4. Similarly $y \in B(a, 2\,t)$ if $\phi_a(y) \neq 0$, and in either case we get that $d(a,x) \le 3\,t$. These are the only a's which are relevant for (10.8), and there is only a bounded number of them, because of Lemma 7.3. Therefore

$$|g_t(x) - g_t(y)| \le C\, \sup\{|\mathrm{m}_{a,t}g - \mathrm{m}_{x,t}g| : a \in A, d(a,x) \le 3\,t\}. \tag{B.10.10}$$

On the other hand

$$|\mathrm{m}_{a,t}g - \mathrm{m}_{x,t}g| \le C\,\mathrm{mo}(x, 4\,t) \tag{B.10.11}$$

when $d(a,x) \le 3\,t$. This can be computed from the definitions, as in (9.11). Combining this with (10.10) we can obtain (10.3) from the assumption that g lies in V.

This leaves (10.4), which is similar. Notice first that

$$g(y) - g_t(y) = \sum_{a \in A} (g(y) - \mathrm{m}_{a,t}g)\,\phi_a(y), \tag{B.10.12}$$

because of (10.7) and (7.5). Thus

$$\begin{aligned} &\tfrac{1}{\mu(B(x,t))} \textstyle\int_{B(x,t)} |g(y) - g_t(y)|\,d\mu(y) \\ &\qquad \le \textstyle\sum_{a \in A} \tfrac{1}{\mu(B(x,t))} \int_{B(x,t)} |g(y) - \mathrm{m}_{a,t}g|\,\phi_a(y)\,d\mu(y). \end{aligned} \tag{B.10.13}$$

In order for a term in the sum on the right to be nonzero we must have that ϕ_a does not vanish identically on $B(x,t)$, which means that there must be a point $y \in B(x,t)$ such that $y \in B(a, 2\,t)$. This means that $x \in B(a, 3\,t)$, and this can happen only for a bounded number of $a \in A$, because of Lemma 7.3. For each of these a's we have that

$$\frac{1}{\mu(B(x,t))} \int_{B(x,t)} |g(y) - \mathrm{m}_{a,t}g|\,d\mu(y) \le C\,\mathrm{mo}(x, 4\,t). \tag{B.10.14}$$

This last inequality is another variant of (9.11), and it can be derived from the definition (9.1) of $\mathrm{mo}(x, 4t)$ and the doubling condition on μ. Once we have (10.14) we can derive (10.4) from (10.13) and the assumption that $g \in V$.

This completes the proof of Lemma 10.2.

The preceding argument also works when g takes values in some $\mathbb{R}^n$, or even in a Banach space. This observation is potentially useful because in some cases one wants g really to take values in some special set, like a manifold, or a metric space. From (10.3) and (10.4) we have that the approximations g_t take values very close to the same set, at least when t is small enough. If the range of g is a nice set, like a finite polyhedron or a manifold, then we can probably deform the g_t's slightly so as to take values in the same set. Under suitable conditions we can also get these approximations to be homotopic to each other, at least for small t. This would use also the uniqueness result Lemma 10.5.

The point of all of this is that one can normally extract from mappings with vanishing mean oscillation practically the same kind of topological information as for continuous mappings. I first encountered this idea in [Doug], specifically Theorem 7.36 on p.189 of [Doug]. This result gave the index of a Toeplitz operator as the winding number of its functional symbol, but where the latter was not continuous so that the concept of the winding number required interpretation. Vanishing mean oscillation was not mentioned in [Doug] in part because the definition had not yet been introduced by Sarason [Sar]. Actually in this setting of complex analysis and operator theory a different description of the function spaces arose naturally, and the connection with the real-variable notion of vanishing mean oscillation was made through the celebrated duality theorem of Fefferman (as discussed in [Sar]). For the purpose of extracting topological information however it is only the real-variable aspects that matter. See [Brez–Nir] for more information about topological properties of mappings with vanishing mean oscillation, especially degree theory.

In trying to understand the computations of this section one should think back to the examples of $|\log |x||^s$ and its cosine as functions on the real line which lie in V. These functions are not continuous, or even bounded in the first case. Lemma 10.2 provides a certain family of approximations to these functions. As t moves these approximations move slowly, because of the uniqueness result Lemma 10.5 (which implies that g_t is very close to g_{2t} when t is small, for instance). As t goes down to 0 these approximations can spin around an infinite amount, that is what happens in these unbounded and discontinuous examples, but the spinning is slow, and the approximations are almost canonical.

B.11. Bounded mean oscillation.

Let $(M, d(x,y), \mu)$ be a space of homogeneous type. Let BMO ("bounded mean oscillation") denote the space of locally integrable functions g on M which satisfy the condition

$$\|g\|_* = \sup\{\mathrm{mo}(x,r) : x \in M, r > 0\} < \infty, \tag{B.11.1}$$

where $\mathrm{mo}(x,r)$ is as in (9.1).

This definition may seem strange, but it shows up naturally in many problems, as we shall see later.

Every bounded function on M satisfies this condition. The function $\log|x|$ on $\mathbb{R}$ also lies in BMO. This is the most basic example of an unbounded BMO function.

Roughly speaking, BMO is to bounded functions as the space V in the previous section is to uniformly continuous functions.

Note that (11.1) is not affected if we add a constant to g. This is a quietly crucial point. If we drop "oscillation" from BMO, and require instead that

$$\sup_{x,r} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |g(y)|\, d\mu(y) < \infty, \tag{B.11.2}$$

then we would simply recover L^∞. By permitting the averages to be unbounded while keeping bounded the mean oscillation we get new functions like the logarithm.

Notice also that (11.1) scales in the same way as the L^∞ norm. If we are working on $\mathbb{R}^n$, for instance, and $a > 0$ is given, then $\|g(x)\|_* = \|g(ax)\|_*$, just as they have also the same L^∞ norm. In practice, if one has a situation where the scaling is the same as for L^∞, but constants can be added without affecting anything, then there is a good chance that BMO is natural.

Here is an example. Let g be a function on $\mathbb{R}^n$ such that $\nabla g \in L^n$. What can we say about g? We can assume that ∇g is taken in the sense of distributions, or we can assume that g is actually smooth and look for a priori estimates. Observe that

$$\int_{\mathbb{R}^n} |\nabla g(x)|^n\, dx \tag{B.11.3}$$

scales in the same manner as the L^∞ norm, i.e., (11.3) does not change if we replace $g(x)$ with $g(ax)$. Also (11.3) does not change if we add a constant to g. The aforementioned principle suggests that BMO should be relevant for this situation. When $n = 1$ this is not really true; we can use the fundamental theorem of calculus to say that $\nabla g \in L^1(\mathbb{R})$ implies that g is bounded, and that the difference between the supremum and infimum of g is controlled by (11.3). When $n > 1$ this is no longer true, and this is the well-known failure of the Sobolev embedding at the critical exponent. We do have in this case that

$$\|g\|_* \leq C \left(\int_{\mathbb{R}^n} |\nabla g(x)|^n \, dx \right)^{\frac{1}{n}}. \tag{B.11.4}$$

This is a standard consequence of the Poincaré inequality, and one can also show that g lies in the space V of the preceding section when $\nabla g \in L^n(\mathbb{R})$ in the sense of distributions. There are more refined versions of these results, but one cannot get away from the fact that the right hand side can be finite for unbounded functions.

For the sake of comparison, recall the Sobolev inequalities

$$\left(\int_{\mathbb{R}^n} |g(x)|^q \, dx \right)^{\frac{1}{q}} \leq C(p,n) \left(\int_{\mathbb{R}^n} |\nabla g(x)|^p \, dx \right)^{\frac{1}{p}}, \tag{B.11.5}$$

where $1/q = 1/p - 1/n$ and $1 \leq p < n$. For this we need to add an assumption about g, that it has compact support for instance, but the bound does not depend on this assumption in a quantitative way. As $p \to n$ we have $q \to \infty$, but when $n > 1$ the estimate degenerates unless we switch to BMO.

It is often the case in analysis that functions are not bounded when one would like them to be, but that they do lie in BMO. Another example of this phenomenon is provided by the fact that standard singular integral operators as mentioned in Section 8 are not bounded on L^∞ but do map L^∞ into BMO, and even BMO into itself under suitable conditions. (See [Garn], [Jour], [Stein]$_{HA}$.)

How close are BMO functions to being bounded? A famous theorem of John and Nirenberg (see [Garn], [Jour], [Stein]$_{HA}$) says that a BMO function on $\mathbb{R}^n$ lies in L^p_{loc} for all $p < \infty$. In fact one can define BMO in terms of an L^p norm in the mean oscillation $\mathrm{mo}(x,r)$ from (9.1) and get an equivalent condition. BMO functions are even exponentially integrable, and one can define BMO in terms of exponential integrals. The logarithm provides the "worst" degree of unboundedness that BMO functions can have.

These facts work not only on Euclidean spaces but on any space of homogeneous type. This provides another example of how spaces of homogeneous type have the right amount of geometry for doing certain kinds of analysis. This is a point worthy of reflection: one can talk about L^p spaces on measure spaces, one can talk about continuous functions and Lipschitz functions and so forth on metric spaces, but to define BMO we need some of each.

Nice theorem of Strömberg (see [Garn], [Jour], [Stein]$_{HA}$): g lies in BMO means that for each ball B in M there is some real number so that at 90% of the points in B $g(x)$ differs from that number by a bounded amount. (The "bounded amount" is not permitted to depend on B, but the "number" is.) This captures well a basic idea of BMO, and of harmonic analysis in general, that it is often natural to say that something behaves well only at most places, and not everywhere, and that if such a thing is true uniformly at all scales and locations, then one can often derive information that appears at first to be much stronger than what one started with.

General references for BMO include [Garn], [Jour], [Stein]$_{HA}$. We shall return to the John-Nirenberg theorem in Section 40. Another kind of bound on the oscillations of BMO functions is given in Lemma 42.1.

III. Rigidity and structure

B.12. Differentiability almost everywhere.

B.12.1. Theorem: *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, then f is differentiable at (Lebesgue) almost all points in $\mathbb{R}^n$.*

This is a fantastic theorem. A proof is described in Section 35 below. See also [Fed]$_{GMT}$, [Stein]$_{SI}$.

Let us think about Lipschitz mappings in terms of snapshots. Given $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$, and $r > 0$, define $f_{x,r} : B(0,1) \to \mathbb{R}$ by

$$f_{x,r}(h) = \frac{f(x + r\,h) - f(x)}{r}. \tag{B.12.2}$$

The Lipschitz condition says exactly that the family of functions $f_{x,r}$, $x \in \mathbb{R}^n$, $r > 0$, is uniformly bounded. Theorem 12.1 says that for almost all $x \in \mathbb{R}^n$ we have that $f_{x,r}$ converges uniformly on $B(0,1)$ as $r \to 0$ to a linear function. Thus we can progress from boundedness to the existence of a limit, at almost all points anyway.

This is fantastic. It is a rigidity statement that is not incorporated explicitly into the definition (unlike smoothness). The potential limit is not known in advance, unlike the notion of Lebesgue point. Let us think about these ideas in the setting of general metric spaces. The concept of Lipschitz mappings makes sense on any metric space. This method (12.2) of blowing up a function at a point does not make sense on an arbitrary metric space, because there is no way to compare one ball with another. For many examples of metric spaces we can do this, for the examples described in Section 2 for instance. So what happens if we look for similar rigidity properties on other metric spaces?

Let us start with a Euclidean snowflake $(\mathbb{R}^n, |x-y|^\alpha)$, $0 < \alpha < 1$. A Lipschitz function on this space is the same as a function on $\mathbb{R}^n$ which is Hölder continuous of order α with respect to the standard metric. For a Hölder continuous function of order α the natural way to blow up at a point x is to take

$$g_{x,r}(h) = \frac{f(x+r\,h) - f(x)}{r^\alpha}. \tag{B.12.3}$$

This family of functions on $B(0,1)$ is bounded if and only if the original function f is Hölder continuous of order α. However there is no good limiting behavior for typical functions when $0 < \alpha < 1$. One can understand this in terms of Theorem 6.3 and Lemma 7.16. These results say that we can practically choose the behavior of Hölder continuous functions at different scales and locations independently of each other, which is incompatible with the existence of a limit.

This idea can be made very precise using Littlewood-Paley theory and wavelets. We shall not go into this here, but let us look at the classical example of lacunary trigonometric series. We work on the real line now, and we take $f : \mathbb{R} \to \mathbb{R}$ to be of the form

$$f(y) = \sum_{j=1}^{\infty} a_j\, 2^{-j\,\alpha}\, \exp(2^j\, i\, y). \tag{B.12.4}$$

Here $\alpha > 0$. If the a_j's are bounded, then this series converges uniformly to a continuous function on $\mathbb{R}$. If $\alpha < 1$ and the a_j's are bounded, then f is Hölder continuous of order α, as one can check using Theorem 6.3. In fact the boundedness of the a_j's is necessary for f to be Hölder continuous of order α.

What happens if we want f to be Lipschitz? This corresponds to $\alpha = 1$, but Theorem 6.3 does not apply to that case. In fact it cannot:

$$\text{If } f \text{ is Lipschitz, then } \sum_j |a_j|^2 < \infty. \tag{B.12.5}$$

In particular we have to have $a_j \to 0$ as $j \to \infty$ in order for f to be Lipschitz, but square-summability is much stronger than that. It is easy to see why (12.5) has to be true. In order for f to be Lipschitz we have to have that $f'(y)$, taken in the sense of distributions, should be bounded.

Setting $\alpha = 1$ in (12.4) we have that $f'(y)$ is represented by the Fourier series

$$f'(y) = \sum_{j=1}^{\infty} i\, a_j \, \exp(2^j\, i\, y). \tag{B.12.6}$$

In order for this to be bounded it should lie in L^2 locally, and this periodic function lies in L^2 if and only if $\sum_j |a_j|^2 < \infty$, by standard results about Fourier series.

It is not true that $\sum_j |a_j|^2 < \infty$ implies that f is Lipschitz when $\alpha = 1$, but this is close to being true, in the sense that $\sum_j |a_j|^2 < \infty$ implies that $f'(y)$ exists almost everywhere and defines a function in L^p_{loc} for all $p < \infty$, and in fact that $f' \in BMO(\mathbb{R})$.

A nice fact: if f is as in (12.4) with $\alpha = 1$, and if f is differentiable at a single point, then $a_j \to 0$ as $j \to \infty$. See the first corollary on p.106 of [Katzn].

The conclusion of this story is that nothing like differentiability almost everywhere holds for Hölder continuous functions of order < 1, and hence there is nothing like differentiability almost everywhere for Lipschitz functions on the Euclidean snowflakes $(\mathbb{R}^n, |x-y|^\alpha)$, $0 < \alpha < 1$. Although Euclidean structure is pleasant for discussing lacunary Fourier series, approximately the same behavior can be found in Theorem 6.3 and Lemma 7.16. We should therefore not expect any rigidity like Theorem 12.1 on any metric space which can be realized as a snowflake.

What about other metric spaces? Is this kind of rigidity peculiar to spaces whose geometry is somehow approximately Euclidean ? The lovely answer is no. There is a version of Theorem 12.1 for the Heisenberg groups. One has to be a little careful in defining the concept of differentiability almost everywhere for functions on the Heisenberg group, but it is basically the same as for Euclidean space. One uses translations and dilations to take snapshots of a given function, all snapshots living on the unit ball, and then one asks that the limit exist as $r \to 0$. The concept of linear mapping has to be modified. One looks for group homomorphisms instead. For real-valued mappings one looks for homomorphisms into the real numbers, but Pansu $[\text{Pan}]_{CC}$ established differentiability almost everywhere for Lipschitz mappings between Heisenberg groups too.

What about other examples? I do not know of any spaces where there is something like differentiability almost everywhere but where the geometry is not practically Euclidean or a relative of the Heisenberg group (or other Carnot groups, as in $[\text{Pan}]_{CC}$). For a space like the Sierpinski carpet described in Section 2 one can hope for better behavior than Cantor sets and snowflakes (but not such good behavior as for Euclidean spaces and Heisenberg groups), but I do not know a nice theorem. (See also the "GWALA" in [Dav–Sem]$_{UR}$, and the discussion in [Sem]$_{Map}$.)

B.13. Pause for reflection.

The preceding discussion of differentiability almost everywhere provides us with a nice opportunity to view everything in Part II with deep suspicion. We see now that there are remarkable rigidity results to be found. There is nothing like that in Part II, and indeed nothing like Theorem 12.1 works at that level of generality.

In the next sections we shall consider other aspects of analysis and geometry more sensitive than those in Part II. Most of it will be about spaces with at least a little Euclidean structure. Measure theory will be crucial, as it is in Theorem 12.1. Good behavior typically does not occur in a uniform way for all points, only at most points. The idea of BMO, with uniform good behavior for 90% percent of the points, uniformly at all scales and locations, will play a large role.

Heisenberg geometry deserves more attention in this way, e.g., for the structure of mappings, recognizing when a space has approximately Heisenberg behavior, etc. Analysis on the Heisenberg group and its similarities to Euclidean analysis has been much studied (see [Fol–Ste]$_{Est}$, [Fol–Ste]$_{HS}$, [Jer], [Stein]$_{HA}$), but less attention has been paid to geometry (but see [Gro]$_{CC}$).

There ought to be more to say about geometry and analysis on certain fractal sets than is available currently. Sierpinski carpets ought to behave better than snowflakes and Cantor sets if not as well as Euclidean spaces and Heisenberg groups. They have lots of rectifiable curves, for instance, although not so many as Euclidean space and the Heisenberg group. Sierpinski gaskets and fractal trees have fewer still but also enough to connect any pair of points. "Guy's birdhouse" [David]$_{MG}$ is another example to consider. See [Barlow] and [Bar–Bass] for an interesting perspective on analysis on fractal sets.

Although Cantor sets have the least amount of structure in the examples that we have discussed, they too enjoy some interesting rigidity properties, as we shall see in Section 22.

B.14. Almost flat curves.

Suppose that we have a curve in the plane whose length is only slightly larger than the distance between its endpoints. What can we say about this curve?

It is not hard to see that the curve must remain close to the line segment between the endpoints. We can hope that its tangents must also be close to parallel to this line segment. This need not be true at all points, but it does have to be true at most points. One can always have a short arc in the curve that winds around as it pleases.

Let us be more precise and make computations. For reasons of simplicity and a larger agenda let us restrict ourselves to curves in the plane, and let us think of the plane as being $\mathbb{C}$ instead of $\mathbb{R}^2$, for notation in particular. Let $z : [a, b] \to \mathbb{C}$ be an arclength parameterization of our curve. This means that $z(t)$ is 1-Lipschitz, $|z'(t)| = 1$ almost everywhere, and the length of the curve is $b - a$. We have that

$$z(b) - z(a) = \int_a^b z'(t)\, dt, \tag{B.14.1}$$

so that $|z(b) - z(a)| \leq b - a$ automatically, and we are interested in the case where

$$b - a \leq (1 + \epsilon)\, |z(b) - z(a)| \tag{B.14.2}$$

for a small $\epsilon > 0$.

Set $\zeta = \frac{z(b)-z(a)}{b-a}$, so that (14.2) is equivalent to

$$(1 + \epsilon)^{-1} \leq |\zeta| \leq 1. \tag{B.14.3}$$

We would like to say that most of the tangents to our curve are almost parallel to ζ, which is practically the same as saying that $z'(t)$ is close to ζ for most t. Indeed this is true, and we can compute in the same manner as [Coif–Mey]$_{IC}$:

$$\int_a^b |z'(t) - \zeta|^2\, dt = \int_a^b \{|z'(t)|^2 - z'(t)\, \overline{\zeta} - \overline{z}'(t)\, \zeta + |\zeta|^2\}\, dt. \tag{B.14.4}$$

We can use (14.2) and the fact that $|z'(t)| = 1$ almost everywhere to convert this into

$$\frac{1}{b-a} \int_a^b |z'(t) - \zeta|^2\, dt = 1 - |\zeta|^2 - |\zeta|^2 + |\zeta|^2 = 1 - |\zeta|^2. \tag{B.14.5}$$

Thus our flatness condition (14.2), (14.3) is equivalent to a bound on the L^2 mean oscillation of $z'(t)$ over $[a, b]$.

When thinking about (14.5) keep in mind that $|z'(t)| = 1$ almost everywhere. The left side of (14.5) is small if and only if $z'(t) - \zeta$ is small except for a set of t's of small measure. Nothing too bad can happen on sets of small measure because $|z'(t)|$ and $|\zeta|$ are both bounded by 1.

In conclusion the flatness condition (14.2) holds with a small ϵ if and only if $z'(t)$ is almost constant on almost all of the interval $[a, b]$. We cannot get rid of the qualification "almost all of the interval", because $z'(t)$ could be practically anything on a small set without affecting (14.2) in a serious way.

Suppose now that we consider curves that are flat at all scales. So that the length of any subarc of the curve is only slightly larger than the distance between its endpoints. Can we say more then?

To be more precise, let us restrict ourselves to unbounded Jordan curves in the plane, and work again with parameterizations by arclength. Let $z : \mathbb{R} \to \mathbb{C}$ be an arclength parameterization of our curve, so that $z(t)$ is 1-Lipschitz and $|z'(t)| = 1$ almost everywhere. We call this curve ϵ-flat if

$$|s - t| \leq (1 + \epsilon)\, |z(s) - z(t)| \qquad \text{for all } s, t \in \mathbb{R}, \tag{B.14.6}$$

where ϵ is some small positive number. This is equivalent to requiring that $z : \mathbb{R} \to \mathbb{C}$ be a $(1 + \epsilon)$-bi-Lipschitz embedding.

With this stronger assumption of flatness at all scales and locations, can we conclude that $z'(t)$ is a small perturbation of a constant everywhere? The answer is no, because there are very flat logarithmic spirals which satisfy this condition. Instead of checking this directly let us make a more general analysis. The flatness condition (14.6) is equivalent to the requirement that

$$\sup_{a,b\in\mathbb{R}} \frac{1}{b-a} \int_a^b |z'(t) - \zeta(a, b)|^2 \, dt \leq 1 - (1 + \epsilon)^{-2}, \tag{B.14.7}$$

where $\zeta(a, b) = \frac{z(b)-z(a)}{b-a}$ is the average of $z'(t)$ over $[a, b]$. The left side of (14.7) is equivalent in size to the square of the BMO norm of $z'(t)$ on $\mathbb{R}$. That is, the BMO norm was defined in (11.1), and it is easy to check that the left side of (14.7) is at least as big as the square of the BMO norm, because of the Cauchy-Schwarz inequality (i.e., the L^2 norm is at least as big as the L^1 norm). Conversely, the left side of (14.7) is bounded by a constant times the square of the BMO norm, because of the John-Nirenberg theorem mentioned in Section 11.

In short our curve is ϵ-flat with a small ϵ if and only if $z'(t)$ has small BMO norm.

The requirement that $z'(t)$ have small BMO norm turns out to be equivalent to the requirement that $z'(t)$ can be written in the form $z'(t) = \exp(ib(t))$, where $b(t)$ is a real-valued function on $\mathbb{R}$ with small BMO norm. The sufficiency of this representation is not hard to see: bound on the mean oscillation of b implies a similar bound for $\exp(ib(t))$, simply because $\exp(ix)$ is a 1-Lipschitz function on $\mathbb{R}$. The converse is more subtle. In particular one has to be careful in taking the logarithm of $z'(t)$, because it can spin around and it will not be continuous in general. See [Coif–Mey]$_{IC}$ for details.

The bottom line here is quite simple. Given a real-valued function $b(t)$ on $\mathbb{R}$ with small BMO norm, we set

$$z(t) = z(0) + \int_0^t \exp(ib(s))\, ds \tag{B.14.8}$$

for some choice of $z(0)$ (it doesn't matter which), and we get an ϵ-flat curve where ϵ is controlled in terms of the BMO norm of b. Conversely all ϵ-flat curves arise this way, with a function $b(t)$ whose BMO norm is controlled in terms of ϵ. A basic example is given by $b(t) = \delta \log |t|$, where δ is small, which corresponds to a flat logarithmic spiral.

Let us proceed now from small distortion to bounded distortion. We say that a curve is a chord-arc curve if the length of the arcs is always bounded by a constant multiple of the lengths of the chords, where now the constant can be large. More precisely, suppose that we are given a curve in the plane which is described by an arclength parameterization $z : \mathbb{R} \to \mathbb{C}$, and let us ask that there exist a constant $k > 1$ such that

$$|s - t| \leq k\,|z(s) - z(t)| \qquad \text{for all } s, t \in \mathbb{R}. \tag{B.14.9}$$

This means that $z : \mathbb{R} \to \mathbb{C}$ is k-bi-Lipschitz . What can we say now?

We can try again to represent $z'(t)$ as $\exp(ib(t))$ in a good way. We can always do this brutally, just taking the principal branch of the logarithm, but this does not respect the geometry very well. For instance we want a representation in which $b(t)$ would be continuous if $z'(t)$ were continuous.

Such a careful choice of logarithm is possible and was carried out in [David]$_{CL}$. Unfortunately there is not a simple characterization of the functions $b(t)$ that arise this way. They lie in BMO, but this necessary condition is not sufficient. Let us call Ω the set of real-valued BMO functions that arise in this manner, through David's logarithm. What can we say about Ω? Not a whole lot, but the basic true fact is that Ω is an open subset of BMO. (The openness comes from the exponential integrability estimates for BMO functions provided by the John-Nirenberg theorem.)

This is a gorgeous little fact. It turns out to be incredibly natural. It gives a topology to the space of chord-arc curves, and the work of Coifman and Meyer shows that certain operators associated to these curves depend continuously on the curves with respect to this topology, even real-analytically. This works in a very sharp way, the BMO topology for curves corresponding exactly to the norm topology for operators. They also have a theorem on the real-analyticity of the correspondence between chord-arc curves and their Riemann mappings. See [Coif–Mey]$_{VR}$, [Coif–Mey]$_{IC}$, [Coif–Mey]$_{LC}$, [Coif–Dav–Mey], and the survey paper [Sem]$_{CIQM}$ for more information. See [Sem]$_{CTCI}$, [Sem]$_{QM}$ for an alternate approach.

It is not known whether the space Ω of chord-arc curves is connected in the BMO topology.

B.15. Mappings that almost preserve distances.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is $(1+\epsilon)$-bi-Lipschitz for some small $\epsilon > 0$, so that f distorts distances by at most a factor of $1+\epsilon$. (Remember (2.7).) What can we say about f?

If $\epsilon = 0$ then f must be a rigid mapping, a combination of a translation and an orthogonal mapping.

When ϵ is small, in what sense does f have to be close to a rigid mapping? To what extent must its differential be almost constant?

Let us consider some examples.

Suppose first that f is of the form $f(x) = x + \phi(x)$, where $\phi : \mathbb{R}^n \to \mathbb{R}^n$ has small Lipschitz norm. Then f is bi-Lipschitz with constant close to 1. These are the simplest mappings of this type.

Now suppose that f is of the form $f(x) = R_{\log|x|}x$ when $x \neq 0$, $f(0) = 0$, where R_t, $-\infty < t < \infty$, is a one-parameter family of rotations on $\mathbb{R}^n$ which is Lipschitz in t with small constant (with respect to any fixed norm on matrices). Then f is bi-Lipschitz with a small constant that depends only on the Lipschitz constant of R_t as a function of t. This is not so hard to check. This shows that a $(1+\epsilon)$-bi-Lipschitz mapping on $\mathbb{R}^n$ can make an infinite amount of spiralling no matter how small ϵ is.

Thus a $(1+\epsilon)$-bi-Lipschitz mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ need not have its differential be almost constant in a uniform way. Slightly sad but true, like the story of the preceding section. So is the differential almost constant in an average sense? This is true, a theorem of John [John]. He did not say it this way, but what he proved was that the differential has small BMO norm. In fact BMO and the John-Nirenberg theorem (discussed in Section 11) began for this problem.

There is a very nice connection here with linear analysis. If one linearizes this problem – think of writing f as $f(x) = x + \delta\, g(x)$, looking at what happens when δ is small, and neglecting terms of order δ^2 or higher – then one faces the following question. Suppose that $g : \mathbb{R}^n \to \mathbb{R}^n$ has the property that the symmetric part of its differential is bounded. (The differential is a matrix-valued function, so that "symmetric part" makes sense.) What can we say about g? About its differential as a whole? It turns out that the antisymmetric part of the differential can be obtained from the symmetric part through singular integral operators, in a manner analogous to the formula (8.4) for obtaining mixed second partial derivatives of a function from the Laplacian. As usual these singular integral operators do not map bounded functions to bounded functions, but map bounded functions into BMO.

If it were true that the differential of a $(1+\epsilon)$-bi-Lipschitz mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ had to be a small perturbation of a constant in the L^∞ norm, then we would be able to conclude that f is a small perturbation of a rigid transformation in the Lipschitz norm. Now that we know that the differential of f is a small perturbation of a constant at most points inside a given ball, what can we say about f itself? In fact it is true that on a given ball B there is a rigid mapping ϕ such that $f - \phi$ has small Lipschitz norm when $f - \phi$ is restricted to a subset of B which contains 90% of the points (in terms of Lebesgue measure). This can be derived from John's theorem and standard facts in analysis (based on maximal functions, as discussed in Sections 35 and 36 below).

Let us now sketch a proof of the fact that if $f : \mathbb{R}^n \to \mathbb{R}^n$ is $(1+\epsilon)$-bi-Lipschitz for some small $\epsilon > 0$, then the differential has small BMO norm. We shall not worry about getting explicit estimates (but see [John]), just the fact that the BMO norm is small. Of course the differential is bounded, so we have to show that the BMO norm is *small* to get any information.

The first observation to make is that if B is a ball in $\mathbb{R}^n$ of radius r, then we can find a rigid mapping $\phi : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$r^{-1} \sup_{y \in B} |f(y) - \phi(y)| \tag{B.15.1}$$

is small, i.e., it is bounded by a function of ϵ which goes to 0 as $\epsilon \to 0$ and which does not depend on f or B. One can give direct proofs of this, but a very simple way to do it is to use compactness. If this is not true, then one can find a sequence of $(1+\epsilon_j)$-bi-Lipschitz mappings $f_j : \mathbb{R}^n \to \mathbb{R}^n$ with $\epsilon_j \to 0$ and a sequence of balls B_j such that the f_j's are not well approximated by rigid mappings on the B_j's, meaning that the quantities analogous to (15.1) are bounded away from 0. One can normalize with translations and dilations to reduce to the case where all the B_j's are the unit ball and $f_j(0) = 0$ for all j. One can then use the Arzela-Ascoli theorem to pass to a subsequence to conclude that the f_j's converge uniformly on compact sets to a mapping $g : \mathbb{R}^n \to \mathbb{R}^n$. This limiting mapping must be

1-bi-Lipschitz and hence rigid. This contradicts the original assumption that the f_j's are uniformly far from rigid mappings on the balls B_j, and so we get our approximation result.

This compactness argument is somewhat revolting for being nonconstructive and not providing explicit bounds, but it is quick and easy.

To show that the differential has small BMO norm we have to show that it is almost constant on most of any ball. This is a stronger statement than (15.1), because it is a condition on the derivative of f. Let a ball B of radius r be given. Let us assume that

$$r^{-1} \sup_{y \in 2B} |f(y) - y| \tag{B.15.2}$$

is small. We can do this because of (15.1), and because we may normalize f by composing it with a rigid mapping without changing the problem. Note that we have replaced B by $2B$ here, to give ourselves some extra room as a precaution.

Fix a point $x \in B$ and another point z with $|x - z| = r$. Consider the line segment S that connects x to z. The restriction of f to S defines a rectifiable curve of length $\leq (1 + \epsilon) r$, since f is $(1 + \epsilon)$-bi-Lipschitz . On the other hand the image of S under f remains very close to S, and in particular

$$r^{-1} \max(|f(x) - x|, |f(z) - z|) \tag{B.15.3}$$

must be small. Thus the distance between the endpoints of this curve is close to r. Let v denote the unit vector $\frac{x-z}{|x-z|}$. If $D_v f(p)$ denotes the directional derivative of f in the direction v at the point p, then it is not hard to show that

$$r^{-1} \int_S |D_v f(p) - v|^2 \, dp \tag{B.15.4}$$

must be small. For instance one can make a calculation like the one in (14.4), (14.5), with the minor additional complication that the restriction of f to S does not provide an arclength parameterization of the image (but almost does, since f is $(1+\epsilon)$-bi-Lipschitz). One can also make geometrical arguments to understand this. At any rate it is true.

Since this is true for all initial choices of x and z one can average over the various segments to show that

$$\frac{1}{|B|} \int_B |df(y) - I|^2 \, dy \tag{B.15.5}$$

is small, where I denotes the identity matrix. This is exactly what we wanted, to show that the mean oscillation of the differential of f was small.

Actually, this is not quite the definition of mean oscillation that we used in (9.1), but it works just as well. We are using here an L^2 integral instead of an L^1 integral, which is all the better (by Cauchy-Schwarz). We have not said that I is the mean value of df on B, but that doesn't matter; as soon as one has a bound for (15.5) one gets a similar bound for the difference between I and the mean value of df on B.

This completes our sketch of the proof that df has small BMO norm.

B.16. Almost flat hypersurfaces.

Let M be a hypersurface in $\mathbb{R}^{d+1}$. That is, M should be a d-dimensional embedded smooth submanifold, and we assume also that it is nice at infinity, asymptotic to a hyperplane for instance. This smoothness should be treated as an *a priori* assumption; we require it for technical convenience, but we want to have estimates which do not depend on this smoothness in a quantitative manner.

Let $\epsilon > 0$ be given. We shall say that M is *ϵ-flat* if the following two conditions are satisfied. First we ask that for each pair of points $x, y \in M$ we have that

$$\text{there is a curve } \gamma \text{ in } M \text{ from } x \text{ to } y \text{ whose length is } \leq (1+\epsilon)\,|x-y|. \tag{B.16.1}$$

This means that the external Euclidean distance and the internal geodesic distance should be almost the same. The second condition is that for each $x \in M$ and $r > 0$ we have that

$$(1-\epsilon)\,\nu_d\, r^d \leq |B(x,r)\cap M| \leq (1+\epsilon)\,\nu_d\, r^d \tag{B.16.2}$$

where $|E|$ denotes the d-dimensional surface measure of $E \subseteq M$ and ν_d denotes the volume of the unit ball in $\mathbb{R}^d$. This means that the volume of a ball in M is practically the same as the volume of a ball of the same radius in flat space. The balls here are defined using the ambient Euclidean metric, but that does not really matter, because of (16.1).

In short M is ϵ-flat if measurements of length and volume are the same to within factors of ϵ of the corresponding measurements for flat space. Note that this concept reduces to the one in (14.6) when $d = 1$. Notice also that this condition is scale-invariant; if we take an ϵ-flat hypersurface and act on it by a translation, dilation, or a rotation, then we get another ϵ-flat hypersurface, with the same choice of ϵ.

So are ϵ-flat hypersurfaces really flat in some reasonable sense? One can show that they are always well-approximated by hyperplanes locally (in the Hausdorff metric), but this is not at all the whole story. The basic question is the following.

B.16.3. Question: *If M is an ϵ-flat hypersurface in $\mathbb{R}^{d+1}$, and if ϵ is small enough, does there exist a $(1+\delta)$-bi-Lipschitz mapping f from $\mathbb{R}^d$ onto M, where δ depends only on ϵ and the dimension, and $\delta \to 0$ as $\epsilon \to 0$?*

A $(1+\delta)$-bi-Lipschitz mapping is just one that distorts distances by at most a factor of $1+\delta$, as in (2.7). If $f : \mathbb{R}^d \to \mathbb{R}^{d+1}$ is $(1+\delta)$-bi-Lipschitz , then $M = f(\mathbb{R}^d)$ satisfies (16.1) and (16.2) with an ϵ which is controlled in terms of δ and which goes to zero when $\delta \to 0$. Question 16.3 asks for the converse to this easy observation.

Notice that ϵ-flat hypersurfaces do not have to be graphs — they can have spirals. We saw this for curves in Section 14, and in higher dimensions we saw in Section 15 that $(1+\delta)$-bi-Lipschitz mappings on Euclidean spaces can spiral around (slowly!) no matter how small δ is.

The answer to Question 16.3 is yes when $d = 1$, because the arc length parameterization provides such a mapping. In higher dimensions this does not work so easily, because there is nothing like the arclength parameterization. Indeed Question 16.3 remains unsolved for all $d > 1$. It is not even known whether an ϵ-flat hypersurface always admits a C-bi-Lipschitz parameterization when ϵ is small enough and C depends only on the dimension.

On the other hand, one can prove directly that ϵ-flat surfaces enjoy many of the properties that would hold if the answer to Question 16.3 is yes. For instance ϵ-flat hypersurfaces are homeomorphic to $\mathbb{R}^d$ when ϵ is small enough. One can build parameterizations with L^p bounds on their differentials and on the differentials of their inverse mappings, with $p \to \infty$ as $\epsilon \to 0$. (See [Sem]$_{SC2}$.) The construction proceeds through an infinite series of approximations, and in order to get bi-Lipschitz estimates one would have to show that the distortion of distance does not get too large anywhere. This seems to be very difficult. If one is willing to settle for L^p bounds for the differential with a $p < \infty$, then it suffices to show that the distortions of distances accumulate only on a small set, and this is much easier to accomplish.

The construction of these parameterizations with L^p bounds proceeds through geometric versions of Calderón-Zygmund approximations. The latter are discussed in a more classical form for analysis in Section 39 below.

There are also nice characterizations of ϵ-flat hypersurfaces in terms of analysis, in terms of the almost-orthogonality of a certain decomposition of $L^2(M)$. See [Sem]$_{SC1}$. These results would be easier to prove if the answer to Question 16.4 were yes, but one can argue directly without knowing that.

I am inclined to believe that the answer to Question 16.3 is no.

There are versions of Question 16.3 for quasisymmetric mappings. One should be a little careful in formulating the appropriate "almost flat" version of the quasisymmetry condition, but this is not a serious problem. In dimension 2 one can find quasisymmetric parameterizations (see $[\text{Sem}]_{SC2}$, $[\text{Hein–Kosk}]_{DQC}$), but this is not known in higher dimensions. The main point of the argument in $[\text{Sem}]_{SC2}$ was to use the uniformization theorem to find a conformal parameterization, and then to use geometric methods to obtain bounds for this conformal mapping. The results of $[\text{Hein–Kosk}]_{DQC}$ permit one to obtain estimates in any dimension as soon as there is a conformal or even a quasiconformal parameterization, but in dimensions larger than 2 they are not so easy to come by. Again I am inclined to believe that they do not exist in general.

To place this question in a larger context let us consider a variant of it formulated in terms of curvature. Let M be a hypersurface in $\mathbb{R}^{d+1}$, and let $k(x)$ denote the maximum of the (absolute values) of the principal curvatures of M at x. This measures the size of the derivative of the unit normal vector to M at x. Consider the condition

$$\int_M k(x)^d \, dx \leq \gamma, \tag{B.16.4}$$

where dx denotes surface measure on M, and γ is a small number. This says that the curvature is small on average, so that the hypersurface ought to be pretty flat. In fact this condition implies ϵ-flatness when γ is small enough, with an ϵ which is controlled by γ and which tends to zero with γ. (See $[\text{Sem}]_{BMO}$.) This condition on curvature is stronger than ϵ-flatness, because it involves more derivatives, but it shares with ϵ-flatness the property of being scale-invariant. When $d = 1$ the integral in (16.4) controls the maximal oscillation of the unit normal to M, and M can be represented as the graph over a line of a Lipschitz function which has small Lipschitz norm. When $d > 1$ (16.4) permits M to have infinite spirals, but they have to be slower than the ones permitted by ϵ-flatness.

When $d = 2$ it turns out that (16.4) is sufficient to imply the existence of a $(1 + \delta)$-bi-Lipschitz parameterization with δ controlled by γ, and with $\delta \to 0$ as $\gamma \to 0$. This was proved in $[\text{Toro}]_{SFF}$. When $d > 2$ this is not known, although there is a related criterion for the existence of such parameterizations in terms of flatness conditions in $[\text{Toro}]_{BP}$.

The curvature condition (16.4) is related to ϵ-flatness through a geometric version of the Sobolev embedding theorem. We have seen in (11.4) how the BMO norm of a function on $\mathbb{R}^d$ is controlled by the L^d norm of its gradient. The situation here is very similar. Let M by a hypersurface in $\mathbb{R}^{d+1}$, and let $n(x)$ denote a (smooth) choice of unit normal vector to M. The curvature condition (16.4) says that the gradient of $n(x)$ as a vector-valued function on M has small L^d-norm on M. If we have something like (11.4) for functions on M, then we can conclude that $n(x)$ has small BMO norm on M, i.e., that $\|n\|_*$ is small, where $\|n\|_*$ is defined as in (11.1), using (9.1) (with μ taken to be the volume measure on M). It is not true that an estimate like (11.4) holds for all hypersurfaces in $\mathbb{R}^{d+1}$, and indeed this becomes wrong when the hypersurface has a big balloon with a small neck. It turns out that this cannot happen when (16.4) is true (with γ small), and there are suitable Poincaré inequalities in this case. See [Simon], $[\text{Sem}]_{BMO}$. Thus the curvature condition implies that $\|n\|_*$ is small. It also happens that ϵ-flatness with small ϵ is equivalent to $\|n\|_*$ being small. (See $[\text{Sem}]_{SC1}$, $[\text{Sem}]_{BMO}$.)

Here again we see how BMO arises naturally in a geometric problem. There is also a natural measure-theoretic aspect of this story. Although we do not know the answer to Question 16.3, we do know that given any ball B centered on M we can parameterize 90% of $B \cap M$ by a subset of $\mathbb{R}^d$ using a $(1+\delta)$-bi-Lipschitz mapping. Actually we can do much better than that, we can realize 90% of $B \cap M$ as a subset of a graph over a hyperplane of a Lipschitz function with small norm. (See $[\text{Sem}]_{SC1}$.)

These geometric statements arise in the following manner. It turns out that for each ball B centered on M there is a hyperplane that approximates M very well in B. One has to understand how this hyperplane turns around as we move from ball to ball, or as we shrink a ball to a point. One can choose the hyperplane so that it is orthogonal to the average of n on B. This leads us to the analytical problem of understanding how the averages of a BMO function can oscillate as we move around. This question can be analyzed using "maximal functions", as discussed in Sections 32 and 42 below.

B.17. The A_∞ condition for doubling measures.

What does it mean for two measures to be almost the same? What is a good way to formulate this idea? We can start with the condition that each be bounded by a constant multiple of the other, but that is too limited.

Recall that a measure α on some measure space is said to be *absolutely continuous* with respect to a measure β if any set with measure zero with respect to β has measure zero with respect to α. The Radon-Nikodym theorem states that if α and β are finite measures, then α is absolutely continuous with respect to β if and only if α is of the form $\alpha = f\,\beta$, where f is an integrable function with respect to β.

If α and β are finite measures, then α is absolutely continuous with respect to β if and only if for every $\epsilon > 0$ there is a $\delta > 0$ so that

$$\beta(E) < \delta \quad \text{implies} \quad \alpha(E) < \epsilon \tag{B.17.1}$$

for each measurable set E. The "if" part is trivial, and the "only if" part can be derived from the Radon-Nikodym theorem.

We shall need here a stronger notion of absolute continuity, a property that takes geometry into account and is scale invariant. Let μ be a doubling measure on $\mathbb{R}^n$. We say that μ is an A_∞ measure, if for each $\epsilon > 0$, there exists a $\delta > 0$ so that given any ball B in $\mathbb{R}^n$ and any measurable subset E of B we have that

$$\frac{|E|}{|B|} < \delta \quad \text{implies} \quad \frac{\mu(E)}{\mu(B)} < \epsilon, \tag{B.17.2}$$

where $|E|$ denotes the Lebesgue measure of E. This sentence is a mouthful, let us take it slowly. The condition (17.2) looks like (17.1), and it implies that μ is absolutely continuous with respect to Lebesgue measure, but it is stronger than that because it requires that a single δ works for all balls B.

Because the A_∞ condition implies absolute continuity we can write an A_∞ measure μ as $\mu = \omega(x)\,dx$ for some locally integrable function $\omega(x)$ on $\mathbb{R}^n$. These functions are called A_∞ weights, and in the harmonic analysis literature the A_∞ condition is normally stated directly in terms of the density.

What are examples of A_∞ weights? Any positive function $\omega(x)$ which is bounded and bounded away from 0 is an A_∞ weight. Some of the examples of doubling measures on $\mathbb{R}^n$ mentioned in Section 3 are actually A_∞ measures. For instance $\omega(x) = |x|^a$ is an A_∞ weight on $\mathbb{R}^n$ for all $a > -n$, and $\omega(x) = |x_1|^b$ is an A_∞ weight on $\mathbb{R}^n$ when $b > -1$. If K denotes the usual middle-thirds Cantor set, then $\operatorname{dist}(x, K)^s$ is an A_∞ weight on $\mathbb{R}$ for a range of s's that one can compute and which is the same as the range for which it is locally integrable. (All $s \geq 0$ are allowed, and some negative s's too.)

The A_∞ condition has the nice property that it is equivalent in nontrivial ways to several other conditions. For instance, it is enough to ask that *there exist* $\epsilon, \delta \in (0,1)$ such that (17.2) holds. A priori this is much weaker than the definition above, in which we ask that for every ϵ there is a δ, but there is a way to iterate this weaker version of (17.2) to get the stronger version. The A_∞ condition is also equivalent to the apparently stronger requirement that there exist $C, a > 0$ so that

$$\frac{\mu(E)}{\mu(B)} \leq C \left(\frac{|E|}{|B|}\right)^a \tag{B.17.3}$$

for all balls B and all measurable subsets E of B. These equivalences are analogous to the results of John-Nirenberg and Strömberg about BMO mentioned in Section 11, and they are discussed again in Section 41.

In fact A_∞ weights enjoy a very close analogy with BMO; they are a multiplicative cousin of BMO. The A_∞ condition is not disturbed if we multiply the weight by a positive constant, the choices of ϵ and δ (or C and a in (17.3)) are unaffected by this multiplication. This corresponds to the fact that the BMO norm of a function is not changed if one adds a constant to the function. If $\omega(x)$ is an A_∞ weight and B is a ball, then there is a positive constant λ (the average of ω over B will work) so that $\omega(x)$ lies within a bounded factor of λ for 90% of the points $x \in B$, where the bounded factor does not depend on B. This corresponds to the theorem of Strömberg mentioned in Section 11.

There are concrete facts behind this similarity between A_∞ weights and BMO. If b is a (real-valued) BMO function on $\mathbb{R}^n$ with sufficiently small BMO norm, then $\omega = e^b$ is an A_∞ weight. This can be proved using the John-Nirenberg theorem. Conversely, if ω is an A_∞ weight, then $\log \omega$ lies in BMO. It is not true that if b is any BMO function, then e^b is an A_∞ weight, and indeed e^b may not even be locally integrable. One is faced with the subset

$$\{\log \omega : \omega \text{ is an } A_\infty \text{ weight}\} \tag{B.17.4}$$

of BMO, which turns out to be convex and open. (Compare with Section 14.)

That (17.4) defines an *open* subset of BMO is especially nice and connected to other basic facts about A_∞ weights. A particularly important result is that a positive locally integrable function $\omega(x)$ is an A_∞ weight if and only if it satisfies a reverse Hölder inequality, which means that there exists a constant $C > 0$ and an exponent $p > 1$ such that

$$\left(\frac{1}{|B|}\int_B \omega(x)^p\,dx\right)^{\frac{1}{p}} \leq C\,\frac{1}{|B|}\int_B \omega(x)\,dx \tag{B.17.5}$$

for all balls B. This may look strange but it is very nice. It says first of all that ω actually lies in L^p_{loc} for some $p > 1$, and not just in L^1_{loc}. The inequality (17.5) would not mean as much if C were allowed to depend on B, but by having uniform estimates for all balls (17.5) imposes much stronger constraints on the oscillation of ω.

Note that (17.3) can be derived from from (17.5) using Hölder's inequality. The openness of (17.4) in BMO can be proved using (17.5) and the John-Nirenberg theorem. See Section 41 for an explanation of how to get the reverse Hölder inequality (17.5).

One can think of the concept of an A_∞ weight as providing a way to perturb Lebesgue measure without making too much of a change. For instance, if we define BMO relative to an A_∞ weight instead of Lebesgue measure, then the resulting space is the same as BMO for Lebesgue measure. (There is a converse to this due to Peter Jones, and also results of Reimann for quasiconformal mappings.) A_∞ weights and their more refined cousins the A_p weights cooperate well with singular integral operators. They also show up on their own, in connection with Riemann mappings and harmonic measure, for instance.

The A_∞ condition has an important symmetry property. We defined the concept of an A_∞ measure relative to Lebesgue measure, but we could have replaced Lebesgue measure by any other doubling measure. Indeed the theory actually works on spaces of homogeneous type, so that it makes sense to say that one doubling measure enjoys the A_∞ property with respect to another. It turns out that the A_∞ property defines an equivalence relation. The transitivity follows easily from the characterization (17.2) with the quantifiers "for every $\epsilon > 0$ there is a $\delta > 0$", while the symmetry follows from (17.2) with the quantifiers "there exist $\epsilon, \delta \in (0,1)$". Note that the transitivity works in the same as way as for absolute continuity but the symmetry is fundamentally different.

Thus we have the concept of deforming a doubling measure by an A_∞ weight, and we expect that this deformation is pretty mild. This idea has a nice geometric version. We saw in (3.6) how to associate a quasimetric to a doubling measure. If we apply this construction to two doubling measures which are A_∞ equivalent, then on 90% of any given ball the two quasimetrics will look approximately the same, except for a scale factor. That is, after making a rescaling (multiplying by a positive constant), the two quasimetrics will each be less than a bounded multiple of the other on 90% of the ball. This can be proved using some information from "maximal functions", as in Lemma 42.5.

The concept of "90% of the points in the ball" captures well the idea of the A_∞ condition. The statement that two doubling measures are A_∞ equivalent means roughly that they both have the same concept of "90% of the points in a ball" even though they may give very different measures to the same set. In this statement the concept of 90% should be taken somewhat figuratively as a way of saying nearly all. The number 90 is not special, just evocative. In passing from one measure to another which is A_∞ equivalent the percentages might have to change, but the point is that the concept of "nearly all" remains the same, in a way that is made precise by (17.2).

See [Garn], [Jour], and [Stein]$_{HA}$ for more information about A_∞ weights.

B.18. Quasisymmetric mappings and doubling measures.

Suppose that f is a quasisymmetric mapping from $\mathbb{R}^n$ onto itself, and let μ be the measure on $\mathbb{R}^n$ obtained by pulling back Lebesgue measure using f, i.e., $\mu(E) = |f(E)|$. What can we say about μ? It is easy to see that μ is a doubling measure. When $n = 1$ we cannot say anything more than that. Given any doubling measure on the line we can integrate it to get a quasisymmetric homeomorphism on the line so that the given doubling measure is the pull-back of Lebesgue measure.

What happens in higher dimensions? It is not so easy to integrate a measure and get a mapping.

B.18.1. Theorem (Gehring [Gehring]): *If $f : \mathbb{R}^n \to \mathbb{R}^n$ is quasisymmetric and the measure μ is defined by $\mu(E) = |f(E)|$, and if $n > 1$, then μ is an A_∞ measure (as defined in the previous section). In particular it is absolutely continuous with respect to Lebesgue measure and its density lies in L^p_{loc} for some $p > 1$.*

This is a wonderful fact.

Why exactly is dimension 1 different from higher dimensions here? It comes down to the ancient length-area principle, that quasisymmetric mappings have to distort length and area in approximately the same way. This does not mean much in dimension 1, where length and "area" are the same. In order to get to the central point let us assume that $f : \mathbb{R}^n \to \mathbb{R}^n$ is quasisymmetric and also smooth. In this case we have that

$$\sup_{|v|=1} |df_x(v)| \leq K \inf_{|v|=1} |df_x(v)|, \tag{B.18.2}$$

where df_x denotes the differential of f at x – the linear mapping provided by calculus which best approximates the behavior of f at x – and K is a constant that depends only on the function that governs the quasisymmetry of f (the function η in Definition 4.1). The supremum and infimum in (18.2) are taken over all unit vectors $v \in \mathbb{R}^n$, and (18.2) says that the maximal stretch of the differential is always bounded by the minimal stretch, even though these numbers may be very small or very large. A diffeomorphism f on $\mathbb{R}^n$ which satisfies (18.2) is said to be *K-quasiconformal*, and it turns out that the concepts of quasiconformality and quasisymmetry are equivalent on Euclidean spaces, modulo technicalities pertaining to differentiability assumptions.

Now let Q be a cube in $\mathbb{R}^n$, and let us study the relationship between the distortion of length and area by f near Q. Let $\omega(x)$ denote the Jacobian of f at x (the determinant of df_x). The measure μ associated to f as in the statement of Theorem 18.1 is given by $\mu = \omega(x)\,dx$. We have that

$$\int_Q \omega(x)\,dx \le C\,(\operatorname{diam} f(Q))^n, \tag{B.18.3}$$

where C depends only on n and the function that governs the quasisymmetry of f. Indeed, the left hand side is just the volume of $f(Q)$, and this volume is bounded by the right side of (18.3).

On the other hand, let σ be a line segment that connects two of the opposite faces of Q. Let us check that

$$\int_\sigma \omega(z)^{\frac{1}{n}}\,dz \ge C^{-1}\ \operatorname{diam} f(Q) \tag{B.18.4}$$

for a suitable constant C, where the dz in the integral denotes the arclength measure. Indeed, we have that $|df_z| \le K\,\omega(z)^{\frac{1}{n}}$, because of (18.2). This implies that the integral on the left side of (18.4) is larger than a multiple of the length of $f(\sigma)$, and hence of $\operatorname{diam} f(\sigma)$. The quasisymmetry condition implies that $\operatorname{diam} f(Q)$ is bounded by a multiple of $\operatorname{diam} f(\sigma)$, and (18.4) follows.

In this way we are able to conclude that

$$\left(\frac{1}{|Q|}\int_Q \omega(x)\,dx\right)^{\frac{1}{n}} \le C\,\frac{1}{\operatorname{diam} Q}\int_\sigma \omega(z)^{\frac{1}{n}}\,dz \tag{B.18.5}$$

whenever σ is a line segment which connects opposite faces of Q. By averaging over σ we can convert this to

$$\left(\frac{1}{|Q|}\int_Q \omega(x)\,dx\right)^{\frac{1}{n}} \le C\,\frac{1}{|Q|}\int_Q \omega(x)^{\frac{1}{n}}\,dx. \tag{B.18.6}$$

This is an interesting inequality. It contains no information whatsoever when $n = 1$. When $n > 1$ it is a reverse Hölder inequality, similar to (17.5). That is, the average of ω over a cube is controlled by the L^s average over the cube, where $s = 1/n < 1$. Gehring proved that once we have such a reverse Hölder inequality we can get a slightly better one, i.e., there exist constants $M, p > 1$ depending on C and n from (18.6) such that

$$\left(\frac{1}{|Q|}\int_Q \omega(x)^p\,dx\right)^{\frac{1}{p}} \le M\left(\frac{1}{|Q|}\int_Q \omega(x)^{\frac{1}{n}}\,dx\right)^n. \tag{B.18.7}$$

(See also Section 41.) It is very important here that we know (18.6) for all cubes Q; it is not true that (18.6) implies something like (18.7) for each fixed cube separately.

It is also important here that the constants in (18.7) depend finally only on the dimension n and the function that governs the quasisymmetry of f. The bottom line is that ω satisfies a reverse Hölder inequality like (17.5), but with balls replaced by cubes, which does not matter. This permits us to conclude that ω is an A_∞ weight, with bounds that depend only on n and the function that governs the quasisymmetry of f.

This does not quite prove Theorem 18.1, because of our a priori smoothness assumption on f, but it does cover the main points.

Gehring's theorem is quite remarkable because it provides very strong restrictions on the way in which quasisymmetric mappings on $\mathbb{R}^n$ can distort distances when $n > 1$. It provides a much stronger restriction than is given directly by the definition, or indeed which is true in general, i.e., than what is true on the real line, or on Cantor sets. For instance, it implies that if B is a ball in $\mathbb{R}^n$, $n > 1$, and $f : \mathbb{R}^n \to \mathbb{R}^n$ is quasisymmetric, then f is actually bi-Lipschitz on 90% of B. In fact, one can normalize f by a scale factor on B to get a uniform bound on the bi-Lipschitz constant. These assertions can be derived from Lemma 42.5 below.

Notice that the quasisymmetry condition is not disturbed by composing the mapping with a dilation. Any estimates that we can hope to get for the distortion of distances by quasisymmetric mappings has to respect that. This corresponds nicely to the fact that the A_∞ condition is not disturbed by multiplying the weight by a multiplicative constant.

One might wonder whether Theorem 18.1 has counterparts for other metric spaces. Other regular metric spaces (Definition 3.8), for instance, for which the formulation of the statement still makes sense (with the measure provided in Definition 3.8 playing the role of Lebesgue measure). It doesn't work for Cantor sets, but it does work for a lot of other spaces. A criterion in terms of the existence of Poincaré inequalities is given in [Hein–Kosk]$_{QCG}$. Gehring's argument can be adapted to situations where there are plenty of curves around that are not too long, e.g., a metric space of the form $N \times \mathbb{R}$ would work, where N is a (regular) metric space. A different variation on Gehring's theme is given in [Sem]$_{Quasi}$. It is not clear yet what is really the right context for Gehring's theorem though.

Presumably one can say more for Sierpinski carpets than for Cantor sets, but less than for Euclidean spaces.

B.19. Metric doubling measures.

Does every A_∞ measure on $\mathbb{R}^n$ arise as in Gehring's theorem? When $n = 1$ the answer is yes, because one can just integrate the measure. When $n > 1$ this is less clear. We should give ourselves some room for maneuver;

let us ask whether it is true that for each A_∞ measure μ on $\mathbb{R}^n$ there is a quasisymmetric mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ and a constant $C > 0$ such that

$$C^{-1}\,\mu(E) \le |f(E)| \le C\,\mu(E) \tag{B.19.1}$$

for all Borel sets E.

This is not true when $n > 1$. Take $\mu = |x_1|\,dx$, where x_1 denotes the first coordinate function. One can check that this defines an A_∞ measure. However it cannot satisfy (19.1) for any quasisymmetric mapping f. If it did, one could show that the differential of f exists everywhere on the $x_1 = 0$ hyperplane and vanishes, and conclude that f is constant there, a contradiction. More generally there is a problem when the density of μ becomes too small on curves that are not too long.

In order to avoid this problem we make an additional assumption.

B.19.2. Definition: *Let μ be a doubling measure on $\mathbb{R}^n$, and define the associated quasimetric $D(x,y)$ on $\mathbb{R}^n$ as in (3.6), with $\alpha = 1/n$. We say that μ is a* metric doubling measure *if there exists a constant $C > 0$ and a metric $\delta(x,y)$ on $\mathbb{R}^n$ such that*

$$C^{-1}\,\delta(x,y) \le D(x,y) \le C\,\delta(x,y) \tag{B.19.3}$$

for all $x, y \in \mathbb{R}^n$.

In other words, the quasimetric $D(x,y)$ associated to μ is likely to not be a metric, but we ask that it be comparable to one in size. For instance, $\mu = |x|^a\,dx$ has this property for all $a > -1$, but when $n > 1$ the measure $\mu = |x_1|^b\,dx$ has this property only when $-1 < b \le 0$, even though we get a doubling measure for all $b > -1$. The problem is the same as the one above; if μ gets too small near a curve that is not too long, and if μ is a metric doubling measure, then we would get into trouble because $\delta(x,y)$ would have to vanish for some pairs of points x, y with $x \ne y$. One can think of this as meaning that the $D(x,y)$-lengths of curves are not permitted to become too small when μ is a metric doubling measure.

This concept was introduced in [Dav–Sem]$_{ASQ}$, with a slightly different name ("strong A_∞ weights"), and with a different but equivalent definition. One of the basic observations was the following.

B.19.4. Lemma: *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is quasisymmetric and μ is a measure on $\mathbb{R}^n$ which satisfies (19.1). Then μ is a metric doubling measure.*

This is an easy consequence of the definitions, with $\delta(x,y) = |f(x) - f(y)|$.

So a necessary condition for (19.1) to hold is that μ be a metric doubling measure. When $n = 1$ this does not contain information beyond the requirement that μ be doubling, because every doubling measure on the line is a metric doubling measure. (Take $\delta(x, y) = \mu([x, y])$.) When $n > 1$ the metric doubling condition does contain nontrivial information.

B.19.5. Theorem: *If μ is a metric doubling measure on $\mathbb{R}^n$, $n > 1$, then μ is an A_∞ measure on $\mathbb{R}^n$.*

This is another observation from [Dav–Sem]$_{ASQ}$. More precisely, in [Dav–Sem]$_{ASQ}$ it is observed that Theorem 19.5 can be proved in the same manner as Gehring proved Theorem 18.1. This is not hard to understand. In the argument outlined in Section 18, we never really needed the mapping, just the measure. The key point is to have a version of (18.4), which should say that if Q is a cube in $\mathbb{R}^n$ and σ is a line segment which connects opposite faces of Q, then the "length" of σ defined with respect to the quasimetric $D(x, y)$ associated to μ is bounded from below by a constant times the diameter of Q with respect to $D(x, y)$. This turns out to be true, because of the metric doubling measure condition. In order to be able to compute everything in a reasonable way, it is better to be able to assume that μ has a density which is positive and continuous, but it turns out to be easy to reduce to that case by an approximation argument. (See [Sem]$_{Bil}$ for details.)

In view of Theorem 19.5 we have to ask ourselves whether every metric doubling measure on $\mathbb{R}^n$ is associated to a quasisymmetric mapping as in (19.1). Let us first try to understand what (19.1) means. Let μ be a metric doubling measure on $\mathbb{R}^n$, and suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is quasisymmetric and satisfies (19.1). Then f defines a bi-Lipschitz mapping from $(\mathbb{R}^n, D(x, y))$ to $(\mathbb{R}^n, |x - y|)$, where $D(x, y)$ is the quasimetric associated to μ as in Definition 19.2. This bi-Lipschitzness is not hard to verify using the definitions. (For this we only need (19.1) applied to balls rather than general subsets E of $\mathbb{R}^n$.)

Conversely, suppose that f is a bi-Lipschitz mapping from $(\mathbb{R}^n, D(x, y))$ to $(\mathbb{R}^n, |x - y|)$. Then f is quasisymmetric with respect to the Euclidean metric, because of Proposition 4.6 and the fact that the composition of quasisymmetric mappings is quasisymmetric. One can then show that (19.1) must hold for balls. Once one has that one can deduce (19.1) for general Borel sets.

Thus the existence of a quasisymmetric mapping associated to μ is equivalent to the bi-Lipschitz equivalence of $(\mathbb{R}^n, D(x, y))$ and $(\mathbb{R}^n, |x - y|)$.

Are these spaces always bi-Lipschitz equivalent? What are the potential obstructions? We should start by asking whether $(\mathbb{R}^n, D(x, y))$ is

Ahlfors regular of dimension n, but it is, as discussed in Section 4. Next we could ask whether a ball in $(\mathbb{R}^n, D(x,y))$ is approximately a topological ball, in the sense that it is contained in a topological ball which is contained in a metric ball whose radius is larger than the original one by only a bounded factor. This is true, as one can see using Proposition 4.6. Are points in $(\mathbb{R}^n, D(x,y))$ connected by curves of length comparable to their $D(\cdot,\cdot)$-distance? True again, a consequence of (17.5), as discussed in [Dav–Sem]$_{ASQ}$, [Sem]$_{Bil}$. More subtle potential obstruction: if a bi-Lipschitz mapping exists, then one should have roughly the same Sobolev and Poincaré inequalities for $(\mathbb{R}^n, D(x,y))$ as for ordinary Euclidean space. Also true, by [Dav–Sem]$_{ASQ}$. (See also [Sem]$_{Rmks}$.)

So maybe it is true that $(\mathbb{R}^n, D(x,y))$ is always bi-Lipschitz equivalent to $(\mathbb{R}^n, |x-y|)$ when μ is a metric doubling measure. As mentioned in Section 17, the A_∞ condition implies that for each ball B in $\mathbb{R}^n$ the identity mapping scaled by an appropriate constant dilation gives a uniformly bi-Lipschitz mapping from $D(x,y)$ to $|x-y|$ on a subset of B containing 90% of its points. So simply the identity provides bi-Lipschitz mappings on large pieces of $(\mathbb{R}^n, D(x,y))$, and maybe one can combine them somehow to get a bi-Lipschitz mapping on the whole space.

This turns out to be wrong, starting in dimension 3 anyway. The counterexample is of the form $\mu = \operatorname{dist}(x,A)^N\,dx$, where A is an Antoine's necklace and N is sufficiently large. Antoine's necklaces are self-similar Cantor sets which have the property that their complement in $\mathbb{R}^n$ is not simply-connected. If one embeds a Cantor set into $\mathbb{R}^n$ in the usual way the complement is simply-connected, but this embedding is not standard. Instead of iterating the rule that a segment is replaced by a disjoint union of segments, or similarly with cubes, Antione's necklaces are obtained by iterating the rule that a solid torus is replaced by a necklace of disjoint solid tori which link. Antoine showed that the complement of such a set has nontrivial fundamental group. By taking N large, though, one can arrange it so that this set is very small in $(\mathbb{R}^n, D(x,y))$, e.g., that it has Hausdorff dimension as small as you like. (One can arrange it so that for each integer $k > 0$ we can cover A by at most 100^k balls with respect to $D(\cdot,\cdot)$ each of whose radius is $\leq \epsilon^k$, where $\epsilon > 0$ can be taken as small as we like by taking N large enough.) If $(\mathbb{R}^n, D(x,y))$ were bi-Lipschitz equivalent to $(\mathbb{R}^n, |x-y|)$, then the image of A in the target space would be a compact subset of $\mathbb{R}^n$ whose complement is not simply-connected and which is very small with respect to the Euclidean metric, i.e., has Hausdorff dimension as small as we like. This contradicts well-known results about general position in Euclidean spaces. By taking N as large as we like, we can even get that there is no homeomorphism from $(\mathbb{R}^n, D(x,y))$ to $(\mathbb{R}^n, |x-y|)$ which is locally Hölder continuous of a positive exponent given in advance. See $[\mathrm{Sem}]_{BP}$ for details.

The problem is unsolved when $n = 2$. Compare also with Question 16.3.

The spaces considered in this section are always quasisymmetrically equivalent to standard Euclidean spaces, but there are examples in $[\mathrm{Sem}]_{Good}$ of very nice spaces which have most of the same properties as above but which do not admit quasisymmetric parameterizations. These examples begin in dimension 3, in dimension 2 there is a positive result based on the uniformization theorem. See $[\mathrm{Sem}]_{SC2}$, $[\mathrm{Dav–Sem}]_{QR}$, and $[\mathrm{Hein–Kosk}]_{DQC}$.

I suspect that it should be possible to show that the difficulties in building bi-Lipschitz mappings in dimensions > 1 and quasisymmetric mappings in dimensions > 2 follow from the fact that one cannot build isometries in general in dimensions > 1 (no arclength parameterization) and that one cannot build conformal mappings in general in dimensions > 2 (the failure of the uniformization theorem), but I do not know how to turn this principle into a precise theorem.

B.20. Bi-Lipschitz embeddings.

Under what conditions is a metric space $(M, d(x, y))$ bi-Lipschitz equivalent to a subset of some $\mathbb{R}^N$? Or quasisymmetrically equivalent?

We can always embed a metric space isometrically into a Banach space, into the space of bounded continuous functions on M. But what if we want to land in a finite dimensional space?

It is not hard to see that it is necessary for M to be doubling in order for this to occur. This follows from Corollary 2.5 and the fact that bi-Lipschitz and quasisymmetric equivalence preserve the doubling property.

Is the converse true? Consider the following result of Assouad ($[\mathrm{Assou}]_{EM}$, $[\mathrm{Assou}]_{DM}$, $[\mathrm{Assou}]_{PL}$).

B.20.1. Theorem: *If $(M, d(x, y))$ is doubling, then for each $0 < s < 1$ there is an $N = N(s)$ such that $(M, d(x, y)^s)$ is bi-Lipschitz equivalent to a subset of $\mathbb{R}^N$.*

This is pretty good. It settles the question of quasisymmetric equivalence, and gives some information about metric spaces in general, namely that snowflakes are always okay.

The snowflake functor can destroy a lot of the most interesting aspects of the geometry of a metric space. We discussed this in Section 12, in the context of the rigidity properties of Lipschitz mappings. It also destroys the existence of curves of finite length.

We have seen before that it is much easier to build Hölder continuous mappings with prescribed behavior at various scales and locations than Lipschitz mappings. In order to build a bi-Lipschitz embedding of an arbitrary metric space $(M, d(x, y))$ which is doubling we would have to confront this problem.

It turns out that the answer to the question is no, that there are metric spaces which are doubling that do not admit bi-Lipschitz embeddings into finite-dimensional Euclidean spaces. The examples are based on the Heisenberg group and were known to Assouad. Recall that the Heisenberg group was defined in Section 2 and given a certain metric which is doubling. As mentioned in Section 12, Pansu $[\text{Pan}]_{CC}$ proved that Lipschitz functions from the Heisenberg group into the real line have the property that they have a derivative at Lebesgue-almost every point which is a group homomorphism. This fact obviously extends to Lipschitz mappings into any $\mathbb{R}^N$. The key point now is that any group homomorphism from the Heisenberg group into $\mathbb{R}^N$ has to have a large kernel, the commutator subgroup, which has dimension 1 in this case. If the mapping were bi-Lipschitz , then the limiting mapping obtained by blowing up would also be bi-Lipschitz , and injective in particular. This gives a contradiction.

Let us consider a more special case. Suppose that μ is a doubling measure on $\mathbb{R}^n$, and let $D(x, y)$ be the quasimetric defined as in (3.6) with $\alpha = 1/n$. Can we embed $(\mathbb{R}^n, D(x, y))$ bi-Lipschitzly into some $\mathbb{R}^N$?

There is an obvious necessary condition, which is that μ be a metric doubling measure (Definition 19.2).

One might hope that the answer for metric doubling measures is yes, since we saw in the previous section that for them $(\mathbb{R}^n, D(x, y))$ is very close to the standard $\mathbb{R}^n$ in its geometry, but in fact this is not true.

B.20.2. Proposition: *If $(M, d(x, y))$ is any metric space which is doubling, then we can find a positive integer n and a metric doubling measure μ on $\mathbb{R}^n$ such that $(M, d(x, y))$ is bi-Lipschitz equivalent to a subset of $(\mathbb{R}^n, D(x, y))$, where $D(x, y)$ is associated to μ as in (3.6), with $\alpha = 1/n$.*

Thus the question of bi-Lipschitz embeddability for doubling metric spaces in general is equivalent to the question for the spaces that arise from deforming Euclidean spaces using metric doubling measures.

Proposition 20.2 is proved in $[\text{Sem}]_{BP}$. The construction is quite simple. Let $(M, d(x, y))$ be given. Using Assouad's theorem we can make $(M, d(x, y)^{1/2})$ be bi-Lipschitz equivalent to a subset E of some $\mathbb{R}^n$. On this $\mathbb{R}^n$ we take μ to be defined by $\mu = \operatorname{dist}(x, E)^n \, dx$. One has to check that this works, but the argument is pretty straightforward.

In $[\text{Sem}]_{Bil}$ it is shown that $(\mathbb{R}^n, D(x, y))$ does admit a bi-Lipschitz embedding into some $\mathbb{R}^N$ for "most" metric doubling measures. The precise condition is a technical, but it says that a bi-Lipschitz embedding exists when there is another metric doubling measure which is a little bit smaller than the given one in a certain sense. In the proof the embedding is constructed in a sequence of layers, where the layers come from an analysis like the Calderón-Zygmund approximation described in Section 39 below. The extra hypothesis on the metric doubling measure ensures that this infinite family of layers can be combined successfully.

One does have bi-Lipschitz embeddability for the examples based on Antoine's necklaces discussed in the preceding section.

B.21. A_1 weights.

B.21.1. Definition: *Let $\omega(x)$ be a positive locally integrable function on $\mathbb{R}^n$. We say that ω is an A_1* weight *if there exists a constant $C > 0$ such that*

$$\frac{1}{|B|}\int_B \omega(x)\,dx \leq C \operatorname*{essinf}_{y\in B} \omega(y) \tag{B.21.2}$$

for all balls B in $\mathbb{R}^n$. Here "essinf" denotes the essential infimum (where one compensates for sets of measure zero).

This is another reverse Hölder inequality, as in (17.5) and (18.6). It turns out that A_1 weights are always A_∞ weights, but the reverse is very much not true. For instance, $\omega(x) = |x|^a$ is an A_∞ weight as soon as $a > -n$ (which is needed for local integrability), but to get an A_1 weight we have to have $-n < a \leq 0$. The point is that A_1 weights are not permitted to vanish in a nontrivial way.

If ω is an A_1 weight and $\mu = \omega(x)\,dx$, then μ is a metric doubling measure. This is not hard to check. (See $[\mathrm{Sem}]_{Bil}$.)

Roughly speaking, A_1 weights are just like A_∞ weights, except that they have only a large part, not a small part. This is made precise by the factorization theorem of Peter Jones, which characterizes A_∞ weights as the product of an A_1 weight and a negative power of an A_1 weight (a large part and a small part). See [Garn], [Jour], $[\mathrm{Stein}]_{HA}$ for more information, and about how A_1 weights can be constructed.

The examples of metric doubling measures which did not correspond to mappings in the two previous sections came from A_∞ weights that were not A_1 weights. In both cases the weights had to vanish somewhere, they were positive powers of the distance to a nonempty set. Thus we have the following open problem: If ω is an A_1 weight on $\mathbb{R}^n$, $\mu = \omega(x)\,dx$, and $D(x,y)$ is the quasimetric associated to μ as in (3.6) with $\alpha = 1/n$, is it true that $(\mathbb{R}^n, D(x,y))$ is bi-Lipschitz equivalent to $(\mathbb{R}^n, |x-y|)$?

It is known that the A_1 condition is sufficient to ensure bi-Lipschitz embeddability into some $\mathbb{R}^N$. (See $[\mathrm{Sem}]_{Bil}$.)

B.22. Interlude: bi-Lipschitz mappings between Cantor sets.

So far Cantor sets have been the least rich in structure for us. What about bi-Lipschitz mappings between them?

Let F_2 and F_3 be sets with 2 and 3 elements (respectively), and define F_2^∞ and F_3^∞ as in Section 2. Define metrics $\rho_2(x,y)$ and $\rho_3(x,y)$ on F_2^∞ and F_3^∞ as in (2.10), with $a = 1/2$ in the case of F_2^∞ and $a = 1/3$ in the case of F_3^∞. This gives a pair of metric spaces, each Ahlfors regular of dimension 1, as in Section 3. Are they bi-Lipschitz equivalent?

The fact that they are both Ahlfors regular with the same dimension implies that they have roughly the same properties in terms of sizes and coverings. Since they are so disconnected one might think that it should be easy to make them bi-Lipschitz equivalent, by sliding pieces around. Remember that Cantor sets are all homeomorphic, and so it seems plausible that one could make the homeomorphisms bi-Lipschitz when the sizes match up.

This turns out to be wrong. See [Fal–Mar], [Co–Pign].

B.23. Another moment of reflection.

We are repeatedly running into situations where we can say that the existence of a mapping implies the presence of some geometric structure, but we do not know how to go backwards and produce a mapping given some geometry. In some cases we know that this is impossible. In general, it is not easy to find homeomorphic parameterizations of spaces, or to find interesting homeomorphisms, bi-Lipschitz mappings, or quasisymmetric mappings. That is to actually obtain them under reasonable circumstances, rather than just making special examples to illustrate a point. It is easier to build interesting spaces. (See also $[\mathrm{Sem}]_{Map}$.) Interesting measures are still easier to make. (See also $[\mathrm{Sem}]_{Quasi}$.)

We turn now to the problem of finding structure in spaces even when we cannot get well-behaved homeomorphic parameterizations.

B.24. Rectifiability.

We began Section III. with the fact that Lipschitz functions are differentiable almost everywhere. Now we want to look at this issue in the context of geometry, in which functions are replaced by sets, derivatives are replaced by tangent planes, and so forth. General references for this section include [Fal], $[\mathrm{Fed}]_{GMT}$, [Mattila].

To begin we need the notion of Hausdorff measure. We shall work always in a Euclidean space $\mathbb{R}^n$ in this section. Given $s \geq 0$, a subset A of $\mathbb{R}^n$, and a $\delta > 0$, set

$$H^s_\delta(A) = \inf \Big\{ \sum_j (\operatorname{diam} E_j)^s : \{E_j\} \text{ is a sequence of sets in } \mathbb{R}^n \text{ which covers } A \text{ and satisfies } \operatorname{diam} E_j < \delta \tag{B.24.1}$$

$$\text{for all } j \Big\}, \tag{B.24.2}$$

and then define the s-dimensional Hausdorff measure of A by

$$H^s(A) = \lim_{\delta \to 0} H^s_\delta(A). \tag{B.24.3}$$

This limit always exists because $H^s_\delta(A)$ is monotone in δ, but the limit might be infinite. If s is an integer and A lies on a nice submanifold of $\mathbb{R}^n$ of dimension s, then $H^s(A)$ coincides with the usual surface measure of A, except perhaps for a normalizing factor. In particular we recapture Lebesgue measure when $s = n$. In general H^s is a countably subadditive outer measure which restricts to a countably additive measure on its collection of measurable subsets, and the Borel sets are measurable for all the Hausdorff measures.

Fix an integer d, $0 < d < n$. Let us call a subset E of $\mathbb{R}^n$ *rectifiable* if we can write E as $E = (\bigcup_i E_i) \cup N$, where $\{E_i\}$ is a sequence of subsets of $\mathbb{R}^n$ which are each bi-Lipschitz equivalent to a subset of $\mathbb{R}^d$, and where $H^d(N) = 0$. This notion of rectifiability depends very much on the choice of d, and we should mention d explicitly if there is any doubt. The reader should beware of the different uses of this terminology, "countable rectifiability" would be the name in [Fed]$_{GMT}$, another name still in [Fal].

What does the concept of rectifiability really mean? One can think of it as a counterpart for sets of the concept of differentiability almost everywhere for functions. It is a way of saying that at almost all points the set behaves well, almost like a Euclidean space.

Before we get too far into that aspect of rectifiability, let us back up and consider some more basic matters of definition. Suppose that we take the same definition as above except that we replace the requirement that the E_i's be bi-Lipschitz equivalent to subsets of $\mathbb{R}^d$ with the requirement that they be subsets of C^1 submanifolds (of dimension d) of $\mathbb{R}^n$. Then we would get an equivalent definition. That may seem surprising, since C^1 mappings are much nicer than Lipschitz mappings, but the point is in the measure theory, and in the differentiability almost everywhere of Lipschitz mappings. To be explicit consider a Lipschitz mapping $f : A \to \mathbb{R}^n$, where A is a subset of $\mathbb{R}^d$. We may as well assume that A is closed, because f extends to the closure of A automatically. We can also extend f to a Lipschitz mapping on all of $\mathbb{R}^d$. A well-known consequence of the differentiability almost everywhere of Lipschitz mappings is that for each $\epsilon > 0$ there is a C^1 mapping $g : \mathbb{R}^d \to \mathbb{R}^n$ which equals f everywhere except on a set of measure $< \epsilon$. (See [Fed]$_{GMT}$.) This is how one can get C^1 mappings into the picture. These error sets have small measure, and the notion of rectifiability is designed to be undisturbed by countable unions or bad sets of measure zero, in such a way that the error sets do not really matter in the end. Similarly the points at which the differential of g has rank less than d are potentially bad, but the image of this set under g has measure zero and so we may forget about it too. From here it is not too hard to see that we can get an equivalent concept of rectifiability if we take the definition of rectifiability above and ask instead that the E_i's lie on C^1 submanifolds.

A similar reasoning shows that we can ask only that the E_i's be the images of subsets of $\mathbb{R}^d$ under Lipschitz mappings (instead of bi-Lipschitz mappings).

It is not hard to believe that rectifiable sets have tangent planes almost everywhere, at least in some sense. The usual notion of tangent plane does not work here, where one takes a point x in E and asks that all points on E near x lie very close to the tangent plane, because we are allowing countable unions and arbitrary sets of measure zero. E might contain a countable dense set, for instance, or a countable dense family of little surfaces. In this measure-theoretic setting one has to work with the concept of "approximate tangent planes", in which one demands that most points in E near x lie very close to the tangent plane, where "most" is computed in terms of Hausdorff measure. See [Fal], [Fed]$_{GMT}$, and [Mattila].

Once we know that almost all points of E are contained in a countable union of C^1 manifolds, it is not so surprising that approximate tangents exist almost everywhere. The main point is to be careful about the measure theory to control the way in which all these submanifolds are intersecting.

Conversely, the existence of approximate tangent planes almost everywhere implies rectifiability. One needs here much less precision than a tangent plane; cones would be fine. This corresponds again to the fact that Lipschitz functions and C^1 functions are almost the same in this context.

To detect rectifiability in practice we need Federer's structure theorem. Before we get to that we need to have the notion of unrectifiability. A subset E of $\mathbb{R}^n$ is called *totally unrectifiable* if $H^d(E \cap R) = 0$ for all rectifiable sets R. This is equivalent to asking that $H^d(E \cap R) = 0$ whenever R is a C^1 submanifold of $\mathbb{R}^n$. Again the dimension d is implicit and should be used throughout, and the terminology is not universal.

A set E is automatically unrectifiable if $H^d(E) = 0$, but that is not much fun. In general one should think of unrectifiable sets as being very scattered or very crinkled. One can make examples from Cantor sets and snowflakes for instance.

A basic fact is that a given set E can be written as $A \cup B$, where A is rectifiable and B is totally unrectifiable, if E is H^d-measurable and $H^d(E) < \infty$. There is a very elegant proof of this fact which is well-known. Set

$$\alpha = \sup\{H^d(R) : R \subseteq E, R \text{ is rectifiable}\}. \tag{B.24.4}$$

The first point is that this supremum is attained. Indeed, we can find a sequence $\{R_j\}$ of rectifiable subsets of E such that $H^d(R_j) \to \alpha$, and we can take $A = \bigcup_j R_j$. We can take the R_j's to be measurable also. With this choice of A we have that $B = E\backslash A$ is totally unrectifiable, since otherwise B would have a nontrivial rectifiable piece which could be added to A to make it larger, in contradiction to the maximality of $H^d(A)$.

This shows that there is a perfect dichotomy between rectifiability and unrectifiability. It is particularly useful in combination with the following theorem of Federer.

B.24.5. Theorem: *Suppose that $E \subseteq \mathbb{R}^n$ is H^d-measurable and $H^d(E) < \infty$. Then E is totally unrectifiable if and only if the projection of E onto almost every d-plane in $\mathbb{R}^n$ has H^d measure zero.*

The Grassmann space of d-planes in $\mathbb{R}^n$ is a smooth manifold and so there is a natural notion of subsets of measure zero in it. It is that notion that we use here.

See [Fed]$_{GMT}$, [Mattila] for a proof of this theorem.

Theorem 24.4 is really great. It can be refined as follows. Given a (measurable) subset E of $\mathbb{R}^n$, its d-dimensional integral-geometric measure is defined in the following manner. Let V be a d-plane in $\mathbb{R}^n$ which passes through the origin. We first compute the measure of the orthogonal projection of E onto V, counting multiplicities. We then take the average of these numbers over all d-planes V. This is the integral-geometric measure of E. If $H^d(E) < \infty$ and E is unrectifiable, then we get 0 as the answer. If E is rectifiable, then we get back $H^d(E)$ as the answer, modulo a normalizing factor perhaps, by a well-known theorem. (See [Fed]$_{GMT}$.) Thus we conclude that if E is H^d-measurable and $H^d(E) < \infty$, then the integral-geometric measure of E computes the size of the rectifiable part of E.

This gives a very useful way to detect the presence of rectifiable sets, through topological information for instance. To illustrate this point, assume that $d = n - 1$ and that E is a closed set with $H^d(E) < \infty$ which separates at least two points in its complement. Then E must have a nontrivial rectifiable part, because one can find a lot of hyperplanes onto which E has a nontrivial projection.

We have the idea that unrectifiable sets are sets which are very scattered, but Federer's Theorem 24.4 makes this precise in a very useful way.

B.25. Uniform rectifiability.

What if we want to have some kind of theory of rectifiability which provides estimates?

We have seen that it can be very hard to find bi-Lipschitz parameterizations of actual given sets. On the other hand one can often verify the rectifiability of such a set, with Federer's Theorem 24.4 used as a key tool. But there is a big gap between knowing that a set is rectifiable and having a bi-Lipschitz parameterization for it.

Let us begin with a quantitative measure-theoretic assumption. Fix integers $0 < d < n$, and let us think about d-dimensional subsets of $\mathbb{R}^n$. A subset E of $\mathbb{R}^n$ is said to be *(Ahlfors) regular* (with dimension d) if it is closed and if there is a constant $C > 0$ such that

$$C^{-1}\, r^d \leq H^d(E \cap B(x,r)) \leq C\, r^d \tag{B.25.1}$$

whenever $x \in E$ and $0 < r \leq \operatorname{diam} E$. This is equivalent to Definition 3.8. This is a necessary condition for a bi-Lipschitz parameterization by Euclidean space, but it is far from sufficient. Self-similar fractals, like Cantor sets and snowflakes, are Ahlfors regular.

Now we define uniform rectifiability, a quantitative and scale invariant version of the notion of rectifiability. A set E in $\mathbb{R}^n$ is said to be *uniformly rectifiable* if it is Ahlfors regular of dimension d, and if there exist constants $\theta, M > 0$ with the property that for each $x \in E$ and $0 < r < \operatorname{diam} E$ we can find a subset A of $E \cap B(x,r)$ such that

$$H^d(A) \geq \theta\, r^d \qquad \text{and} \tag{B.25.2}$$

$$A \text{ is } M\text{-bi-Lipschitz equivalent to a subset of } \mathbb{R}^d. \tag{B.25.3}$$

In other words, inside of each ball centered on E we wish to have a substantial portion which is bi-Lipschitz equivalent to a subset of $\mathbb{R}^d$, with uniform bounds.

This is a complicated definition, let us try to understand it slowly. Think of the basic problem of deciding when a set E is itself bi-Lipschitz equivalent to $\mathbb{R}^d$. We have decided that this problem is practically unsolvable, and so we are asking for less, while still trying to keep as much of the same flavor of the bi-Lipschitz parameterization as possible. If we do not ask for an actual bi-Lipschitz parameterization, then it is reasonable to ask that a substantial proportion of E admit such a parameterization. On the other hand one of the particular features of bi-Lipschitz mappings is their scale invariance, take a snapshot at any location and scale and you have the same estimates. Uniform rectifiability merges these two ideas: the idea of saying that we can find a good parameterization for most points, and the idea of having uniform bounds at all scales and locations. It is very close to the ideas of BMO and A_∞ weights.

Like the notion of rectifiability, uniform rectifiability enjoys some nice stability properties. In the definition above we permit θ to be as small as we want, so long as it is fixed. Suppose that we prescribe a small $\epsilon > 0$ in advance, and ask instead of (25.2) that

$$H^d((E \cap B(x,r))\backslash A) \leq \epsilon\, r^d. \tag{B.25.4}$$

It turns out that if we make this change then we get a condition which is equivalent to the one above. The price to pay for taking ϵ small is to increase M in (25.3).

Regular subsets of uniformly rectifiable sets are uniformly rectifiable. The union of two uniformly rectifiable sets is uniformly rectifiable.

If we replace (25.3) with the requirement that A be the M-Lipschitz image of a subset of $\mathbb{R}^d$ of diameter $\leq r$, then we also get an equivalent condition. This relies on a theorem of Peter Jones [Jones]$_{Lip}$.

It turns out that if E is Ahlfors regular and is unbounded, then E is uniformly rectifiable if and only if $E \times \mathbb{R}$ is (as a subset of $\mathbb{R}^{n+1}$). (If E is bounded one should take the product with a finite interval.) This type of statement is at best unclear for the question of bi-Lipschitz parameterizations by $\mathbb{R}^d$.

One of the basic questions is whether one can detect uniform rectifiability under reasonable geometric conditions. The ideal would be to have something like Federer's Theorem 24.4. One would be happy to know that if a set has bounded Hausdorff measure and if there is a lower bound on the measure of its projections (counted without multiplicities) onto a large collection of d-planes, then the set has a substantial piece which is bi-Lipschitz equivalent to a subset of $\mathbb{R}^d$, with uniform bounds. This is a natural quantitative conjecture motivated by Federer's theorem. Unfortunately it remains unknown. There is a method of David ([David]$_{MG}$, see also [David]$_{WSI}$) which provides uniform rectifiability results in many of the cases in which one would like to apply a quantitative version of Federer's result, in many situations in which one has suitable "lower bounds on topology", for instance. See also [Dav–Jer], [Dav–Sem]$_{QR}$, [Sem]$_{Find}$, [Jon–Ka–Var], [DS5].

B.26. Stories from the past.

One of the original motivations for studying uniform rectifiability came from analysis, namely the problem of knowing on which sets certain classes of singular integral operators were bounded on L^2. This question grew out of the Calderón program (see [Cald]), and earlier work of Calderón, Coifman, MacIntosh, Meyer, and David showed that lots of singular integral operators were bounded on sets that were bi-Lipschitz equivalent to Euclidean spaces. (See [Coif–Dav–Mey], $[\text{David}]_{WSI}$ for more information and further references.) Such sets have little smoothness, but they come with parameterizations, which is a lot. The question then was what kind of sets were really allowed, what kind of smoothness was needed, what kind of parameterization, and whether there were simple geometric criteria for singular integrals to be well behaved.

Much of the previous work focussed on 1-dimensional sets, in which the issue of the existence of a parameterization is much less significant, because of arclength parameterizations. In higher dimensions there is no arclength parameterization, and the relationship between geometry and parameterization is much less clear.

For one-dimensional sets the best result is due to David $[\text{David}]_{PC}$, who showed that singular integral operators are well behaved on curves which are Ahlfors regular of dimension 1. He did this by proving two things: that such curves are uniformly rectifiable, and that singular integral operators are well behaved on uniformly rectifiable sets. (Actually he worked with a slightly stronger version of uniform rectifiability.) The second statement, that uniform rectifiability implies good behavior for singular integral operators, works in all dimensions. It is not immediately clear how to formulate the first statement for higher dimensional sets, because of the lack of arclength parameterizations. For one formulation, based on a kind of parameterization discussed in the next section, David proved in $[\text{David}]_{SR}$ that singular integral operators are well behaved, using something slightly weaker than uniform rectifiability.

A very different approach to the problem was given in $[\text{Sem}]_{BSI}$. A geometric condition was given there for sets of codimension 1 to be well behaved for singular integral operators. This condition requires that the set be topologically nontrivial in a uniform way at all scales and locations, as measured by separating points in the complement. The methods used were analytical, and did not provide much insight into the geometry of the set.

The story of ϵ-flat hypersurfaces in $[\text{Sem}]_{SC1}$ was also somewhat disturbing for the idea of parameterizations. Here was a class of hypersurfaces which were much better than uniformly rectifiable but for which it was not at all clear that there was a nice global parameterization, even of the type in [David]$_{SR}$.

Another disturbing fact was a criterion in [Sem]$_{SC2}$ for the existence of well-behaved parameterizations of 2-dimensional sets. This criterion provided the existence of a global parameterization which was strong enough to imply uniform rectifiability but which was not accommodated by the class of parameterizations permitted in the first version of [David]$_{SR}$ (a situation David later rectified by enlarging his class of parameterizations). This criterion had also the annoying feature that it relied on the uniformization theorem, and so gave information in a mysterious way which did not work in higher dimensions.

These were the circumstances at the time that David did [David]$_{MG}$. It was not clear then that there was a common geometric thread to these examples. Something like uniform rectifiability was a clear guess, but it was not clear that one should always have it, or that it was reasonable to ask always for good parameterizations of large pieces of sets. One of the main problems was that it was not known that the sets considered in [Sem]$_{BSI}$ had any uniform rectifiability properties or good parameterizations. Rectifiability was clear because of Federer's structure theorem. One of the main results of [David]$_{MGL}$ was that these sets are uniformly rectifiable. Simpler proofs of this were given later in [Dav–Jer] and [Dav–Sem]$_{QR}$, and [Sem]$_{Rect}$ provided an approach to the geometry of these sets through analysis.

David also showed in [David]$_{MGL}$ that uniform rectifiability is implied by the existence of the kind of parameterizations considered in [David]$_{SR}$. He gave in addition a criterion for uniform rectifiability for sets of codimension larger than 1 analogous to the condition in [Sem]$_{BSI}$ in codimension 1 (in terms of uniform lower bounds on topology). No other method is known for doing this in higher codimensions. Analytical methods as in [Sem]$_{BSI}$ do not seem to work very well in higher codimensions.

In [Dav–Sem]$_{SI}$ a converse was obtained, to the effect that if you have an Ahlfors regular set on which sufficiently many singular integral operators are well behaved, then the set must be uniformly rectifiable. Some other characterizations were established, including some of the ones mentioned earlier. These results showed that there was really only one natural notion of uniform rectifiability. It was also shown that uniformly rectifiable sets are always contained in sets which admit a parameterization of the type in [David]$_{SR}$.

For more information about these topics see [Dav–Sem]$_{UR}$ and [Sem]$_{Find}$.

B.27. Regular mappings.

B.27.1. Definition: *Let $(M, d(x,y))$ and $(N, \rho(u,v))$ be metric spaces (or quasimetric spaces). We say that a mapping $f : M \to N$ is regular if it is Lipschitz and if for each ball B in N we can cover $f^{-1}(B)$ by at most C balls in M with the same radius as B, where C does not depend on B.*

For example, $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$ is regular. A bi-Lipschitz embedding is regular if M is doubling, but regular mappings need not be bi-Lipschitz . Note that a regular mapping can have only a bounded number of preimages for each point in the range.

Regular mappings provide us with a variation on the theme of what constitutes a good parameterization. Instead of asking that a set admit a bi-Lipschitz parameterization by a Euclidean space, we can ask for a regular parameterization, which allows some crossings. We could also ask that a given set be contained in a regular image of a Euclidean space, i.e., we could allow some holes, for the sake of balance.

One of the results of $[\text{David}]_{MG}$ is that the image of $\mathbb{R}^d$ in any $\mathbb{R}^n$ under a regular mapping is uniformly rectifiable. It is not known whether every uniformly rectifiable set is contained in such a regular image.

This situation provides a nice example of how metric doubling measures (Definition 19.2) can be used, as a way to perturb Euclidean geometry modestly. Let μ be a metric doubling measure on $\mathbb{R}^d$, and let $D(x,y)$ denote the associated quasidistance, as in (3.6), with $\alpha = 1/d$. It turns out that the image of $(\mathbb{R}^d, D(x,y))$ in $\mathbb{R}^n$ under a regular mapping is always uniformly rectifiable. This comes from the A_∞ property of metric doubling measures (Theorem 19.5), which helps to ensure that the geometry of $(\mathbb{R}^d, D(x,y))$ is sufficiently close to being Euclidean. Conversely, given a uniformly rectifiable subset of $\mathbb{R}^n$ with dimension d, then it is contained in the image of such a regular mapping for some metric doubling measure μ. For this we need to assume that n is large enough compared to d, or we have to settle for a regular mapping into $\mathbb{R}^{n+1}$. See $[\text{Dav–Sem}]_{SI}$.

For the converse, for the construction of a regular mapping which contains a given uniformly rectifiable set in its image, we can take the metric doubling measure μ to be of the form $\omega(x)\,dx$, where ω is actually an A_1 weight (Definition 21.1). It remains unknown whether $(\mathbb{R}^d, D(x,y))$ is always bi-Lipschitz equivalent to $(\mathbb{R}^d, |x-y|)$ when the density of μ is an A_1 weight (as discussed in Section 21), and if this is true one could get rid of the weight in this story of regular mappings.

When does a metric space $(M, d(x,y))$ admit a regular mapping into some finite dimensional Euclidean space? A necessary condition is that M

$(M, d(x,y))$ which admits a regular mapping into some finite dimensional Euclidean space also admits a bi-Lipschitz embedding into some Euclidean space of larger but still finite dimension.

B.28. Big pieces of bi-Lipschitz mappings.

B.28.1. Theorem ([David]$_{MG}$): *Let Q denote the unit cube in $\mathbb{R}^n$. Suppose that $f : Q \to \mathbb{R}^n$ is Lipschitz, and that $|f(Q)| \geq \delta > 0$. Then there exists an $\epsilon > 0$ such that there is a subset E of Q with $|E| \geq \epsilon$ such that the restriction of f to E is ϵ^{-1}-bi-Lipschitz . This ϵ depends only on the dimension n, the Lipschitz constant for f, and δ, and not on the specific choice of f.*

I like this theorem very much. A crucial point is to have universal bounds.

This is not true in general for metric spaces, even under suitable measure-theoretic assumptions so that it all makes sense. It is not true for mappings on Cantor sets for instance. (David and I found counterexamples.) Another example: consider the snowflake $([0,1], |x-y|^{1/2})$. A well-known construction of a space-filling curve provides a Lipschitz mapping from this space onto the unit square in $\mathbb{R}^2$, but no such mapping can be bi-Lipschitz on a set of positive measure. (Lipschitz and bi-Lipschitz maps from $([0,1], |x-y|^{1/2})$ into itself are the same as Lipschitz and bi-Lipschitz maps with respect to the Euclidean metric, so there is nothing new there.)

A third example: Start with the set $F = \{0, 1\}$, construct the Cantor set F^∞ of infinite sequences of elements of F, as in Section 2, and give it a metric $d(x, y)$ as in (2.10), with $a = 1/2$. This defines a metric space which is Ahlfors regular of dimension 1 (Definition 3.8). There is an obvious mapping from F^∞ onto $[0,1]$, obtained by taking an element of F and interpreting it as a binary sequence of a real number. This mapping is Lipschitz, and even regular, but one can show that it cannot be bi-Lipschitz on a set of positive measure.

Thus David's Theorem 28.1 again reflects rigidity properties of Lipschitz mappings on Euclidean spaces which are not true abstractly. This rigidity property has the nice feature that it is easier to formulate for general spaces (like Ahlfors regular metric spaces) than something like differentiability almost everywhere.

Theorem 28.1 is closely related to differentiability almost everywhere. This point is easier to understand in terms of the fact that for each Lipschitz mapping $f : Q \to \mathbb{R}^n$ (where Q is still the unit cube in $\mathbb{R}^n$) and every $\delta > 0$ there is a C^1 mapping $g : Q \to \mathbb{R}^n$ which equals f except on a set of measure mapping $f : Q \to \mathbb{R}^n$ (where Q is still the unit cube in $\mathbb{R}^n$) and every $\delta > 0$ there is a C^1 mapping $g : Q \to \mathbb{R}^n$ which equals f except on a set of measure $< \delta$. Using this fact it is not hard to show that if $f(Q)$ has positive measure, then f must be bi-Lipschitz on a set of positive measure. Indeed, this comes down to the inverse function theorem, which implies that a C^1 function is bi-Lipschitz on a neighborhood of any point at which the differential is invertible, and the fact that the set of points where the differential of a C^1 mapping is not invertible gets mapped to a set of measure zero in the image.

This analysis implies Theorem 28.1 but without the uniform bound on ϵ. The remarkable feature of David's theorem is that it does provide such a bound.

Peter Jones gave the following improvement of David's theorem.

B.28.2. Theorem ($[\text{Jones}]_{Rec}$): *Let Q denote the unit cube in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}^n$ be a 1-Lipschitz mapping. Let $\eta > 0$ be given. Then there are compact subsets $K_1, \ldots, K_N$ of Q such that the restriction of f to each K_j is N-bi-Lipschitz , and such that*

$$\left| f(Q \setminus \bigcup_{i=1}^{N} K_i) \right| < \eta. \tag{B.28.3}$$

Here N depends only on the dimension n and on η.

In other words, we can break up the domain Q into a bounded number of pieces, in such a way that one of these pieces has a small image, and the restriction of f to each of the others is bi-Lipschitz. As an illustration think about the mapping $z \mapsto z^k$ on the unit disk in $\mathbb{C}$. Note that the Lipschitz norm of this mapping is k.

If we did not want to have a bound on the number of pieces K_i, then Theorem 28.2 would not be so exciting. It would follow from much simpler arguments, about approximating Lipschitz functions by C^1 functions.

Although we do not know much about building bi-Lipschitz parameterizations of whole spaces, there is some pretty good technology for finding bi-Lipschitz mappings on sets of a definite size. This is one of the reasons why the notion of uniform rectifiability works well.

One of the main points in Jones' proof is to have some quantitative understanding of how often Lipschitz functions are well approximated by affine functions. We shall take a closer look at this topic in the next section. For now let us just say that a key feature of affine functions is that if you control their Lipschitz norm, and if you know that they do not shrink volumes too much, then you can control their bi-Lipschitz constant. Jones combines this trivial observation with the good approximations of Lipschitz functions by affine functions and a coding argument to prove his Theorem 28.2. (See Section 43 for slightly more information.)

Jones' argument also works when the target space is $\mathbb{R}^m$ and $m > n$, but then one must be careful about the kind of "measure" to use in (28.3) (Hausdorff content instead of Hausdorff measure).

See [Dav–Sem]$_{QR}$ for some variations on these themes.

B.29. Quantitative smoothness for Lipschitz functions.

How smooth are Lipschitz functions on Euclidean spaces? We know that they are differentiable almost everywhere, but can we say more than that? Can we say something quantitative? The Lipschitz condition is a quantitative condition; it is not nice to have differentiability almost everywhere but not a quantitative statement. One has a bound on the size of the derivative, but what about some kind of bound on the "differentiability", on the extent to which Lipschitz functions are approximately affine?

What kind of quantitative statement might we be able to hope for? We cannot hope for bounds on the rate of convergence of the limit in the derivative.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given, and consider the quantity

$$\alpha(x,t) = t^{-1} \inf \left\{ \sup_{y \in B(x,t)} |f(y) - a(y)| : \ a \text{ is an affine function} \right\}. \tag{B.29.1}$$

Here $x \in \mathbb{R}^n$ and $t > 0$, and $\alpha(x,t)$ provides a measure of the oscillation of f on the ball $B(x,t)$. It measures the extent to which the snapshot of f on $B(x,t)$ is almost affine. The normalization here – dividing by t – is the natural one for taking the snapshot of a Lipschitz function, just like in the definition of the derivative. With this normalization $\alpha(x,t)$ is always bounded by the Lipschitz norm of f (just take $a(y)$ to be the constant function $f(x)$). The question is whether $\alpha(x,t)$ is small as a dimensionless quantity.

If f is differentiable at a point x, then

$$\lim_{t\to 0} \alpha(x,t) = 0. \tag{B.29.2}$$

The vanishing of this limit is not equivalent to the existence of a derivative. If $\alpha(x,t)$ is small, then f is well-approximated by an affine function on $B(x,t)$, but this affine function can spin around as $t \to 0$. The existence of a derivative means that (29.2) holds and that these affine approximations do not spin around, that in fact a single affine function works well for all (small) t.

If f is Lipschitz, then what can we say about the $\alpha(x,t)$'s? How often are they small? We know that (29.2) must hold for almost all x, but can we do better than this? Can we get a bound for the number of times that the $\alpha(x,t)$'s are not small? It turns out that we can do this, but to do this properly we need a new concept, the concept of Carleson sets. Before we get to that let us make some more naive observations.

The $\alpha(x,t)$'s fit well with our idea of snapshots of a function — they are measuring the size of the oscillations of these snapshots. Our collection of snapshots is parameterized by $\mathbb{R}^n \times \mathbb{R}_+$, where $\mathbb{R}_+$ denotes the set of positive real numbers. That is, $\mathbb{R}^n \times \mathbb{R}_+$ parameterizes the space of balls in $\mathbb{R}^n$, and we have a snapshot of f for each ball.

In Section 1 we saw how hyperbolic geometry on $\mathbb{R}^n \times \mathbb{R}_+$ can be natural in analysis, and this is one of those occasions. In this case this means that $\alpha(x,t)$ does not change a whole lot when we move (x,t) by a modest amount in the hyperbolic metric. Actually, for our purposes of getting bounds on $\alpha(x,t)$, it is a little better to use an observation that does not quite fit as well with the idea of hyperbolic geometry. Namely, given $x, y \in \mathbb{R}^n$ and $s, t \in \mathbb{R}_+$, we have that

$$\alpha(y,s) \le 4\,\alpha(x,2\,t) \qquad \text{whenever } |x-y| \le t \text{ and } \frac{t}{2} \le s \le t. \tag{B.29.3}$$

This is not hard to check from the definitions, the main point being that $B(y,s) \subseteq B(x,2\,t)$ under the assumptions above. This observation says that if $\alpha(x,2t)$ is small for some (x,t), then $\alpha(y,s)$ is small for a lot of (y,s) which are close to (x,t) hyperbolically. The extra factor of 2 in front of the t is a little unpleasant for crisp geometric statements, but this observation implies nonetheless that we have to respect hyperbolic geometry on $\mathbb{R}^n \times \mathbb{R}_+$ to some extent when we try to understand how often the $\alpha(x,t)$'s are small. We should not expect anything too interesting to happen to $\alpha(x,t)$ at small distances in the hyperbolic metric. We should instead look at large hyperbolic distances. For instance, (29.3) implies that if (29.2) holds, so that $\alpha(x,t)$ is small for (x,t) on a small vertical ray (i.e., for x fixed and all small $t > 0$), then $\alpha(y,s)$ is automatically small for all (y,s) in a cone around that ray (a uniform hyperbolic neighborhood of the ray).

Carleson sets provide a useful notion of "small" subsets of $\mathbb{R}^n \times \mathbb{R}_+$. A measurable subset $\mathcal{A}$ of $\mathbb{R}^n \times \mathbb{R}_+$ is said to be a *Carleson set* if there is a constant $C > 0$ such that

$$\int_0^t \int_{B(x,t)} \mathbf{1}_{\mathcal{A}}(y,s)\,\frac{dy\,ds}{s} \le C\,t^n \tag{B.29.4}$$

for all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}_+$. Here $\mathbf{1}_{\mathcal{A}}(y,s)$ denotes the characteristic function of $\mathcal{A}$. This condition is a mouthful, let us take it slowly.

We are integrating over the region $B(x,t) \times (0,t)$ in $\mathbb{R}^n \times \mathbb{R}_+$. If we think of $\mathbb{R}^n \times \mathbb{R}_+$ as parameterizing the locations and scales in $\mathbb{R}^n$, then $B(x,t) \times (0,t)$ parameterizes the locations and scales near x and finer than t. In practice we shall look at sets $\mathcal{A}$ like

$$\{(x,t) \in \mathbb{R}^n \times \mathbb{R}_+ : \alpha(x,t) > \epsilon\}, \tag{B.29.5}$$

for fixed $\epsilon > 0$. In this case the membership of (x,t) in $\mathcal{A}$ depends only on the values of f in $B(x,t)$. The integral in (29.4) depends only on the values of f in $B(x,2t)$. This integral measures the total oscillations of f at all scales and locations inside $B(x,2t)$.

If the integrand $\mathbf{1}_{\mathcal{A}}(y,s)$ were not present in (29.4), then the integral would diverge, because of the ds/s integral. Indeed,

$$\int_a^b \int_{B(x,t)} \frac{dy\,ds}{s} = c(n)\,t^n \log\frac{b}{a}, \tag{B.29.6}$$

where $c(n)$ is in fact the volume of the unit ball in $\mathbb{R}^n$. This is why the Carleson condition is a smallness condition. Already the convergence of the integral in (29.4) implies that $\mathcal{A}$ is small in a nontrivial way.

Here is a basic example. Suppose that F is a subset of the integers, and set

$$\mathcal{A} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R}_+ : 2^j < t < 2^{j+1} \text{ for some } j \in F\}. \tag{B.29.7}$$

Then $\mathcal{A}$ is a Carleson set if and only if F is finite, and in this case the Carleson constant (the best constant C for (29.4)) is proportional to the number of elements of F.

The Carleson condition (29.4) is a way of trying to count the average number of layers in $\mathcal{A}$. Fix $\mathcal{A}$ and $t > 0$, and set

$$N_t(y) = \int_0^t \mathbf{1}_{\mathcal{A}}(y,s)\,\frac{ds}{s}. \tag{B.29.8}$$

This tries to count the number of layers in $\mathcal{A}$ directly above y up to the level t. The ds/s in the integral is natural here, natural for the hyperbolic geometry, a "layer" should be something like $[s,2s]$ for some $s > 0$. Different layers in $\mathbb{R}^n \times \mathbb{R}_+$ correspond to different scales in $\mathbb{R}^n$. We can rewrite (29.4) as

$$t^{-n} \int_{B(x,t)} N_t(y)\,dy \le C. \tag{B.29.9}$$

In other words the average of $N_t(y)$ over $B(x,t)$ should be bounded.

If $\mathcal{A}$ is as in (29.5), and f is Lipschitz, then (29.2) holds almost everywhere. This implies that $N_t(y) < \infty$ almost everywhere. If $\mathcal{A}$ is a Carleson set, then one gets much more than $N_t < \infty$ a.e.

B.29.10. Theorem: *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz and $\epsilon > 0$ is arbitrary, then (29.5) defines a Carleson set, with a Carleson constant bounded by a constant that depends only on n, ϵ, and the Lipschitz constant of f.*

This is a very interesting fact. I am not sure exactly how to attribute it. Jones $[\text{Jones}]_{Lip}$ was the first person that I know to explicitly state and use it. Results roughly like this have been stated before – it is easy to derive Theorem 29.10 from [Dorr] – but $[\text{Jones}]_{Lip}$ is the first place where I saw this kind of fact used in a meaningful way, to prove Theorem 28.2.

Let us consider some more examples. Let H be a hyperplane in $\mathbb{R}^n$, and set $f(x) = \operatorname{dist}(x,H)$. Notice that f is affine on each of the two complementary components of H. This implies that

$$\alpha(x,t) = 0 \qquad \text{when } \operatorname{dist}(x,H) \geq t, \tag{B.29.11}$$

where $\alpha(x,t)$ is again as in (29.1). If we take $\mathcal{A}$ to be as in (29.5) (for some $\epsilon > 0$), then

$$\mathcal{A} \subseteq \{(x,t) \in \mathbb{R}^n \times \mathbb{R}_+ : \operatorname{dist}(x,H) < t\}. \tag{B.29.12}$$

It is not hard to check directly that the right hand side is a Carleson set. In this case the multiplicity functions $N_t(x)$ are not bounded, but we do have the average bounds (29.9). Indeed, $N_t(x) = 0$ when $\operatorname{dist}(x,H) \geq t$, and when $\operatorname{dist}(x,H) < t$, we have that $N_t(x)$ is roughly like $\log(t/\operatorname{dist}(x,H))$.

A more elaborate example to consider is $f(x) = \operatorname{dist}(x,E)$, where E is any subset of $\mathbb{R}^n$. We might as well take E to be closed. This gives a 1-Lipschitz function on $\mathbb{R}^n$ which can be a little more complicated for the $\alpha(x,t)$'s, and I do not think that Theorem 29.10 is so trivial in this case. A relevant Carleson set for this example is

$$\{(x,t) \in \mathbb{R}^n \times \mathbb{R}_+ : \operatorname{dist}(x,E) < t < 2 \operatorname{dist}(x,E)\}. \tag{B.29.13}$$

For this set, the functions $N_t(x)$ are all bounded, and the set has only one "layer", but it is not at a uniform height. The Carleson constant does not depend on the choice of E.

As another illustration of Theorem 29.10 let us consider the following consequence of it. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz with norm 1, say, and let $\epsilon > 0$ be given. Then there is a constant k which depends only on ϵ and n so that for any ball $B(x,t)$ in $\mathbb{R}^n$, there is a ball $B(y,s) \subseteq B(x,t)$ such that $s \geq t/k$ and $\alpha(y,s) < \epsilon$ (f is almost affine on $B(y,s)$). On average we should expect that there are plenty of balls $B(y,s)$ like this, but even in the worst case there is always one which is not too different from $B(x,t)$. It is not hard to derive this from Theorem 29.10, using (29.6), for instance. This fact is slightly crude compared to the whole truth, but it provides a nice manifestation of the uniform and scale-invariant nature of the smoothness of Lipschitz functions.

If a function on $\mathbb{R}^n$ satisfies the conclusions of Theorem 29.10, to what extent must it be like a Lipschitz function? Not so much, as in the next example. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \sum_{j=1}^{\infty} a_j \, 2^{-j} \, \exp(2^j \, i \, x), \tag{B.29.14}$$

where $\{a_j\}$ is a bounded sequence of numbers. (Back to lacunary series, as in Section 12.) One can compute that

$$\lim_{t \to 0} \sup_{x \in \mathbb{R}^n} \alpha(x,t) = 0 \qquad \text{if } \lim_{j \to \infty} a_j = 0, \tag{B.29.15}$$

where $\alpha(x,t)$ is as in (29.1). Also, this function f is bounded, and we have that

$$\alpha(x,t) \leq t^{-1} \, \|f\|_\infty, \tag{B.29.16}$$

whence $\lim_{t \to \infty} \sup_{x \in \mathbb{R}^n} \alpha(x,t) = 0$. Therefore $\lim_{j \to \infty} a_j = 0$ implies that for each $\epsilon > 0$ there exist $\delta, R > 0$ such that

$$\alpha(x,t) > \epsilon \quad \text{implies} \quad \delta < t < R. \tag{B.29.17}$$

In this case the conclusions of Theorem 29.10 certainly hold, because the N_t's are all uniformly bounded (for a fixed ϵ). On the other hand, in order for the distributional derivative of f to lie in $L^2_{loc}(\mathbb{R})$ it is necessary for $\sum_j |a_j|^2 < \infty$. This follows from basic facts about Fourier series. By taking sequences $\{a_j\}$ which tend to 0 but which are not square summable we can get functions which satisfy the conclusion of Theorem 29.10 very well but which are badly behaved in terms of differentiability properties. In fact more refined results from harmonic analysis imply that in this case f' will not look like a function in any reasonable sense, not just that it will not be in $L^2_{loc}(\mathbb{R})$.

So we cannot really characterize anything like the Lipschitz property in terms of the conclusions of Theorem 29.10. To get such a characterization we have to impose stronger conditions. There are basically two reasonable ways to do that. The first is to impose stronger conditions on the size of the $\alpha(x,t)$'s. We shall not enter into the details, but Theorem 29.10 is a reduction of a better estimate on quantities like the $\alpha(x,t)$'s, and the reduction is like Chebychev's theorem (controlling the measure of a set where a function is large rather than the integral). The better estimate can be derived from [Dorr], and involves L^2 integrals. The second reasonable way to strengthen the conclusion of Theorem 29.10 is to impose conditions on how fast the good affine approximations to f spin around. This second approach is based on Carleson's corona construction, and is exposed in Chapter 2 of Part IV of [Dav–Sem]$_{UR}$. Actually, these strengthenings of the conclusion of Theorem 29.10 do not capture quite the Lipschitz property of a function, but are better adapted to something like functions whose gradient lies in BMO. This is not a catastrophic difference though.

So why should we expect any kind of good estimates for the $\alpha(x,t)$'s at all? Why should we expect them to be small most of the time, when the Lipschitz condition doesn't seem to say much more than their boundedness? All the good estimates that I know come down to some kind of orthogonality, e.g., Plancherel's theorem asserting that the Fourier transform preserves the L^2 norm, or estimates derived from some other Hilbert space consideration. Let us consider functions on the real line, specifically a function f in L^2 of the real line. It is well known that $f' \in L^2(\mathbb{R})$ if and only if

$$\sup_{t>0} \int_{\mathbb{R}} \left| \frac{f(x+t) - f(x)}{t} \right|^2 dx < \infty. \tag{B.29.18}$$

In this case

$$\frac{f(x+t) - f(x)}{t} \to f'(x) \quad \text{as } t \to 0 \tag{B.29.19}$$

in L^2. On the other hand $f' \in L^2(\mathbb{R})$ if and only if

$$\int_0^\infty \int_{\mathbb{R}} \left| \frac{f(x+t) + f(x-t) - 2\,f(x)}{t} \right|^2 \frac{dx\,dt}{t} < \infty. \tag{B.29.20}$$

This is not too hard to verify, using the Fourier transform. Plancherel's theorem transforms the x-integral into an integral on the Fourier side, and then the t-integral splits off. The conclusion is that the double integral in (29.20) is equal to a constant multiple of $\int |f'|^2$.

Let us compare (29.20) and (29.18). We are interested in the quantities

$$\int_{\mathbb{R}} \left| \frac{f(x+t) - f(x)}{t} \right|^2 dx \quad \text{and} \quad \int_{\mathbb{R}} \left| \frac{f(x+t) + f(x-t) - 2\,f(x)}{t} \right|^2 dx, \tag{B.29.21}$$

as functions of t. The first is bounded as a function of t in the situation of interest, and we cannot say more as $t \to 0$ because (29.19) ensures that it tends to a nonzero limit when f is not constant. The second quantity is also bounded but in fact it does tend to 0 as $t \to 0$ when $f' \in L^2$. It has to be small most of the time in order for the double integral in (29.20) to be finite, because the dt/t gives infinite measure to $(0, 1)$. A bound on (29.20) gives a bound on how often it is small, on average.

The second difference $f(x+t) + f(x-t) - 2\,f(x)$ is closely related to $\alpha(x, t)$. If $\alpha(x, t) = 0$, then f equals an affine function on $B(x, t)$, and hence the second difference $f(x + t) + f(x - t) - 2\,f(x)$ vanishes. One cannot control $\alpha(x, t)$ directly in terms of $f(x+t) + f(x-t) - 2\,f(x)$, but one can control $\alpha(x, t)$ in terms of averages of second differences.

As a practical matter there are more convenient ways to control the $\alpha(x, t)$'s than using second differences, but the principles remain the same. One first controls some quantities that measure the oscillation of functions using the Fourier transform or other orthogonality methods, and then one controls the $\alpha(x, t)$'s in terms of averages of these quantities.

Although Theorem 29.10 does not provide the definitive information about the smoothness of Lipschitz functions, it does provide nontrivial information that is useful for geometry. The $\alpha(x, t)$'s are better for geometry than other common measurements of oscillation in analysis, and simply the knowledge that they are smaller than a small threshold is often more useful than more subtle measurements of their size. Theorem 29.10 is pretty sharp for its formulation. Practically all Carleson sets arise in this manner (although it would be easier to allow functions f with $\nabla f \in BMO$ instead of just Lipschitz functions).

See [Dav–Sem]$_{QR}$, [Dav–Sem]$_{UR}$ for variations on the theme of Theorem 29.10.

B.30. Smoothness of uniformly rectifiable sets.

How smooth are uniformly rectifiable sets? The short answer is that they are practically as smooth as Lipschitz functions, with the "smoothness" of Lipschitz functions interpreted as in the preceding section. There is a particularly nice formulation of this statement which corresponds to Theorem 29.10, which we shall discuss now.

Let integers $0 < d < n$ be given, and let E be a d-dimensional Ahlfors regular set in $\mathbb{R}^n$ (as in (25.1)). The concept of Carleson sets in $E \times \mathbb{R}_+$ can be formulated in exactly the same manner as before, with E playing the role of $\mathbb{R}^n$, and with the restriction of Hausdorff measure H^d to E playing the role of Lebesgue measure on $\mathbb{R}^n$. Ahlfors regularity ensures that Carleson sets behave in the same way as on Euclidean spaces.

How do we measure the smoothness of a set? For functions we like to measure the extent to which they are well-approximated by affine functions, while for sets we like to measure the extent to which they are well-approximated by planes. Given $x \in E$ and $t > 0$ set

$$b\beta(x,t) = \inf_P \Big\{ t^{-1} \sup_{y \in E \cap B(x,t)} \operatorname{dist}(y,P) + t^{-1} \sup_{z \in P \cap B(x,t)} \operatorname{dist}(z,E) \Big\}. \tag{B.30.1}$$

The "b" here stands for "bilateral", and the β was inherited from Peter Jones $[\text{Jones}]_{Squ}$, $[\text{Jones}]_{Rec}$. The infimum in (30.1) is taken over all d-planes P, and $b\beta(x,t)$ measures the extent to which E is approximately equal to a d-plane in $B(x,t)$. Note that $b\beta(x,t) = 0$ if and only if E coincides with a d-plane inside $B(x,t)$. The factor of t^{-1} is a useful normalization which ensures that $b\beta(x,t) \leq 1$. $b\beta(x,t)$ is a dimensionless quantity.

We say that E satisfies the "Bilateral Weak Geometric Lemma (BWGL)" if

$$\{(x,t) \in E \times \mathbb{R}_+ : b\beta(x,t) > \epsilon\} \tag{B.30.2}$$

is a Carleson set inside $E \times \mathbb{R}_+$ for each $\epsilon > 0$. (Compare with (29.5).)

B.30.3. Theorem: *A regular set E is uniformly rectifiable (Section 25) if and only if it satisfies the BWGL.*

This is proved in $[\text{Dav–Sem}]_{UR}$ (Theorem 2.4 in Part I).

The fact that uniform rectifiability implies the BWGL can be seen as a geometric version of Theorem 29.10. The converse, the fact that the BWGL implies uniform rectifiability, can be seen as the geometric analogue of false statements about functions. We saw in the previous section how the conclusions of Theorem 29.10 do not come too close to capturing the Lipschitzness of a function, and it is amusing that when we switch to geometry this problem goes away. This phenomenon can be better understood in terms of a "unilateral" version of $b\beta(x,t)$, namely

$$\beta(x,t) = \inf_P \Big\{ t^{-1} \sup_{y \in E \cap B(x,t)} \operatorname{dist}(y,P) \Big\}. \tag{B.30.4}$$

Here again the infimum is taken over all d-planes. This quantity measures the extent to which E lies close to a d-plane inside $B(x,t)$, but it does not require that points in $B(x,t)$ which lie on the d-plane also lie close to E. Holes are allowed. Thus $\beta(x,t) = 0$ means that $E \cap B(x,t)$ lies on a d-plane, but it need not coincide with the intersection of $B(x,t)$ with the d-plane. If we ask that

$$\{(x,t) \in E \times \mathbb{R}_+ : \beta(x,t) > \epsilon\} \tag{B.30.5}$$

be a Carleson set inside $E \times \mathbb{R}_+$ for each $\epsilon > 0$, then we get a condition called the Weak Geometric Lemma (WGL) which is necessary but not sufficient for uniform rectifiability. (See [Dav–Sem]$_{SI}$ for an example.) If we want to characterize rectifiability properties of sets in terms of the size of the $\beta(x,t)$, then we have to work with stronger quadratic Carleson measure estimates, as in [Jones]$_{Squ}$, [Jones]$_{Rec}$, and [Dav–Sem]$_{SI}$, and this fact is completely analogous to the relevant quadratic estimates for the $\alpha(x,t)$'s for functions in [Dorr], [Jones]$_{Squ}$. When we work with the $b\beta(x,t)$'s we get a different phenomenon, and the analogy with functions breaks down. See [Dav–Sem]$_{UR}$.

The distinction between $\beta(x,t)$ and $b\beta(x,t)$ does not have an obvious counterpart in the context of functions.

The BWGL says that a uniformly rectifiable set looks like a d-plane at most scales and locations. One cannot really do much better than say that the set of exceptions is a Carleson set. There are seemingly stronger properties that uniformly rectifiable sets enjoy, including one modeled on Carleson's corona construction which controls how fast the approximating d-planes spin around. (See [Sem]$_{Rect}$, [Dav–Sem]$_{SI}$, [Dav–Sem]$_{UR}$.)

One can try to detect the rectifiability properties of sets by measuring approximations against other sets of models besides d-planes. This issue arises in connection with various problems in analysis and is taken up in [Dav–Sem]$_{UR}$. Reasonable sets of models include convex sets, connected sets, Lipschitz graphs, minimal surfaces, and sets whose complementary components are convex. There is a much larger variety of reasonable models to consider in geometry than in linear analysis.

B.31. Comments about geometric complexity.

How can one control the geometric complexity of a set in $\mathbb{R}^n$, or of a metric space?

If one has a specific model in mind, like Euclidean space, it is reasonable to ask first for controlled parameterizations. We have seen some limitations to this approach. Even in the best situations in topology – compact smooth manifolds or piecewise-linear manifolds – one always has problems with complexity which stem from the fundamental group (perhaps local fundamental groups).

So we give up on parameterizations, what next then? Uniform rectifiability provides one answer. It captures pretty much all the "smoothness" that a set should have if it were to have a bi-Lipschitz parameterization, capturing this smoothness in a more flexible way that is much easier to detect. Measurements of smoothness like the bilateral weak geometric lemma (BWGL) and the corona decomposition discussed in Section 30 are very suggestive of certain types of constructions, in which one says that the geometry of the set is usually very tame, the number of times that it is bad can be controlled in terms of a Carleson set, and ad hoc arguments in the bad cases give some good control overall. There are plenty of examples where this type of argument works pretty well. (See $[\mathrm{Sem}]_{Rect}$, $[\mathrm{Dav–Sem}]_{SI}$, $[\mathrm{Dav–Sem}]_{UR}$.) Nonetheless this type of approach clearly misses a lot. Uniform rectifiability does not do such a good job of capturing information that is more topological in nature. Perhaps it could be combined with topological information in a more productive way. One can view $[\mathrm{Dav–Sem}]_{ASQ}$, $[\mathrm{Sem}]_{Diff}$, $[\mathrm{Sem}]_{QT}$, $[\mathrm{Sem}]_{Rmks}$ as providing some alternate approaches to controlling the geometric complexity of spaces which are roughly Euclidean but maybe not parameterized well.

This issue of geometric complexity is fundamentally not understood. It is not at all clear even what language to use. In this appendix we have seen some examples, but still the basic perspectives rest on a lot of assumptions.

IV. An introduction to real-variable methods

In this last part we describe some basic techniques from harmonic analysis and we indicate some of the arguments needed for statements mentioned earlier. These real-variable methods reflect the interaction between measure theory and the geometry of snapshots.

B.32. The Maximal function.

Let $(M, d(x,y), \mu)$ be a space of homogeneous type, as in Section 8. For simplicity one can just think about Euclidean spaces with the usual metric and Lebesgue measure. At any rate we ask that $d(x,y)$ be an actual metric, instead of a quasimetric, but this is only for convenience and entails no loss of generality because of (2.2).

Given a locally integrable function f on M, define the *(Hardy-Littlewood) maximal function* f^* of f by

$$f^*(x) = \sup_{B \ni x} \frac{1}{\mu(B)} \int_B |f(y)| \, d\mu(y). \tag{B.32.1}$$

The supremum here is taken over all (open) balls B which contain x. Sometimes people prefer to use the "centered" maximal function, given by

$$\sup_{r>0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y)|\, d\mu(y). \tag{B.32.2}$$

This is no greater than $f^*(x)$, and $f^*(x)$ is bounded by a constant multiple of the centered maximal function, because of the doubling condition on μ. Thus the two are practically equivalent, but the uncentered version is a little nicer for geometric interpretation.

Maximal functions arise very naturally in analysis, for proving theorems about the existence almost everywhere of limits and for controlling singular integral operators. One can appreciate them at a more naive level by thinking about trying to control the snapshots of a function in terms of simpler information (like integrability).

When I first saw the definition of the maximal function as a student I was impressed by its brashness. It seems to be asking a lot to expect that f^* be finite in nontrivial situations, to take the supremum over *all* balls that way. Nonetheless it works out well.

B.32.3. Theorem: *Notation and assumptions as above. If $f \in L^1(\mu)$ then we have the "weak type" inequality*

$$\mu(\{x \in M : f^*(x) > \lambda\}) \leq C\,\lambda^{-1}\,\|f\|_1 \tag{B.32.4}$$

for some constant C (which depends only on the doubling constant for μ) and all $\lambda > 0$. If $f \in L^p(\mu)$ for some $p > 1$, then $f^* \in L^p$ and

$$\|f^*\|_p \leq C(p)\,\|f\|_p, \tag{B.32.5}$$

where C depends only on p and the doubling constant for μ.

This theorem probably looks strange if one is not accustomed to these things, so let us take it slowly. To understand the weak-type inequality it is helpful to look at an example. Actually, let us agree to extend the definition of the maximal function to locally finite measures in the obvious manner. If we do that, and if we take $(M, d(x,y), \mu)$ to be the real line equipped with Lebesgue measure, then the maximal function of a Dirac mass at the origin is the function given by $1/|x|$. This is not integrable near the origin or at infinity. This is why $f \mapsto f^*$ cannot be bounded on L^1, or even take L^1 into itself. Strictly speaking measures are not the same as L^1 functions, but in any concrete situation like this you cannot have good behavior on L^1 while having such bad behavior for measures. This fact can be made precise by building L^1 functions that contain little copies of approximate Dirac masses, and then showing that the maximal function has suitably bad behavior.

However this example is perfectly compatible with the weak-type inequality (32.4), and it shows that the weak-type inequality is optimal in a natural way.

Note that the weak-type inequality would hold automatically if the L^1 norm of f^* were controlled by the L^1 norm of f. The weak-type inequality is what we would get from the L^1 bound using Chebychev's inequality.

The weak-type inequality also works for finite measures, and with the same proof, which we give in the next section.

What about the other L^p spaces? The L^∞ estimate is trivial and not very useful. The L^p estimates are more remarkable for being able to take the supremum in (32.1) inside the integral. For $1 < p < \infty$ it turns out that there is a general "interpolation" theorem due to Marcinkeiwicz which permits one to obtain the L^p bounds from the L^1 and L^∞ estimates. Note that we have less than boundedness at L^1 but we are recapturing boundedness in between. See [Stein]$_{SI}$, [Stein]$_{HA}$, and [Stein–Weiss] for more details. The penalty for the unboundedness at $p = 1$ is that the constants $C(p)$ in (32.5) blow up as $p \to 1$, at a rate like $(p-1)^{-1}$.

This phenomenon occurs repeatedly in analysis, that one has bounds for $1 < p < \infty$ but not at the endpoints. Here $p = \infty$ is trivially okay, but normally it is not, as in the context of singular integral operators. For linear operators the $p = 1$ and $p = \infty$ cases correspond to each other under duality.

For our purposes it is the weak-type inequality for L^1 that matters most, and not so much the L^p estimates which are very important in analysis in general.

An amusing fact: if f is a locally integrable function with $f^* \not\equiv \infty$ and if $0 < \delta < 1$, then $(f^*)^\delta$ is an A_1 weight (Definition 21.1), and one can get all A_1 weights in this manner, modulo multiplication by functions that are bounded and bounded away from 0. This was proved by Coifman and Rochberg. See [Garn], [Jour], [Stein]$_{HA}$.

B.33. Covering lemmas.

Again let $(M, d(x,y), \mu)$ be a space of homogeneous type, with $d(x,y)$ an actual metric for simplicity. Suppose that $f \in L^1(\mu)$, and set

$$E_\lambda = \{x \in M : f^*(x) > \lambda\}, \tag{B.33.1}$$

where $\lambda > 0$ is given. Notice that E_λ is an *open* set. We want to control $\mu(E_\lambda)$.

What do we know about E_λ? If $x \in E_\lambda$, then there is a ball B such that $x \in B$ and

$$\frac{1}{\mu(B)} \int_B |f(y)| d\mu(y) > \lambda. \tag{B.33.2}$$

Thus E_λ is covered by balls of this type. In order to control $\mu(E_\lambda)$ we need a covering lemma, and the following works.

B.33.3. Lemma: *Let $(M, d(x,y))$ be a metric space, and let $\mathcal{B}$ be a collection of balls (open or closed) of bounded radius which covers a set E. Then we may extract from $\mathcal{B}$ a sequence of pairwise disjoint balls $\{B_i\}$, possibly finite, such that either*

$$E \subseteq \bigcup_i 5B_i \tag{B.33.4}$$

or there are infinitely many B_i's and they all have radii uniformly bounded away from 0.

This is well known, and the argument on p. 9-10 of [Stein]$_{SI}$ works, for instance. The idea is easy enough. One chooses B_1 in $\mathcal{B}$ so that its radius is as large as possible, to within a factor of 2, and then one chooses the B_i's recursively so that at each stage the new ball is disjoint from its predecessors and with radius as large as possible, to within a factor of 2. One then checks that either the radii tend to 0, in which case one gets the covering property (33.4), or they remain bounded away from 0. This proves the lemma.

Let us now use this to prove that

$$\mu(E_\lambda) \leq C\,\lambda^{-1} \int_M |f(y)|\, d\mu(y). \tag{B.33.5}$$

To avoid a small technical problem let F be an arbitrary bounded (measurable) subset of E_λ, and let us get an estimate for $\mu(F)$ which does not depend on the choice of F. Let $\mathcal{B}$ be the set of balls which satisfy (33.2), and let $\mathcal{B}_F$ denote the elements of $\mathcal{B}$ which intersect F. We claim that there is a sequence of pairwise disjoint balls $\{B_i\}$ in $\mathcal{B}_F$ such that

$$\mu(F) \leq \sum_i \mu(5B_i). \tag{B.33.6}$$

This basically follows from the lemma, but with a couple of small qualifications. If the balls in $\mathcal{B}_F$ do not have bounded radii, then we just take one very large ball B_1 in $\mathcal{B}_F$, and that will satisfy $F \subseteq 5B_1$. If the balls in $\mathcal{B}_F$ do have bounded radii, then we may apply Lemma 33.3. Either we get the covering property (33.4), in which case we are in business, or there are infinitely many B_i's and their radii are bounded away from 0. In this case the right side of (33.6) is infinite, because of the doubling property for μ, and because all the B_i's intersect the bounded set F by assumption.

Because the B_i's are disjoint and satisfy (33.2) we have that

$$\begin{aligned}\mu(F) \leq \sum_i \mu(5B_i) &\leq C \sum_i \mu(B_i) \\ &\leq C \sum_i \lambda^{-1} \int_{B_i} |f(y)|\, d\mu(y) \qquad \text{(B.33.7)} \\ &\leq C\,\lambda^{-1} \int_M |f(y)|\, d\mu(y).\end{aligned}$$

This constant C depends on the doubling constant for μ and on nothing else. Since F is an arbitrary bounded subset of E_λ we conclude that

$$\mu(E_\lambda) \leq C\,\lambda^{-1} \int_M |f(y)|\, d\mu(y), \qquad \text{(B.33.8)}$$

as desired.

This completes the proof of the weak-type inequality (32.4) in Theorem 32.3.

On the real line there is a very special covering lemma. Given any three intervals which contain a point in common, one of the intervals is contained in the union of the other two. This is easy to check. It implies that if a compact set K in $\mathbb{R}$ is covered by a family of open intervals, then one can cover K by finitely many of these intervals, and with the additional restriction that no point in $\mathbb{R}$ is contained in more than two of the intervals. This fact leads to another development of the weak-type inequality on the real line.

B.34. Lebesgue points.

B.34.1. Theorem: *Let $(M, d(x,y), \mu)$ be a space of homogeneous type, with $d(x,y)$ a metric (rather than a quasimetric) for convenience. If $f \in L^1_{loc}(\mu)$, then*

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|\, d\mu(y) = 0 \qquad \text{(B.34.2)}$$

for μ-almost all $x \in M$.

To prove this we may as well assume that $f \in L^1$ since the matter is entirely local. Set

$$\Lambda(f)(x) = \limsup_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|\, d\mu(y). \qquad \text{(B.34.3)}$$

That is, Λ is a (nonlinear) operator which takes locally integrable functions to measurable functions. We want to show that $\Lambda(f) \equiv 0$ μ-a.e. for all $f \in L^1(\mu)$.

The operator Λ has three important features. The first is that it is sublinear, which means that

$$\Lambda(f_1 + f_2) \leq \Lambda(f_1) + \Lambda(f_2) \tag{B.34.4}$$

for all $f_1, f_2 \in L^1$. This is easy to check. The second is that

$$\Lambda(g) \equiv 0 \tag{B.34.5}$$

when g is continuous, by definitions. The third is that

$$\Lambda(f) \leq 2\, f^*. \tag{B.34.6}$$

Thus we have a bound on the size of $\Lambda(f)$ coming from (32.4).

Every $f \in L^1$ can be approximated by continuous functions. That is, for each $\epsilon > 0$ there is a continuous function g on M such that

$$\int_M |f(z) - g(z)|\, d\mu(z) < \epsilon. \tag{B.34.7}$$

(This uses the Borel regularity of μ.) From the sublinearity property (34.4) and (34.5) we have that

$$\Lambda(f) = \Lambda(f - g). \tag{B.34.8}$$

Using these two facts together with (34.6) and (32.4) we obtain that

$$\{x \in M : \Lambda(f)(x) > t\} \leq C\, t^{-1}\, \epsilon \tag{B.34.9}$$

for some $C > 0$ and all $t > 0$, $\epsilon > 0$. We have applied (32.4) here with f replaced by $f - g$, and it is important that C does not depend on either t or ϵ. Since $\epsilon > 0$ is arbitrary we have that

$$\{x \in M : \Lambda(f)(x) > t\} = 0 \tag{B.34.10}$$

for all $t > 0$, and hence that $\Lambda(f) \equiv 0$ almost everywhere. This proves the theorem.

The proof of this theorem indicates a very general recipe for establishing the existence of limits almost everywhere. The main point is to have a maximal function estimate like (32.4). In the presence of such a bound the existence of limits almost everywhere can be derived from the existence of limits for a dense class, which is typically much easier to get.

B.35. Differentiability almost everywhere.

Let us now specialize to the case of $\mathbb{R}^n$ and look at another question of the existence of limits, namely differentiability almost everywhere for Lipschitz functions. We are going to prove Theorem 12.1 using an argument like the one in the previous section.

Roughly speaking, the functions in the previous section correspond to the derivatives of the functions that we shall consider in this section. Note well that we shall use a lot of structure of Euclidean spaces here that does not work on spaces of homogeneous type except in very special situations (like the Heisenberg group).

As in the preceding section we need to have a maximal function to work with. In the present case it is given by

$$N(f)(x) = \sup_{r>0} r^{-1} \sup_{y\in B(x,r)} |f(y) - f(x)|. \tag{B.35.1}$$

This function measures the oscillations of f near x with the same scaling as the derivative. We want to get a bound for this operator, and then use it to establish the existence almost everywhere of the derivative of a Lipschitz function.

If f is Lipschitz, then $N(f)$ is uniformly bounded. This is a bound, but it is not a useful one. In order to get differentiability almost everywhere we need to understand the behavior of limits for a dense class. We would like to use smooth functions for our dense class, but they are not dense in the space of Lipschitz functions with respect to the Lipschitz norm. We have to use a weaker norm. The same issue is implicit in the preceding section; continuous functions are not dense in L^∞ with respect to the L^∞ norm, but they are dense in L^p when $p < \infty$.

The first thing that we need to do here is to get our hands on some L^∞ functions so that we can drop down to L^p spaces. For that matter we have to produce a candidate for the derivative, which is not given to us a priori.

For this we use the notion of weak derivatives. Given two locally integrable functions f and g on $\mathbb{R}^n$, we say that $g = \frac{\partial}{\partial x_j} f$ in the weak sense if

$$\int_{\mathbb{R}^n} f(x)\, \frac{\partial}{\partial x_j}\phi(x)\, dx = -\int_{\mathbb{R}^n} g(x)\, \phi(x)\, dx \tag{B.35.2}$$

for all smooth functions ϕ on $\mathbb{R}^n$ with compact support. This amounts to saying that $g = \frac{\partial}{\partial x_j} f$ "in the sense of distributions". Note that (35.2) holds when f is continuously differentiable and we use its ordinary derivative.

B.35.3. Lemma: *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, then for each $j = 1, 2, \ldots, n$ there is a function $g_j \in L^\infty(\mathbb{R}^n)$ such that $g_j = \frac{\partial}{\partial x_j} f$ in the weak sense.*

This is well known, but let us sketch the argument. Let e_j denote the standard basis vector in $\mathbb{R}^n$ in the direction of x_j, and consider the quantity

$$F_j(x,h) = \frac{f(x + h\, e_j) - f(x)}{h} \tag{B.35.4}$$

for $h \in \mathbb{R}$. We would like to take a limit of this as $h \to 0$. A priori we do not know that this limit exists, only that these functions are uniformly bounded in L^∞. Standard results in functional analysis imply that we can find a sequence $\{h_k\}$ of real numbers which tend to 0 such that $F_j(x, h_k)$ converges to a function $g_j \in L^\infty$, where the convergence is in the sense of the *weak** topology on L^∞ with respect to the duality with L^1. As a practical matter this means that

$$\lim_{k\to\infty} \int_{\mathbb{R}^n} F_j(x,h_k)\,\phi(x)\,dx = \int_{\mathbb{R}^n} g_j(x)\,\phi(x)\,dx \tag{B.35.5}$$

for all $\phi \in L^1$. The key point behind this existence result is to realize the limiting object initially as an element of the dual space of L^1, and not directly as an L^∞ function. One first obtains the existence of the limit on the left side of (35.5) for a countable dense collection of ϕ's, using a diagonalization argument to choose the sequence $\{h_k\}$ correctly, and then one shows that the limit exists as an element of the dual space, etc.

Once one has g_j as in (35.5), it is easy to check that $g_j = \frac{\partial}{\partial x_j} f$ in the weak sense. This proves the lemma.

The derivative in the weak sense (35.2) is unique if it exists. This is amusing in the context of the preceding proof, for which it implies that the *weak** limit g_j does not really depend on the choice of $\{h_k\}$. In fact one can conclude that $\lim_{h\to 0} F_j(x,h) = g_j(x)$, where again the convergence is taken in the *weak** topology.

How exactly can we use this notion of weak derivative in concrete terms? The answer lies in regularization, the process by which we take a function on $\mathbb{R}^n$ which is not smooth and approximate it by a smooth function. Regularization through convolution, as follows. Let θ be a smooth function on $\mathbb{R}^n$ which vanishes outside the unit ball and which satisfies $\int_{\mathbb{R}^n} \theta(x)\,dx = 1$. Set $\theta_t(x) = t^{-n}\,\theta(x/t)$, so that θ_t is supported in $B(0,t)$ and also has $\int_{\mathbb{R}^n} \theta_t(x)\,dx = 1$. Convolution of f with θ_t is given by

$$f * \theta_t(x) = \int_{\mathbb{R}^n} f(y)\,\theta_t(x-y)\,dy. \tag{B.35.6}$$

Thus $f * \theta_t(x)$ is really an average of values of f near x, "near" meaning points in $B(x,t)$.

The main points now are these. We have that

$$\lim_{t \to 0} f * \theta_t = f \tag{B.35.7}$$

uniformly when f is uniformly continuous, in particular when f is Lipschitz. This is easy to verify directly. If $h(x) \in L^p_{loc}$, $1 \leq p < \infty$, then

$$\lim_{t \to 0} h * \theta_t = h \tag{B.35.8}$$

in L^p_{loc}. This is a standard fact and also not difficult to prove. The version for L^p_{loc} follows easily from the version for L^p because the problem is local. For L^p we use the fact that continuous functions with compact support are dense to reduce to (35.7). This requires also the simple fact that $h \mapsto h * \theta_t$ is a uniformly bounded family of linear operators on each L^p space.

For any locally integrable function f we have that $f * \theta_t$ is smooth. This follows from (35.6) and the smoothness of θ. Suppose we know also that $\frac{\partial}{\partial x_j} f$ exists as a locally integrable function in the weak sense described above. Then

$$\frac{\partial}{\partial x_j}(f * \theta_t) = \left(\frac{\partial}{\partial x_j} f\right) * \theta_t. \tag{B.35.9}$$

This is a nice exercise, using (35.6) and (35.2).

Let us summarize our conclusions in the following lemma.

B.35.10. Lemma: *Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz. Then there is a family of smooth functions $f_t : \mathbb{R}^n \to \mathbb{R}$, $t > 0$, such that $f_t \to f$ uniformly as $t \to 0$ and $\frac{\partial}{\partial x_j} f_t \to \frac{\partial}{\partial x_j} f$ in $L^p_{loc}(\mathbb{R}^n)$ for each $p < \infty$. Here $\frac{\partial}{\partial x_j} f$ is taken in the weak sense (35.2). If f has compact support, then we have convergence in L^p and not just L^p_{loc}.*

For the convergence of the derivative we combine (35.9) with (35.8).

This lemma tells us that we can approximate Lipschitz functions by smooth functions in a nice way. (We could easily derive additional information, such as the uniform boundedness of $\frac{\partial}{\partial x_j} f_t$.) Next we want a maximal function estimate.

B.35.11. Theorem: *Suppose that g is a smooth function on $\mathbb{R}^n$ with $\nabla g \in L^p(\mathbb{R}^n)$ for some $n < p \leq \infty$. Then*

$$\|N(g)\|_p \leq C(p, n)\, \|\nabla g\|_p, \tag{B.35.12}$$

where $C(p)$ does not depend on g.

Remember that $N(\cdot)$ is defined in (35.1).

The restriction $p > n$ is sharp when $n > 1$, and it reflects the famous Sobolev embedding theorem. This point is more prominent in the following lemma.

B.35.13. Lemma: *Let B be an open ball in $\mathbb{R}^n$ and let g be a smooth function on B. Then*

$$|g(z) - g_B| \le C \int_B \frac{1}{|u-z|^{n-1}} |\nabla g(u)|\, du \tag{B.35.14}$$

for all $z \in B$, where g_B denotes the average of g over B. This constant C depends only on the dimension n.

The proof will show that g is integrable on B as soon as the right side is finite.

When $n = 1$ one can reduce (35.14) to the fundamental theorem of calculus using Fubini's theorem, and when $n > 1$ we do the same except we must average over paths in all directions. Specifically, if $w \in B$, then

$$|g(z) - g(w)| \le \int_0^1 |\nabla g((1-s)z + sw)|\, ds, \tag{B.35.15}$$

by the fundamental theorem of calculus. To get (35.14) one simply averages this over w. One can compute that the right side of (35.14) comes out, and we leave this as an exercise. (Reduce to $z = 0$ and then make a change the variables $u = sw$.) This proves the lemma.

Let us convert this inequality into a more directly useful form.

B.35.16. Lemma: *Let B be an open ball in $\mathbb{R}^n$ with radius $r > 0$ and let g be a smooth function on B. If $n < q \le \infty$ ($q = 1$ is allowed when $n = 1$), then*

$$r^{-1} \sup_{u,v \in B} |g(u) - g(v)| \le C(q,n) \left(\frac{1}{|B|} \int_B |\nabla g(w)|^q\, dw \right)^{\frac{1}{q}}. \tag{B.35.17}$$

This is an easy consequence of (35.14) and Hölder's inequality. The main point is that $|x|^{1-n}$ lies in L^q_{loc} when $q > n$ (and not in L^n_{loc} when $n > 1$).

Let us prove now Theorem 35.11. Let $n < q \le p$ be given, with $q = 1$ permitted when $n = 1$, and set $f = |\nabla g|^q$. Then Lemma 35.16 implies that

$$N(g)(x)^q \le C(q,n)\, f^*(x) \tag{B.35.18}$$

for all x. If we are careful to choose $q < p$, then $f \in L^s(\mathbb{R}^n)$ with $s = p/q > 1$, and Theorem 32.3 implies that $f^* \in L^s(\mathbb{R}^n)$. The norm bound (35.12) then follows from (32.5).

Since we did not actually give a proof of (32.5) here let us observe that we can also take $q = p$ and apply (32.4) to obtain that

$$|\{x \in \mathbb{R}^n : N(g)(x)^p > \lambda\}| \leq C(n)\, \lambda^{-1} \int_{\mathbb{R}^n} |\nabla g(y)|^p \, dy. \tag{B.35.19}$$

This weaker inequality is sufficient for the present proof of Theorem 12.1.

The assumption of smoothness of g in Theorem 35.11 is made only for simplicity. It suffices to take a locally integrable function g on $\mathbb{R}^n$ such that $\nabla g \in L^p$ exists in the weak sense. In this case one has to accommodate the technical nuisance that g could behave badly on a set of measure zero and require modification there, but this is not a serious issue. The point is that one can approximate g by smooth functions using convolutions as above, apply Theorem 35.11 to the approximations, and then take limits.

Let us now use Theorem 35.11 to show that Lipschitz functions on $\mathbb{R}^n$ are differentiable almost everywhere. Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz, and assume without loss of generality that f has compact support. Fix any $p > n$. Let $\{f_t\}$ be a family of smooth approximations to f as in Lemma 35.10. We want to use the differentiability of the f_t's to derive the differentiability of f, and for this we need more information about the approximation. Let us check that

$$\lim_{t \to 0} \|N(f - f_t)\|_p = 0. \tag{B.35.20}$$

Morally this follows from Theorem 35.11, but technically we have a small problem since f is not smooth. This is resolved by using Fatou's lemma to show that

$$\|N(f - f_t)\|_p \leq \liminf_{s \to 0} \|N(f_s - f_t)\|_p. \tag{B.35.21}$$

From here (35.20) follows easily.

Define an operator $L(g)$ by

$$L(g)(x) = \lim_{r \to 0} r^{-1} \sup_{y \in B(x,r)} |g(y) - g(x) - \nabla g(x) \cdot (y - x)|. \tag{B.35.22}$$

To say that $L(g)(x) = 0$ means exactly that g is differentiable at x and that its derivative is given by $\nabla g(x)$. We assume here that ∇g exists in the weak sense as a locally integrable function, for instance. We want to show that $L(f) = 0$ almost everywhere. It is easy to see that L is sublinear, so that

$$L(f) \leq L(f - f_t) + L(f_t) \tag{B.35.23}$$

for all $t > 0$. The smoothness of f_t implies that $L(f_t) \equiv 0$ for all $t > 0$. Thus

$$L(f) \leq L(f - f_t). \tag{B.35.24}$$

We also have that

$$L(g) \leq N(g) + |\nabla g| \tag{B.35.25}$$

for all functions g. From (35.20) and the fact that $\nabla f_t \longrightarrow \nabla f$ in L^p norm we conclude that $\|L(f)\|_p = 0$, so that $L(f) = 0$ almost everywhere.

This proves that f is differentiable almost everywhere and that its gradient is given by the gradient in the weak sense obtained before.

There is a lot more to say about functions with locally integrable derivatives in the weak sense (see [Ev–Gar], [Stein]$_{SI}$), but the preceding arguments touch many of the basic techniques.

There are other reasonable ways to approach the differentiability almost everywhere of Lipschitz functions. One argument (going back to Calderón?) relies on the observation that if f is Lipschitz, then for each $p < \infty$ we have that

$$\lim_{r \to 0} r^{-n} \int_{B(x,r)} |\nabla f(y) - \nabla f(x)|^p \, dy = 0 \tag{B.35.26}$$

for almost all $x \in \mathbb{R}^n$. This uses only the fact that $\nabla f \in L^p(\mathbb{R}^n)$, and it can be derived from a Lebesgue point argument much like the one in the proof of Theorem 34.1. From this one can get

$$\lim_{r \to 0} r^{-1} \sup_{y \in B(x,r)} |f(y) - f(x) - \nabla f(x) \cdot (y - x)| = 0 \tag{B.35.27}$$

almost everywhere using (35.17) (generalized from smooth functions to Lipschitz functions). (I learned this method from Pekka Koskela.)

B.36. Finding Lipschitz pieces inside functions.

Let $(M, d(x,y))$ be a metric space and let $f : M \longrightarrow \mathbb{R}$ be a function. Suppose that we are in a situation where f is not Lipschitz but we wish that it were. What to do?

We can begin by defining the function $N(f)$ as in (35.1). This makes sense on any metric space. Let us suppose that we have some kind of control on $N(f)$, so that $N(f) \not\equiv \infty$. We saw in the preceding section how we can sometimes control the size of $N(f)$, in terms of its L^p norm for instance, when we are working on a Euclidean space. There are similar results in other contexts.

Let us assume also that f is continuous, for simplicity.

Let $\lambda > 0$ be given, and consider the set

$$F_\lambda = \{x \in M : N(f)(x) \leq \lambda\}. \tag{B.36.1}$$

It is easy to see that F_λ is closed when f is continuous. We think of F_λ as being the set where f behaves like it is Lipschitz with constant $\leq \lambda$.

Let us assume that $F_\lambda \neq \emptyset$ and be precise about the idea that f behaves like it is λ-Lipschitz on F_λ, even near F_λ. We have that

$$|f(x) - f(y)| \leq \lambda\, d(x, y) \tag{B.36.2}$$

for all $x \in F_\lambda$ and all $y \in M$. This follows easily from the definition of $N(f)$. In particular the restriction of f to F_λ is λ-Lipschitz.

A real-valued Lipschitz function on a subset of a metric space can always be extended to a Lipschitz function on the whole space, without increasing the norm. In this case we wish to use this fact to obtain a λ-Lipschitz function $g_\lambda : M \to \mathbb{R}$ which equals f on F_λ. According to the usual recipe, we can obtain g_λ from the formula

$$g_\lambda(y) = \inf_{x \in F_\lambda} \{f(x) + \lambda\, d(x, y)\}, \qquad y \in M. \tag{B.36.3}$$

This function is λ-Lipschitz, as in Lemma 6.20, and it is easy to check that $g_\lambda = f$ on F_λ.

How well does g_λ approximate f off of F_λ? We have that

$$|g_\lambda(y) - f(y)| \leq 2\,\lambda\, \mathrm{dist}(y, F_\lambda) \qquad \text{for all } y \in M. \tag{B.36.4}$$

To see this let x be any element of F_λ. Then

$$|g_\lambda(y) - f(y)| \leq |g_\lambda(y) - g_\lambda(x)| + |f(y) - f(x)| \tag{B.36.5}$$

since $g_\lambda(x) = f(x)$. Thus

$$|g_\lambda(y) - f(y)| \leq 2\,\lambda\, d(x, y), \tag{B.36.6}$$

since g_λ is λ-Lipschitz and we have (36.2). Taking the infimum over $x \in F_\lambda$ we get (36.4).

Thus g_λ provides a good approximation to f on all of M, with a very precise estimate. In practical situations on Euclidean spaces and some other spaces we have a good bound on the measure of $M \backslash F_\lambda$, and we can use this approximation to derive results about f from results about Lipschitz functions. Here is an example.

B.36.7. Proposition: *Suppose that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuous. Then f is differentiable at almost every point for which $N(f) < \infty$.*

To prove this we let F_λ be defined as in (36.1). It suffices to show that f is differentiable at almost every point in F_λ for each $\lambda > 0$.

Let $g_\lambda : \mathbb{R}^n \longrightarrow \mathbb{R}$ be as above. Then g_λ is Lipschitz, and hence differentiable a.e.

Each F_λ is closed and therefore measurable, and this means that almost every element of F_λ is a point of density of F_λ. That is,

$$\lim_{r \to 0} \frac{|F_\lambda \cap B(x,r)|}{|B(x,r)|} = 1 \tag{B.36.8}$$

for almost all $x \in F_\lambda$. This follows from Theorem 34.1, applied to the characteristic function of F_λ.

Suppose now that $x \in F_\lambda$ is a point of density of F_λ and a point of differentiability of g. Let us show that f is also differentiable at x, and with the same gradient as g. It suffices to show that

$$\lim_{r \to 0} r^{-1} \sup_{y \in B(x,r)} |f(y) - g(y)| = 0. \tag{B.36.9}$$

Because of (36.4) it suffices to show that

$$\lim_{r \to 0} r^{-1} \sup_{y \in B(x,r)} \operatorname{dist}(y, F_\lambda) = 0. \tag{B.36.10}$$

This last is an easy consequence of (36.8). (To see this it is easier to think about showing that the failure of (36.10) implies the failure of (36.8).)

Thus f is differentiable almost everywhere on each F_λ, and the proposition follows.

See [Stein]$_{SI}$ for more information about pointwise differentiability results.

There are many variations and refinements of the techniques indicated in this section, using other types of maximal functions and more refined extension techniques.

B.37. Maximal functions and snapshots.

Let us come back to the idea of snapshots of a function. Let f be a real-valued function in $\mathbb{R}^n$, and let us think about f in terms of Lipschitz or "almost-Lipschitz" properties. Each ball B in $\mathbb{R}^n$ gives a snapshot of our function f, in the manner of (12.2). The Lipschitz condition means exactly that these snapshots are all bounded. The condition that $N(f)(x) < \infty$ means that all of the snapshots of f on the balls $B(x,r)$ are bounded. This implies that the snapshots of f on all balls containing x are then bounded, with a slightly worse constant.

Let $\lambda > 0$ be given, and let F_λ be as in (36.1). Think of $\mathbb{R}^n \times (0,\infty)$ as parameterizing the set of balls in $\mathbb{R}^n$, and hence the set of snapshots of f. Let $\widehat{F}_\lambda$ denote the set of points in $\mathbb{R}^n \times (0,\infty)$ which correspond to balls which contain an element of F_λ. On these balls we control the size of the snapshot of f in terms of λ.

One can picture $\widehat{F}_\lambda$ as the union of the cones in $\mathbb{R}^n \times (0,\infty)$ with aperture 1 and vertex in F_λ. If F_λ is large, then $\widehat{F}_\lambda$ will be a large piece of $\mathbb{R}^n \times (0,\infty)$.

In the story of g_λ we are trying to build a function on $\mathbb{R}^n$ whose snapshots associated to elements of $\widehat{F}_\lambda$ look like the snapshots of f but whose behavior is good everywhere. This is an important point; we want to match the behavior of f at all scales, not just the values of f at individual points.

These principles make sense in greater generality. Maximal functions control the behavior of functions for all snapshots over balls containing a given point. When the maximal function is controlled on a set, the snapshots are controlled for a subset of $\mathbb{R}^n \times (0,\infty)$ which is a union of cones. We can try to build a function which has approximately the same snapshots as the original function in the good region but which has better behavior in the bad region.

To understand this properly one should understand the Whitney extension theorem, Whitney cubes and their associated partitions of unity, as in [Stein]$_{SI}$. They provide the tools for building the approximations that one needs on "the bad set" (e.g., the complement of F_λ) with much more care and control. We shall not pursue this here.

We shall discuss however the "Calderón-Zygmund approximation" in Section 39, which provides a version of this idea corresponding to the size of functions rather than there smoothness. Before we do this we take up the preliminary topic of dyadic cubes.

B.38. Dyadic cubes.

An interval I in the real line is said to be *dyadic* if it is of the form $[j\,2^k, (j+1)\,2^k)$, where j and k are integers. The funny business with the end points is made to ensure that the dyadic intervals of a fixed length provide a partition of $\mathbb{R}$. A cube Q in $\mathbb{R}^n$ is said to be *dyadic* if it is a product of dyadic intervals of the same length. Thus $\mathbb{R}^n$ is partitioned by the cubes of sidelength 2^k for each integer k.

Dyadic cubes enjoy the following combinatorial property.

B.38.1. Lemma: *If Q and R are dyadic cubes in $\mathbb{R}^n$, then either $Q \subseteq R$, $R \subseteq Q$, or $Q \cap R = \emptyset$.*

This is not very difficult to check. It is very useful because it permits us to avoid covering lemmas when working with dyadic cubes, using the following instead.

B.38.2. Lemma: *Given any collection $\mathcal{C}$ of dyadic cubes in $\mathbb{R}^n$, the maximal elements of $\mathcal{C}$ are pairwise disjoint.*

This is not to say that maximal elements always exist. They will under mild conditions, like a uniform bound on the size of the cubes in $\mathcal{C}$.

Lemma 38.2 is an easy consequence of Lemma 38.1. Let us give an application to illustrate how it can be used. Given a locally integrable function f on $\mathbb{R}^n$, define the *dyadic maximal function of f* by

$$f_\delta^*(x) = \sup_{Q \ni x} \frac{1}{|Q|} \int_Q |f(y)|\, dy, \tag{B.38.3}$$

where the supremum is taken over all dyadic cubes that contain x.

This maximal function is bounded by a constant multiple of the maximal function given in (32.1), but the converse is not quite true. Thus we have the same estimates for f_δ^* as for f^*, but let us take a closer look at (32.4) and give a more direct argument.

Assume that $f \in L^1(\mathbb{R}^n)$, let $\lambda > 0$ be given, and set

$$E_\lambda = \{x \in \mathbb{R}^n : f_\delta^*(x) > \lambda\}, \tag{B.38.4}$$

in analogy with (33.1). We can rewrite this as

$$E_\lambda = \bigcup_{Q \in \mathcal{C}} Q, \qquad \text{where} \tag{B.38.5}$$

$$\mathcal{C} = \left\{ Q \in \Delta : \frac{1}{|Q|} \int_Q |f(y)|\, dy > \lambda \right\}, \tag{B.38.6}$$

and where Δ denotes the collection of all dyadic cubes in $\mathbb{R}^n$. Let $\mathcal{M}$ denote the collection of maximal elements of $\mathcal{C}$. We have that

$$E_\lambda = \bigcup_{Q \in \mathcal{M}} Q, \tag{B.38.7}$$

because the definition of $\mathcal{C}$ and the assumption that $f \in L^1(\mathbb{R}^n)$ ensure that the elements of $\mathcal{C}$ have bounded size, so that each element of $\mathcal{C}$ is contained in a maximal element. Using Lemma 38.2 we obtain the disjointness of the elements of $\mathcal{M}$, and hence

$$\begin{aligned} |E_\lambda| = \sum_{Q \in \mathcal{M}} |Q| &\leq \lambda^{-1} \sum_{Q \in \mathcal{M}} \int_Q |f(y)|\, dy \\ &= \lambda^{-1} \int_{E_\lambda} |f(y)|\, dy \leq \lambda^{-1} \int_{\mathbb{R}^n} |f(y)|\, dy. \quad \text{(B.38.8)} \end{aligned}$$

This gives us the analogue of (32.4) for the dyadic case.

In defining dyadic cubes we have used the special structure of Euclidean spaces, but in fact one can make similar constructions in much greater generality, as in [David]$_{MG}$, [David]$_{WSI}$, [Christ]$_{SIO}$, and [Christ]$_{CI}$. The constructions in [David]$_{MG}$ and [David]$_{WSI}$ are given for Ahlfors regular subsets of Euclidean spaces, but Assouad's embedding theorem (Theorem 20.1) permits them to be applied to general metric spaces.

A lot of analysis with respect to dyadic cubes can be formulated very abstractly through the idea of "martingales" in probability theory, but we shall not go into that here.

B.39. The Calderón-Zygmund approximation.

Let f be an integrable function on $\mathbb{R}^n$, and let $\lambda > 0$ be given. We would like to approximate f by a function which is never much larger than λ, and to do this in a way which respects the geometry of $\mathbb{R}^n$.

Let f_δ^* denote the dyadic maximal function of f, as in (38.3). Let E_λ be as in (38.4), and set $F_\lambda = \mathbb{R}^n \backslash E_\lambda$. Let $\mathcal{C}$ be as in (38.6), and let $\mathcal{M}$ denote again the maximal elements of $\mathcal{C}$, so that (38.7) holds. Let us check that

$$\frac{1}{|Q|} \int_Q |f(y)|\, dy \leq 2^n \lambda \quad \text{when } Q \in \mathcal{M}. \qquad \text{(B.39.1)}$$

Let $Q \in \mathcal{M}$ be given, and let T be the parent of Q, the (unique) dyadic cube in $\mathbb{R}^n$ which contains Q and has sidelength twice that of Q. Then

$$\frac{1}{|T|} \int_T |f(y)|\, dy \leq \lambda \qquad \text{(B.39.2)}$$

because Q is a maximal element of $\mathcal{C}$ and therefore T does not lie in $\mathcal{C}$. This implies (39.1), by an elementary calculation.

Define an approximation g to f by

$$\begin{aligned} g(x) &= f(x) \quad \text{when } x \in F_\lambda \\ &= \frac{1}{|Q|} \int_Q |f(y)|\, dy \quad \text{when } x \in Q, Q \in \mathcal{M}. \end{aligned} \qquad \text{(B.39.3)}$$

This is the Calderón-Zygmund approximation to f (for this choice of λ).

We have that

$$|g(x)| \leq 2^n \lambda \quad \text{almost everywhere on } \mathbb{R}^n. \tag{B.39.4}$$

This follows from (39.1) when $x \notin F_\lambda$. When $x \in F_\lambda$ we use the fact that $|f(x)| \leq f_\delta^*(x)$ almost everywhere. This last is easy to check for Lebesgue points, for instance.

Note that $g = f$ except on E_λ, and that the measure of E_λ is controlled by (38.8). In fact we have that

$$\int_Q g(y)\, dy = \int_Q f(y)\, dy \tag{B.39.5}$$

whenever Q is a dyadic cube which intersects F_λ. Actually (39.5) holds trivially for $Q \in \mathcal{M}$, which is not quite contained in the preceding assertion. We want to derive the general case from this one. Let Q be given, with $Q \cap F_\lambda \neq \emptyset$. Observe that $Q \cap E_\lambda$ is the disjoint union of the elements of $\mathcal{M}$ which are contained in Q. This is easy to check, as in Lemmas 38.1 and 38.2. This implies that

$$\int_{Q \cap E_\lambda} g(y)\, dy = \int_{Q \cap E_\lambda} f(y)\, dy, \tag{B.39.6}$$

since we already know that (39.5) holds for the elements of $\mathcal{M}$. This yields (39.5) for Q, since f and g agree on the complement of E_λ.

We can think of (39.5) as saying that g approximates f well in terms of snapshots and not just point values, but snapshots in terms of dyadic cubes instead of balls.

We shall give applications of the Calderón-Zygmund approximation in the next two sections. See [Stein]$_{SI}$, [Stein]$_{HA}$ for more information and other applications.

The constructions given here and in Section 36 are particularly close for the real line, where we can take a locally integrable function and integrate it to get a differentiable function. This case is captured better by the "rising sun lemma" (§5.5 on p.24 of [Stein]$_{SI}$).

B.40. The John-Nirenberg theorem.

Remember from Section 11 that the John-Nirenberg theorem provides higher integrability properties of BMO functions than are offered initially by the definition. In this section we sketch a proof of the theorem, which is based on the Calderón-Zygmund approximation.

Let us first put the *BMO* condition into a more convenient form. Let f be a locally integrable function on $\mathbb{R}^n$. Given a cube Q in $\mathbb{R}^n$, set

$$m_Q f = \frac{1}{|Q|} \int_Q f(y)\, dy, \tag{B.40.1}$$

the average of f over Q. To say that f lies in *BMO* means that

$$\|f\|_{**} = \sup_Q \frac{1}{|Q|} \int_Q |f(y) - m_Q f|\, dy < \infty, \tag{B.40.2}$$

where the supremum is taken over all cubes Q. It is not hard to check that this is equivalent to (11.1), and that the two seminorms are each bounded by constant multiples of the other.

We shall use the following formulation of the John-Nirenberg theorem. Suppose that $\|f\|_{**} \le 1$. Then for any cube Q in $\mathbb{R}^n$ and any integer $j \ge 1$ we have that

$$|\{x \in Q : |f(x) - m_Q f| > j\, 2^{n+1}\}| \le 2^{-j}\, |Q|. \tag{B.40.3}$$

The point is to get exponential decay on the right hand side.

To prove this we apply the Calderón-Zygmund approximation from the preceding section. Let Q and j be given, and let us assume that Q is a dyadic cube. We can always reduce to that case by making a translation and dilation (which cause no trouble). Set

$$f_0(x) = (f(x) - m_Q f)\, \mathbf{1}_Q(x). \tag{B.40.4}$$

Thus $f_0 \in L^1(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} |f_0(x)|\, dx \le |Q|$.

We apply the construction of the preceding section with f replaced by f_0 and with $\lambda = 2$. If $\mathcal{C}$ is as in (38.6), then Q is not an element of $\mathcal{C}$, and in fact all elements of $\mathcal{C}$ are proper subcubes of Q. If $E_\lambda = E_2$ and g are as before, then we have that

$$|E_2| \le \frac{1}{2}\, |Q| \tag{B.40.5}$$

$$|g| \le 2^{n+1} \quad \text{a.e.} \tag{B.40.6}$$

If $j = 1$ then we get (40.3), because E_2 contains the set in the left side of (40.3). What if $j = 2$? In this case we repeat the argument on each cube $T \in \mathcal{M}$, where $\mathcal{M}$ denotes the collection of maximal elements of $\mathcal{C}$. For each $T \in \mathcal{M}$ define a new function f_T by

$$f_T(x) = (f(x) - m_T f)\, \mathbf{1}_T(x). \tag{B.40.7}$$

For this function we have essentially the same information that we did for f_0 on Q. We apply again the construction of the preceding section with $f = f_T$ and $\lambda = 2$ to obtain a set $E_2(T)$ and an approximation g_T to f_T such that

$$|E_2(T)| \leq \frac{1}{2}|T| \tag{B.40.8}$$

$$|g_T| \leq 2^{n+1} \quad \text{a.e.} \tag{B.40.9}$$

By construction we have that $f_T = g_T$ off of $E_2(T)$, and this implies that

$$|\{x \in T : |f(x) - m_T f| > 2^{n+1}\}| \leq \frac{1}{2}|T| \tag{B.40.10}$$

for each $T \in \mathcal{M}$.

Notice that

$$|m_Q f - m_T f| \leq 2^{n+1} \qquad \text{when } T \in \mathcal{M}. \tag{B.40.11}$$

This is implicit in (40.6) with our definitions (and comes from (39.1)). Thus we obtain

$$|\{x \in T : |f(x) - m_Q f| > 2^{n+1} \cdot 2\}| \leq \frac{1}{2}|T| \tag{B.40.12}$$

for each $T \in \mathcal{M}$, using (40.10). Now $X_2 = \{x \in Q : |f(x) - m_Q f| > 2 \cdot 2^{n+1}\}$ is necessarily contained in $E(2)$, and hence in the union of the T's, and we have that

$$|X_2| \leq \sum_{T \in \mathcal{M}} |X_2 \cap T| \leq \sum_{T \in \mathcal{M}} \frac{1}{2}|T|, \tag{B.40.13}$$

by (40.12). Since the T's are disjoint and their union is $E(2)$ (as in (38.7)) we have that

$$|X_2| \leq \frac{1}{2}|E(2)| \leq \frac{1}{4}|Q|. \tag{B.40.14}$$

This is exactly what we wanted. The point is that we won a second factor of 1/2.

Now for each T we have a family of cubes $\mathcal{M}(T)$ with the same properties as before, and in particular the average of $|f_T|$ over any of the elements of $\mathcal{M}(T)$ is bounded by 2^{n+1}. This means that we can repeat the argument for each of them. Repeating the argument indefinitely gives (40.3) for each

j, as desired. This completes our sketch of the proof of the John-Nirenberg theorem.

B.41. Reverse Hölder inequalities.

We shall sketch now the proofs of some results discussed in Sections 17 and 18.

Let ω be a positive locally integrable function on $\mathbb{R}^n$. We shall write

$$\omega(A) \quad \text{for} \quad \int_A \omega(x)\,dx \tag{B.41.1}$$

when A is a measurable subset of $\mathbb{R}^n$, and we shall continue to write $|A|$ for the Lebesgue measure of A. Let us assume that there exists an $\epsilon > 0$ so that for each cube Q and each measurable subset E of Q we have that

$$\frac{|E|}{|Q|} \le \epsilon \quad \text{implies} \quad \frac{\omega(E)}{\omega(Q)} \le 1 - \epsilon. \tag{B.41.2}$$

This is one of the forms of the A_∞ condition mentioned in Section 17, and in fact it is the weakest version. For the convenience of our proofs we shall be working with cubes instead of balls in this section though.

We would like to sketch here a proof of the fact that this assumption on ω implies the existence of $K, p > 1$ such that $\omega \in L^p_{loc}$ and

$$\left(\frac{1}{|Q|}\int_Q \omega(x)^p\,dx\right)^{\frac{1}{p}} \le K\,\frac{1}{|Q|}\int_Q \omega(x)\,dx \tag{B.41.3}$$

for all cubes Q in $\mathbb{R}^n$. Here K and p depend only on ϵ and n.

In order to prove (41.3) we shall show that there is a large constant $L > 1$ so that if Q is any cube and

$$Z_j = \left\{x \in Q : \omega(x) > L^j\,\frac{\omega(Q)}{|Q|}\right\}, \qquad \text{then} \tag{B.41.4}$$

$$\omega(Z_j) \le (1-\epsilon)^j\,\omega(Q), \tag{B.41.5}$$

$j = 1, 2, 3, \ldots$. Actually we can take $L = 2^n/\epsilon$. It is not too difficult to show that this implies (41.3), and that (41.3) also implies (41.5) with a suitable choice of L.

To prove (41.5) we shall use an argument much like the one in the previous section for (40.3). Let a cube Q be given, which we may assume to be dyadic. Set

$$\lambda = 2^{-n}\,L\,\frac{\omega(Q)}{|Q|}, \tag{B.41.6}$$

and let us apply the Calderón-Zygmund approximation (Section 39) to $f(x) = \omega(x)\,\mathbf{1}_Q(x)$. This gives a set E_λ and an approximation g of f with the usual properties. First

$$|E_\lambda| \le \lambda^{-1} \int_{\mathbb{R}^n} f(x)\,dx = 2^n\,L^{-1}\,|Q|, \tag{B.41.7}$$

because of (38.8). We also have that $f = g$ on the complement of E_λ, so that

$$\omega(x) \le 2^n\,\lambda = L\,\frac{\omega(Q)}{|Q|} \quad \text{on } Q\backslash E_\lambda. \tag{B.41.8}$$

Let $\mathcal{C}$ again be the collection of dyadic cubes defined by (38.6). These cubes must all be contained in Q in this case, because of our choices of λ and L. If $\mathcal{M}$ denotes the collection of maximal elements of $\mathcal{C}$, then E_λ is the union of the cubes in $\mathcal{M}$, and for each cube $R \in \mathcal{M}$ we have that

$$\frac{1}{|R|}\int_R f(y)\,dy \le 2^n\,\lambda, \tag{B.41.9}$$

as in (39.1). In the present case this converts to

$$\frac{\omega(R)}{|R|} \le L\,\frac{\omega(Q)}{|Q|}. \tag{B.41.10}$$

With $L = 2^n/\epsilon$, as above, (41.7) implies that

$$|E_\lambda| \le \epsilon\,|Q|. \tag{B.41.11}$$

Our assumption (41.2) implies now that

$$\omega(E_\lambda) \le (1-\epsilon)\,\omega(Q). \tag{B.41.12}$$

This gives (41.5) when $j = 1$, because of (41.8).

What about $j = 2$, and larger values of j? We can iterate the procedure, as in the previous section. We can take the cubes in $\mathcal{M}$, apply the Calderón-Zygmund approximation to each of them, and make the same computations as before, to obtain the analogue of (41.12) for each of those cubes, and then (41.5) with $j = 2$ by a computation as in (40.13), (40.14). The Calderón-Zygmund approximations for the cubes in $\mathcal{M}$ provides us with a new family of cubes, then we repeat the process for them to get (41.5) with $j = 3$, etc.

This completes our sketch of the proof of (41.5).

Let us explain how to use this fact for the results of Gehring described in Section 18. Specifically, we would like to show that if ω is a positive locally integrable function on $\mathbb{R}^n$ for which there exist constants $N > 0$ and $0 < s < 1$ such that

$$\frac{1}{|Q|}\int_Q \omega(x)\,dx \le N\left(\frac{1}{|Q|}\int_Q \omega(x)^s\,dx\right)^{\frac{1}{s}} \tag{B.41.13}$$

for all cubes Q, then (41.3) holds for some choices of $K, p > 1$. (Compare with (18.6) and (18.7).) It suffices to show that (41.2) holds for some $\epsilon > 0$ and all cubes Q.

Let $\epsilon > 0$ be small, to be computed in a moment. Suppose to the contrary that there exists a cube Q and a measurable subset E of Q such that

$$\frac{|E|}{|Q|} \le \epsilon \quad \text{and} \quad \frac{\omega(Q\backslash E)}{\omega(Q)} < \epsilon. \tag{B.41.14}$$

We want to get a contradiction with (41.13) by showing that the right-hand side has to be too small. From Hölder's inequality we obtain that

$$\left(\frac{1}{|Q|}\int_{Q\backslash E} \omega(x)^s\,dx\right)^{\frac{1}{s}} \le \frac{1}{|Q|}\int_{Q\backslash E} \omega(x)\,dx < \epsilon\,\frac{\omega(Q)}{|Q|}. \tag{B.41.15}$$

For the other piece, we use Hölder's inequality to obtain

$$\begin{aligned}\left(\frac{1}{|Q|}\int_E \omega(x)^s\,dx\right)^{\frac{1}{s}} &\le \left(\frac{1}{|Q|}\int_E \omega(x)\,dx\right)\cdot\left(\frac{1}{|Q|}\int_E dx\right)^{1-\frac{1}{s}} \quad \text{(B.41.16)}\\ &\le \frac{\omega(Q)}{|Q|}\,\epsilon^{1-\frac{1}{s}}.\end{aligned}$$

Combining this with (41.15) we get a contradiction to (41.13) if ϵ is small enough, depending on N and s. Therefore (41.13) implies (41.2), and hence (41.3), as desired.

B.42. Two useful lemmas.

In this section we record a couple of lemmas which capture information from maximal functions in convenient ways. These lemmas came up in Part III.

B.42.1. Lemma: *Suppose that f is a function in BMO on $\mathbb{R}^n$ (or any space of homogeneous type for that matter). Given a ball B let $m_B f$ denote the average of f over B. Let B_0 be a fixed ball. Then there is a set $E \subseteq B_0$ such that*

$$|B_0\backslash E| \le .01\,|B| \qquad \textit{and} \tag{B.42.2}$$

$$|m_{B_0} f - m_B f| \le C(n)\,\|f\|_* \tag{B.42.3}$$

whenever B is a ball which intersects E and radius $B \leq$ radius B_0. Here $\|f\|_*$ denotes the BMO norm of f, as in (11.1), and $C(n)$ is independent of f, B_0, and B.

That is, most of the averages of f over balls in a given ball B_0 deviate from the average of f over B_0 by a bounded amount. There is nothing special about .01 here, it was simply chosen for definiteness.

The proof of the lemma is quite easy. Given B_0 define the auxiliary function f_0 by

$$f_0(x) = (f(x) - m_{B_0} f)\, \mathbf{1}_{2B_0}(x). \tag{B.42.4}$$

One can check that $\frac{1}{|B_0|}\int_{B_0} |f_0(x)|\, dx$ is bounded by a constant multiple of the BMO norm of f. The lemma follows by applying the estimate (32.4) for the maximal function of f_0 (with a suitable choice of λ).

Of course Theorem 32.3 provides better estimates than the one that we stated, particularly if one combines it with the John-Nirenberg theorem.

There is a nice analogue of Lemma 42.1 for A_∞ weights.

B.42.5. Lemma: *Let ω be an A_∞ weight on $\mathbb{R}^n$. Then there is a constant K such that for each ball B_0 in $\mathbb{R}^n$, there is a set $E \subseteq B_0$ such that (42.2) holds and*

$$K^{-1}\, m_{B_0}\omega \leq m_B\omega \leq K\, m_{B_0}\omega \tag{B.42.6}$$

for all balls B which intersect E and have radius less than the radius of B_0. The constant K depends only on n and the A_∞ constant for ω.

The fact that we can get the upper bound is an easy consequence of (32.4), applied to $f(x) = \omega(x)\, \mathbf{1}_{2B_0}(x)$, and with λ taken to be a large multiple of $m_{B_0}\omega$. The lower bound is more tricky. It can be derived from estimates for the maximal function of $f(x) = \mathbf{1}_{2B_0}(x)$ taken with respect to ω measure rather than Lebesgue measure. We can do this, because Section 32 applies to any space of homogeneous type, $(\mathbb{R}^n, |x-y|, \omega(x)\, dx)$ in particular. We conclude that the lower bound in (42.6) is true for all balls B which touch a set $E' \subseteq B_0$, where $B_0 \backslash E'$ has small measure with respect to ω. The A_∞ property permits us to obtain that $B_0 \backslash E'$ has small Lebesgue measure, although for this assertion we are using more of the theory about A_∞ weights.

Alternatively one can derive Lemma 42.5 from the fact that the logarithm of an A_∞ weight lies in BMO, which permits a reduction to Lemma 42.1.

In any case we shall not give the details. They are not difficult given the theory that exists, but in the present context it is better to just give a flavor and move on.

B.43. Better methods for small oscillations.

There were two main ideas in the preceding sections, maximal functions and Calderón-Zygmund approximations. They provide very powerful tools for managing good behavior that occurs only some of the time.

A basic deficiency of these methods is that they are not as efficient at detecting small oscillations as they are at controlling large oscillations. Indeed, consider the following question. Suppose that f is an L^1 function on $\mathbb{R}^n$, so that

$$\lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy = f(x) \tag{B.43.1}$$

for almost all points $x \in \mathbb{R}^n$. What can we say about this limiting process? The limit exists, but can we get an estimate? We cannot get any kind of uniform control on the rate of convergence, but can we say something about the number of oscillations? For instance, let $\epsilon > 0$ be given, pick a point x, start at $r = 1$, and steadily shrink r to 0. As r is shrinking, ask how many times the quantity

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy \tag{B.43.2}$$

oscillates by at least ϵ. The answer is finite for almost all points, because of the existence of the limit, but can we get an estimate? One can ask analogous questions for the differentiability of Lipschitz functions.

The method of maximal functions does not deal with this question so well. The matter of small oscillations there is hidden in the approximation of L^1 functions by continuous functions, which does not come with quantitative estimates. (Maximal functions and Calderón-Zygmund approximations do however cooperate well with objects that are more sensitive to oscillations, such as singular integral operators, even if they are not so useful in detecting small oscillations directly. See [Stein]$_{SI}$, [Stein]$_{HA}$.)

Carleson's solution of the Corona problem provides a way to deal with this kind of question. Garnett's book [Garn] is the basic reference for this topic. See also Chapter 2 of Part IV of [Dav–Sem]$_{UR}$, which is aimed more at Lipschitz functions and geometry. Carleson's method is not very difficult to understand, but we do not have room for it here. Still it would be nice to give a flavor of it. Suppose that $f \in L^2(\mathbb{R}^n)$. Given any integer k let f_k be defined as follows: for each dyadic cube Q in $\mathbb{R}^n$ of sidelength 2^k we set

$$f_k(x) = \frac{1}{|Q|} \int_Q f(y)\, dy \qquad \text{when } x \in Q. \tag{B.43.3}$$

We have that $f_k \to f$ in L^2 as $k \to -\infty$ and $f_k \to 0$ in L^2 as $k \to \infty$. Thus we can write f as a telescoping series

$$f = \sum_{k=-\infty}^{\infty} (f_k - f_{k+1}). \tag{B.43.4}$$

(Compare with (6.31).) In fact the functions $f_k - f_{k+1}$ are orthogonal to each other. This is easy to check, using the observations that $f_k - f_{k+1}$ is constant on dyadic cubes of size 2^k and has integral zero over dyadic cubes of larger size. Orthogonality implies that

$$\|f\|_2^2 = \sum_{k=-\infty}^{\infty} \|f_k - f_{k+1}\|_2^2. \tag{B.43.5}$$

(Compare with Theorem 6.3 and (12.5).) One can compute from this that differences between averages of f over successive dyadic cubes are often small. The bound is not so strong that it implies immediately the existence of $\lim_{k\to-\infty} f_k(x)$ almost everywhere, but it provides quantitative bounds not given by the existence of the pointwise limits.

In (43.4) we decompose f according to the obvious layering of dyadic cubes according to size. In fact any decomposition leads to orthogonality. In Carleson's Corona construction one chooses a decomposition adapted to the oscillations of the given function through a stopping-time argument.

The question of measuring oscillations arose before in Section 29, where we sought quantitative bounds on the extent to which a Lipschitz function is approximated by affine functions. In that case too there were L^2 estimates at the heart of the matter (as in (29.20)). Once one can account for oscillations in a meaningful way, as in Theorem 29.10, one can try to make geometrical arguments using real-variable methods. Jones' proof of Theorem 28.2 provides a nice example of this. It works roughly as follows. We start with a Lipschitz mapping on the unit cube Q in $\mathbb{R}^n$. Given a dyadic subcube T of Q, consider the question of whether

$$\frac{|f(T)|}{|T|} \tag{B.43.6}$$

is small or not, where "small" means less than some threshold that has to be computed in the proof. Let E denote the union of the cubes for which this ratio is small. Then E is the disjoint union of the maximal cubes with this property, and one can show that the image of E is small under f. This basically means that it can be ignored.

Now look at the dyadic cubes T with the property that f is very well approximated by an affine function on the double of T, say. The degree of approximation should be normalized properly, as in (29.1), and then one has a good approximation for most cubes, in the sense of Carleson sets. The conclusion is that for most points $x \in Q$, the number of cubes T containing x which are bad for approximating f by an affine function is uniformly bounded, except for a set of x's of small measure. Again we can ignore the bad set because of the formulation of Theorem 28.2.

If you have a cube T on which f is sufficiently well approximated by an affine function, and for which (43.6) is not too small, then f itself must be approximately bi-Lipschitz on that cube. The reason is that the affine function that makes the approximation will have to be bi-Lipschitz with a reasonable bound, because otherwise (43.6) would have to be small. (Compare with Sard's theorem.)

If x is a point that has not been excluded so far, then all but a bounded number of the dyadic cubes T containing x will have the good property just mentioned, of approximate bi-Lipschitzness. To complete the proof of Theorem 28.2 one makes a coding argument to sort these points into a bounded number of sets on which f is bi-Lipschitz.

See [Dav–Sem]$_{QR}$ for variations on this theme.

Notice that these arguments of Carleson and Jones treat the geometry of snapshots differently from the method of maximal functions and Calderón-Zygmund approximations.

B.44. Real-variable methods and geometry.

We have discussed real-variable methods mostly in terms of analysis of functions, but they can be employed in geometric problems too. We have already seen some examples of this, in studying the behavior of Lipschitz or quasisymmetric mappings, or in cases where the geometry of a space is governed by a function (like the unit normal vectors on an almost-flat hypersurface, or metric doubling measures). One can also try to work directly with the geometry of sets and spaces. The idea of Calderón-Zygmund approximations can be useful in studying the geometry of sets. This occurs in [David]$_{PC}$, [David]$_{SR}$, [David]$_{WSI}$, for example. Maximal functions have not been as effective as in the context of functions. They are useful in matters of mass distribution, but in questions of smoothness one has the problem that the regularity of a function is reflected in the size of its derivative while the size of a tangent plane has less meaning (unless there is a question of mass distribution). In geometry it often seems better to study measurements of oscillation than of size. One can measure the extent to which a set can be approximated in a ball by a plane or some other standard model, as in [Jones]$_{Squ}$, [Jones]$_{Rec}$, [Dav–Sem]$_{SI}$, [Dav–Sem]$_{UR}$,

$[\mathrm{Sem}]_{Rect}$.

For the purposes of geometry, my favorite real-variable methods come from Carleson's corona construction. It is sensitive to small oscillations while not depending so much on linear structure as other methods from analysis. Its basic language of Carleson sets and stopping-time regions provides good tools for organizing information, for separating "good parts" from "bad parts", and for measuring the amount of bad behavior and the size of the transition region. This basic language is often a good place to start in a problem, an initial framework with broad applicability and a penchant for clearing away the dust so that one can see the main issues more clearly.

At first the corona construction was applied to functions that govern the geometry of a set, such as Riemann mappings and unit normal vectors, as in [GaJ], $[\mathrm{Jones}]_{Squ}$, $[\mathrm{Jones}]_{Rec}$, $[\mathrm{Sem}]_{Rect}$. In $[\text{Dav–Sem}]_{SI}$ and $[\text{Dav–Sem}]_{UR}$ it is applied more directly to sets themselves. See Chapter 3 of Part I of $[\text{Dav–Sem}]_{UR}$ for the general language of the corona method.

The methods from the corona construction are very general but do not normally finish a problem on their own. The strongest real-variable method for actually resolving a problem in geometry that I know is that of David $[\mathrm{David}]_{MG}$. It addresses the question of when a Lipschitz mapping between two spaces admits a large bi-Lipschitz piece, which can provide a way to parameterize part of the domain. It proceeds by analyzing the possible configuration of snapshots, reducing the matter to one scenario that can cause trouble but which can sometimes be prevented by natural assumptions. The method is purely geometric, in that it can function without starting with some information from linear analysis. The principles are quite simple even if the details are more complicated. See also $[\mathrm{David}]_{WSI}$, $[\text{Dav–Sem}]_{UR}$.

One should keep in mind that these real-variable methods are best suited to finding good behavior most of the time, granting the existence of singularities. For some purposes this is a high price to pay, in topology, for instance. Real variable methods provide a certain kind of language with which to address geometric complexity. They are very good for what they do, but they can only do certain things.

Acknowledgements

I am grateful to J. Heinonen and R. Rochberg for their comments and suggestions.

Appendix C

Paul Levy's Isoperimetric Inequality

by M. Gromov

Consider the standard $(n+1)$-dimensional sphere S^{n+1} and a domain $V_0 \subset S^{n+1}$ with smooth boundary. The classical isoperimetric inequality provides the following lower bound for the n-dimensional volume $\mathrm{vol}(\partial V_0)$ of the boundary ∂V_0.

Denote by α the ratio $\mathrm{vol}(V_0)/\mathrm{vol}(S^{n+1})$, where "vol" means the $(n+1)$-dimensional volume, and take a ball $B_\alpha \subset S^{n+1}$ with $\mathrm{vol}(B_\alpha) = \alpha \,\mathrm{vol}(S^{n+1})$. Denote by $s(\alpha)$ the n-dimensional volume of the boundary sphere ∂B_α and denote by $\mathrm{Is}_{n+1}(\alpha)$ the ratio $s(\alpha)/\mathrm{vol}(S^{n+1})$.

Classical Isoperimetric Inequality.

$$\frac{\mathrm{vol}(\partial V_0)}{\mathrm{vol}(S^{n+1})} \geq \mathrm{Is}_{n+1}(\alpha), \qquad \alpha = \frac{\mathrm{vol}(V_0)}{\mathrm{vol}\, S^{n+1}}.$$

Observe that by applying this inequality to domains with diameters converging to zero, we come to the isoperimetric inequality in the Euclidean space $\mathbb{R}^{n+1}$.

Back in 1919, Paul Levy extended the classical isoperimetric inequality to the convex hypersurfaces in $\mathbb{R}^{n+2}$, and he found a striking infinite dimensional application of his result. (See Chapter IV of the third part of his book [Levy]).

We show in this note that Levy's method works for all Riemannian manifolds. As an application, we obtain some estimates for the eigenvalues of the Laplacian on a Riemannian manifold.

1. Isoperimetric inequality. For a compact Riemannian manifold V, we denote by $\mathrm{vol}(V)$ its total volume and by Ricci the Ricci tensor. We set

$$R(V) = \inf_{\tau} \mathrm{Ricci}(\tau, \tau),$$

where τ runs over all unit tangent vectors of V.

Let V be a closed $(n+1)$-dimensional manifold and let $V_0 \subset V$ be a domain with smooth boundary. If $R(V) \geq n = R(S^{n+1})$, then

$$\frac{\mathrm{vol}(\partial V_0)}{\mathrm{vol}(V)} \geq \mathrm{Is}_{n+1}(\alpha) \geq \mathrm{Is}_{n+1}(\alpha), \qquad \alpha = \frac{\mathrm{vol}(V_0)}{\mathrm{vol}(V)}, \tag{$*$}$$

where $\mathrm{Is}_{n+1}(\alpha)$ is the same function as in the classical isoperimetric inequality.

Remarks and corollaries. When $V = S^{n+1}$, the inequality $(*)$ becomes the classical isoperimetric inequality.

Our proof (see §4) shows that equality in $(*)$ holds only for the standard pair $(V, V_0) = (S^{n+1}, B_\alpha)$.

If one applies $(*)$ to the ε-neighborhoods U_ε of V_0 and integrates over ε, one gets

$$\frac{\mathrm{vol}(U_\varepsilon)}{\mathrm{vol}(V_0)} \geq \frac{\mathrm{vol}(A_\varepsilon)}{\mathrm{vol}(B_\alpha)},$$

where A_ε denotes the ε-neighborhood of B_α in S^{n+1}.

2. The first eigenvalue. Denote by λ_1 the first eigenvalue of the Laplacian on V and by D the diameter of V.

If $R(V) \geq -n$, $0 < n = \dim(V) - 1$, then

$$\lambda_1 \geq \exp(-2nD). \tag{$**$}$$

This inequality (in a sharper form) follows from the inequality (Is$'$) of §4 and the following theorem of Cheeger (see [Che]$_{\mathrm{RBLD}}$):

$$\lambda_1 \geq \frac{1}{4}\left(\inf_{V_0} \frac{\mathrm{vol}(\partial V_0)}{\mathrm{vol}(V_0)}\right)^2,$$

where V_0 runs over all domains in V with smooth boundary such that $\mathrm{vol}(V_0)/\mathrm{vol}(V) \leq 1/2$. Some other relations of this type can be found in [Maz] and [Bur–Maz].

The inequality $(**)$ was conjectured by Cheeger, and it was proved in some cases by Yau [Yau]$_{\text{ICFE}}$ and Li [Li], see also [Yau]$_{FTPCRM}$.

I was recently informed that Yau found an independent proof [1] of $(**)$.

We refer to [Ber–Gau–Maz] and [Osser] for further information on the spectrum of the Laplacian. In §5 we generalize this inequality to all eigenvalues λ_i.

3. Levy-Heintze-Karcher comparison theorem. Let $H \subset V$ be a smooth hypersurface with a normal unit vector field ν. Denote by $\exp : H \times \mathbb{R}_+ \to V$ the normal exponential map in the direction ν and denote by $J = J(h,t)$, $h \in H$, $t \in \mathbb{R}_+$, the Jacobian of this map.

Consider also a model pair $(\overline{V}, \overline{H})$, where $\overline{V}$ has constant sectional curvature δ, and $\overline{H}$ is a totally umbilic hypersurface of mean curvature η (relative to the preferred direction $\overline{\nu}$). Denote by $\overline{J}(t) = \overline{J}_{\eta\delta}(t)$ the Jacobian of the corresponding exponential map.

Take a point $h \in H$ such that at this point the mean curvature of H (in the direction ν) is not less than η. Take a point $t \in \mathbb{R}_+$ such that the segment $h \times [0,t]$ is sent by the exponential map to a distance minimizing geodesic segment, i.e., $\operatorname{dist}(H, \exp(h,t)) = t$.

If $R(V) \geq n\delta = R(\overline{V})$, $n = \dim H = \dim \overline{H}$, *then* $|J(h,t)| \leq \overline{J}(t)$.

The proof can be found in [Heint–Kar]. See also [Buy] for some related results.

The explicit formula for $\overline{J}(t)$ is as follows. Set $s_\delta(t) = \delta^{-1/2}\sin(\delta^{-1/2}t)$ for $\delta > 0$, $s_0(\delta) = t$, and $s_\delta(t) = |\delta|^{-1/2}\sinh(|\delta|^{-1/2}t)$ when $\delta < 0$. Using these notations, we have (see [Heint–Kar])

$$\overline{J}_{\eta\delta}(t) = \left(\frac{ds_\delta(t)}{dt} - ns_\delta(t)\right)^n .$$

It is convenient to use another function $\overline{J}^+(t)$ defined as follows: $\overline{J}^+(t) = \overline{J}(t)$ when t is less than the first zero t_0 of $\overline{J}(t)$ and $\overline{J}^+(t) = 0$ for $t \geq t_0$.

Corollary: (See [Heint–Kar]). *Let a domain $V_0 \subset V$ be bounded by a smooth hypersurface $H = \partial V_0$ whose mean curvature is everywhere not less than η (relative to the inward normal). Then*

$$\operatorname{vol}(V_0) \leq \operatorname{vol}(\partial V_0) \int_0^d \overline{J}^+(t)\,dt,$$

where $d = \sup_{v \in V_0} \operatorname{dist}(v, \partial V_0)$. *(Observe that* $d \leq D = \operatorname{diam}(V)$.*)*

[1] See [Li–Yau], where the treatment is, in many respects, more complete and precise than ours.

4. Main inequalities. Let V be a closed $(n+1)$-dimensional manifold with $R(V) = n\delta$, $-\infty < \delta < \infty$, and let $H \subset V$ be an arbitrary smooth hypersurface which divides V into two domains V_0 and V_1 with common boundary H.

There exist positive numbers d_0 and d_1 with $d_0 + d_1 \leq D = \operatorname{diam} V$ and an $\eta \in (-\infty, \infty)$ such that

$$\operatorname{vol}(V_0) \leq \operatorname{vol}(H) \int_0^{d_0} \overline{J}^+_{\eta\delta}(t)\, dt, \tag{Is}$$

$$\operatorname{vol}(V_1) \leq \operatorname{vol}(H) \int_0^{d_1} \overline{J}^+_{\eta\delta}(t)\, dt.$$

Corollaries: *Observe that for $\eta \geq 0$ we have $\overline{J}^+_{\eta\delta}(t) \leq \overline{J}^+_{0\delta}(t)$, and in view of the explicit formula for $\overline{J}$, the inequality* (Is) *implies:*

If $\operatorname{vol}(V_0) \leq \operatorname{vol}(V_1)$, *then*

$$\operatorname{vol}(V_0) < \operatorname{vol}(\partial V_0) \int_0^D \left(\cosh\left(|\delta|^{-1/2} t\right)\right)^n dt. \tag{Is$'$}$$

Notice that this inequality is interesting only when $R(V) \leq 0$. The following inequality is most interesting when $R(V) > 0$.

Set $\alpha = \operatorname{vol}(V_0)/\operatorname{vol}(V)$, $\beta = 1 - \alpha$, and denote

$$\mu(\eta) = \min\left(\alpha^{-1} \int_0^D \overline{J}^+_{\eta\delta}(t)\, dt\ ,\ \beta^{-1} \int_0^D \overline{J}^+_{-\eta\delta}(t)\, dt\right).$$

By combining the inequalities (Is), we immediately get

$$\frac{\operatorname{vol}(H)}{\operatorname{vol}(V)} \geq \inf_{-\infty<\eta<\infty} (\mu(\eta))^{-1}. \tag{Is$''$}$$

When $\delta = 1$ and $D = \infty$, a straightforward calculation shows that

$$\inf_{-\infty<\eta<\infty} (\mu(\eta))^{-1} = \operatorname{Is}_{n+1}(\alpha) = \operatorname{Is}_{n+1}(\beta),$$

and hence, (Is$''$) implies the inequality $(*)$ from §1.

Proof of (Is). Fix an α, $0 < \alpha < 1$, and consider all hypersurfaces in V which divide V into two parts V_0, V_1 with $\operatorname{vol}(V_0) = \alpha$, $\operatorname{vol} V_1 = 1 - \alpha$. Among these hypersurfaces, there is one with minimal n-dimensional volume (see $[\mathrm{Alm}]_{\mathrm{Exi}}$). Suppose for a minute that this minimal hypersurfaces is nonsingular. In such a case, it has *constant* mean curvature. We denote by η the value of this curvature relative to the normal looking into V_0. The curvature in the direction looking into V_1 is η, and applying the Corollary of §3 to V_0 and V_1, we get (Is).

About the singularity. When $\dim V \leq 7$, the minimal hypersurface H dividing V in the given volume proportion is known to be nonsingular (see [Alm]$_{\text{Exi}}$), and the proof of (Is) is complete. In the general case, H may have a singular locus, but it does not affect our proof due to the following

Lemma: *Take a point $v \in V$ and a geodesic segment $\gamma \in V$ having v as one endpoint and $h \in H$ as the other, such that* $\text{length}(\gamma) = \text{dist}(v, H)$. *Then the point $h \in H$ is nonsingular.*

Proof. Take the sphere (relative to the Riemannian metric in V) centered at the center of γ and having radius $\text{dist}(v, H)/2$. This sphere meets H only at h, and it is smooth near this point. It follows that the tangent cone to H at h is contained in a half-space and hence (see [Alm]$_{\text{Exi}}$), the point h is nonsingular.

Estimates for λ_i, $i > 1$. Our method of estimating λ_1 can be extended to other eigenvalues of the Laplacian on V, but we shall establish these estimates by a more elementary method because for the higher λ_i, the more refined Levy's method does not give much better constants.

Notations. Let V be a compact Riemannian manifold of dimension $n+1 \geq 2$. Denote by $N(\varepsilon)$, $\varepsilon > 0$, the minimal number N such that V can be covered by N balls of radius ε. Notice that

$$N(\varepsilon) \geq \frac{D}{2}$$

$$N(D) = 1, \quad D = \text{diam}(V).$$

Denote by $b_{n+1}(\varepsilon)$ the volume of the ε-ball in the $(n+1)$ dimensional hyperbolic space of curvature -1. The Rauch comparison theorem (see [Che–Ebin]) implies that for a closed V with $R(V) \geq -n$, one has

$$N(\varepsilon) \geq (b_{n+1}(\varepsilon))^{-1}\,\text{vol}(V).$$

On the other hand, we shall see below that

$$N(\varepsilon) \leq b_{n+1}(D)(b_{n+1}(\varepsilon/2))^{-1},$$

and, more generally, for $\varepsilon_1 \geq \varepsilon$, one has

$$N(\varepsilon_1) \leq N(\varepsilon) \leq b_{n+1}(2\varepsilon_1 + \varepsilon/2)(b_{n+1}(\varepsilon/2))^{-1} N(\varepsilon_1).$$

Theorem: *There are two positive constants C_1 and C_2 depending on n, such that for any closed Riemannian manifold V of dimension $n+1 \geq 2$ with $R(V) \geq -n$, and for each positive $\varepsilon \leq D = \mathrm{diam}(V)$, one has the following estimates for the eigenvalue λ_i with $i = N(\varepsilon)$*

$$\varepsilon^{-2}C_1^{1+\varepsilon} \geq \lambda_i \geq \varepsilon^{-2}C_2^{1+\varepsilon}.$$

Corollary: Applying these inequalities to $\varepsilon = 1$ and using the estimates for $N(\varepsilon)$ above, we get

(a) if $i \leq D/2$, then $\lambda_i \leq \mathrm{const}_n = C_1^2$,

(a') if $i \leq \mathrm{vol}(V)/b_{n+1}(1)$, then $\lambda_i \leq \mathrm{const}_n = C_1^2$,

(b) if $i \geq b_{n+1}(D)(b_{n+1}(1/2))^{-1}$, then $\lambda_i \geq \mathrm{const}_n = C_2^2$.

Remark: The inequality $\lambda_i \leq \varepsilon^{-2}C_1^{1+\varepsilon}$ and its corollaries (a) and (a′) are not new. A more precise result is due to Cheng (see [Cheng]), but we give the proof here for the sake of completeness.

Proof. We start with several simple lemmas.

(A) (Rauch Theorem): *Take a point $v \in V$ and two balls B, B_1 centered at v of radii r and $r_1 > r$. Let $A \subset B_1$ be a star-convex set, i.e., a Borel set such that for each point $a \in A$, each geodesic segment which joins v with a and has length $= \mathrm{dist}(v,a)$ is contained in A. If* $\mathrm{Ricci}(V) \geq -n$, *then*

$$\frac{\mathrm{vol}(A \cap B_1)}{\mathrm{vol}(A \cap B)} \leq \frac{b_{n+1}(r_1)}{b_{n+1}(r)},$$

where "vol" denotes the $(n+1)$-dimensional measure in V. In particular, one has

$$\frac{\mathrm{vol}(B_1)}{\mathrm{vol}(B)} \leq \frac{b_{n+1}(r_1)}{b_{n+1}(r)}.$$

When $B' \subset V$ is a radius-r ball centered at $v' \in V$ with $\mathrm{dist}(v,v') = d$, we get

$$\frac{\mathrm{vol}(B_1)}{\mathrm{vol}(B')} \leq \frac{b_{n+1}(r_1+d)}{b_{n+1}(r)}.$$

It follows that each ε_1-ball in V can be covered by

$$p \leq b_{n+1}(2\varepsilon_1 + \varepsilon/2)(b_{n+1}(\varepsilon/2))^{-1}$$

balls of radius ε. It implies the relation between $N(\varepsilon)$ and $N(\varepsilon_1)$ which was stated above.

(A′) Denote by $a_n(\varepsilon)$ the n-dimensional volume of the ε-sphere in the $(n+1)$-dimensional hyperbolic space of curvature -1. The n dimensional volume of the set $A \cap \partial B$ and the $(n+1)$-dimensional volume of the part A' of A contained in the complement $B_1 \setminus B$ (i.e., $A' = A \cap (B_1 \setminus B)$) are related as follows:

$$\frac{\mathrm{vol}(A')}{\mathrm{vol}(A \cap \partial B)} \le \frac{b_{n+1}(r_1) - b_{n+1}(r)}{a_n(r)}.$$

Proof. Both theorems (A) and (A′) follow from the local Rauch theorem (see [Che–Ebin]). Observe that they can also be easily reduced to the Levy-Heintze-Karcher comparison theorem.

(B) Let (V, v) be as above and consider an arbitrary Borel set $W \subset V$. Let $H \subset V$ be a smooth hypersurface (possibly with a boundary) such that for each $w \in W$, each distance-minimizing geodesic segment γ between w and v ("distance-minimizing" means $\mathrm{length}(\gamma) = \mathrm{dist}(v, w)$) meets H at a point $h \in H \cap \gamma$. Let $d_1 = d_1(\gamma)$ denote the distance $\mathrm{dist}(v, h)$ and $d_2 = \mathrm{dist}(h, w) = \mathrm{length}(\gamma) - d_1$. (When γ meets H at several points, we take for h the nearest point to w.) The $(n+1)$-dimensional volume of W and the n dimensional volume of H are related by the inequality

$$\frac{\mathrm{vol}(W)}{\mathrm{vol}(H)} \le \sup_{\gamma} \frac{b_{n+1}(d_1(\gamma) + d_2(\gamma)) - b_{n+1}(d_1(\gamma))}{a_n(d_1(\gamma))}.$$

Proof. This immediately follows from (A'). Observe that (A') implies a more precise inequality. Denote by $H' \subset H$ the set of all intersection points $\gamma \cap H$ and for an $h \in H'$ denoted by $\alpha(h)$ the angle between H and the corresponding γ. Set $d_1(h) = \mathrm{dist}(h, v)$, $d_2(h) = \mathrm{dist}(h, w)$. Now one has

$$\mathrm{vol}(W) \le \int_{H'} \left(\frac{b_{n+1}(d_1(h) + d_2(h)) - b_{n+1}(d_1(h))}{a_n(d_1(h))} \right) \sin(\alpha(h))\, d\mu,$$

where $d\mu$ denotes the n-dimensional measure in H' induced from H.

(C) Let V be an arbitrary closed Riemannian manifold and let H be a closed hypersurface dividing V into two parts V_0 and V_1. Let $W_0 \subset V_0$, $W_1 \subset V_1$ be two Borel sets of positive measure. Then there is a point w_0 in one of the sets W_0, W_1, say in W_0, and a subset W in another set, say in W_1, such that each distance minimizing segment joining w_0 with an arbitrary point $w_1 \in W$ meets H at a point h with $\mathrm{dist}(w_0, h) \ge \mathrm{dist}(w_1, h)$, and such that

$$\mathrm{vol}(W) \ge \frac{1}{2} \mathrm{vol}(W_1).$$

Proof. Consider the product $W_0 \times W_1 \subset V \times V$ and the set $X \subset W_0 \times W_1$ of the points (w_0, w_1) which can be joined (in V) by the only one shortest segment γ. This set has full measure in $W_0 \times W_1$. Denote by $Y \subset X$ the set of pairs (w_0, w_1) such that for an intersection point h of H with the corresponding γ, we have $\operatorname{dist}(w_0, h) \geq \operatorname{dist}(w_1, h)$. Denote by $X' \subset W_1 \times W_0$ the set corresponding to X under the natural involution in $V \times V$. Denote by $\tilde{Y} \subset X'$ the set of pairs (w_1, w_0) such that for the corresponding h we have $\operatorname{dist}(w_1, h) \geq \operatorname{dist}(w_0, h)$. One of the sets $Y, \tilde{Y}$ must contain at least one half of the total measure of X (or of X'), and we can assume that this is the case with Y. By the Fubini theorem, there exists a point $w_0 \in W_0$ such that the intersection $w_0 \times W_1 \cap Y$ contains at least one half of the total measure of W_1.

We are now ready to prove a version of the inequality (Is$'$) of §4. We apply (C) to the manifolds V_0, V_1 themselves, and we find a point $w_0 \subset V_0$ and a set $W \subset V_1$ as in (C). By using (B), we get

$$\operatorname{vol}(V_1) \leq 2\operatorname{vol}(W) \leq 2\operatorname{vol}(H)\frac{b_{n+1}(D) - b_{n+1}(D/2)}{a_n(D/2)}.$$

Of course, if W happens to be in V_0, we get such estimates for $\operatorname{vol}(V_0)$, but, in any case, we have an estimate for the smaller of the manifolds V_0 and V_1. According to the Cheeger theorem (see §2), this is sufficient for an estimate of λ_1 from below. In order to get an estimate for the rest of the λ_i, we need one more lemma.

(D) Let V be an arbitrary compact Riemannian manfold, and let $B_1, B_2, \dots, B_i$ be arbitrary Borel sets in V of positive measure. Let $f_1, f_2, \dots, f_i$ be linearly independent continuous functions orthogonal to the constant function. Then there are some constants $c_0, c_1, \dots, c_i$ such that the function $f = c_0 + \sum_{j=1}^{i} c_j f_j$ is not constant and has the following property:

For each $j = 1, \dots, i$, both intersections $B_j \cap f^{-1}((-\infty, 0])$ and $B_j \cap f^{-1}([0, \infty))$ have volume not less than $\operatorname{vol}(B_j)/2$.

Remark: When the f_j are nice smooth functions, the level set $f^{-1}(0)$ has measure zero, and we get the inequalities

$$\operatorname{vol}(B_j \cap f^{-1}((-\infty, 0]) = \operatorname{vol}(B_j \cap f^{-1}([0, \infty)) = \frac{1}{2}\operatorname{vol}(B_j).$$

Proof. The functions f_j map V into $\mathbb{R}^i$. By the Borsuk-Ulam theorem, there is a hyperplane which divides each B_j into two pieces of measure $\geq \operatorname{vol}(B_j)/2$. This hyperplane is defined by a combination $c_0 + \sum_{j=1}^{i} c_j f_j$.

Proof of the inequality $\lambda_i \geq \varepsilon^{-2} C_2^{1+\varepsilon}$. Cover V by $i = N(\varepsilon)$ balls of radius ε. Take the eigenfunctions $f_1, \dots, f_i$ and form $f = c_0 + \sum_{j=1}^{i} c_j f_j$ as above. For each ball B_j, take the concentric ball $\tilde{B}_j$ of radius 2ε and set $g = f^2$. *Then for each $j = 1, 2, \dots, i$, we have*

$$\int_{B_j} g\, dv \leq c(n, \varepsilon) \int_{\tilde{B}_j} \|\operatorname{grad} g\|\, dv, \qquad (i)$$

where $c(n, \varepsilon) = 2(b_{n+1}(2\varepsilon) - b_{n+1}(\varepsilon))/a_n(\varepsilon)$.

Proof. Set $B_j(t) = B_j \cap g^{-1}([t, \infty))$ and $\tilde{B}_j(t) = \tilde{B}_j \cap g^{-1}([t, \infty))$. One obviously has (see [Ber–Gau–Maz], for example)

$$\int_{B_j} g\, dv = \int_0^\infty \operatorname{vol}(B_j(t))\, dt,$$

$$\int_{\tilde{B}_j} \|\operatorname{grad} g\|\, dv = \int_0^\infty \operatorname{vol}(\partial \tilde{B}_j(t))\, dt,$$

where in the first case, "vol" denotes the $(n+1)$-dimensional volume in V, and in the second formula, we use the n-dimensional volume of the hypersurface $\partial \tilde{B}_j(t)$. In order to prove (i), we must only show that for each noncritical value $t \in (0, \infty)$, one has

$$\operatorname{vol}(B_j(t)) \leq c(n, \varepsilon) \operatorname{vol}(\tilde{B}_j(t)).$$

Denote by $B^+ = B_j^+(t) \subset B_j(t)$ the set where $f \geq 0$ and denote by $B^- \subset B_j(t)$ the set where $f \leq 0$. When $t > 0$, these sets are disjoint, and their volumes do not exceed $\operatorname{vol}(B_j(t))/2$.

The hypersurface $\partial \tilde{B}_j(t)$ is also divided according to the sign of f into two parts which we denote by H^+ and H^- correspondingly. Let us show that

$$\operatorname{vol}(B^+) \leq c(n, \varepsilon) \operatorname{vol}(H^+).$$

Set $W_0 = B^+$ and $W_1 = B_j(t) \setminus B^+$. For any two points $w_0 \in W_0$, $w_1 \in W_1$, the shortest geodesic segment between these points has length at most 2ε, and it meets H^+ at some point. By applying (C) to W_0, W_1 and to the hypersurface $f^{-1}(\sqrt{t}) \supset H^+$, and by applying (B) to H^+, we get our inequality.

The same argument shows that

$$\operatorname{vol}(B^-) \leq c(n, \varepsilon) \operatorname{vol}(H^-),$$

and by adding the two inequalities, we obtain our estimate for $\operatorname{vol}(B_j(t))$ and hence for $\int_{B_j} g\, dv$.

Denote by M the maximal multiplicity of the covering $\{\tilde{B}_j\}_{1,2,\dots,i}$. This number M satisfies

$$M \leq M_n(\varepsilon) = b_{n+1}\left(\frac{13}{2}\varepsilon\right)(b_{n+1}(\varepsilon/2))^{-1}.$$

Really, take a point $v \in V$ which is contained in M balls $\tilde{B}_j$. The ball $B_{2\varepsilon}$ of radius 2ε around v contains M centers of the balls B_j. But the ball $B_{3\varepsilon}$ around v can be covered by $p \leq M_n(\varepsilon)$ balls of radius ε, and the inequality $M > p$ would contradict the minimality of the covering $\{B_j\}$. (The definition of $N(\varepsilon)$ implies that V cannot be covered by $N(\varepsilon) - 1$ balls of radius ε.) Now we have

$$\begin{aligned}\int_V g\,dv &\leq \sum_{j=1}^{i}\int_{B_j} g\,dv \\ &\leq c(n,\varepsilon)\sum_{j=1}^{i}\int_{\tilde{B}_j}\|\operatorname{grad} g\|\,dv \\ &\leq M_n(\varepsilon)c(n,\varepsilon)\int_V \|\operatorname{grad} g\|\,dv.\end{aligned}$$

We now recall that $g = f^2$, and hence

$$\int_V \|\operatorname{grad} g\|\,dv \leq 2\left(\left(\int\|\operatorname{grad} f\|^2\,dv\right)\left(\int f^2\,dv\right)\right)^{1/2}.$$

So we have

$$\int_V f^2\,dv \leq E\int_V \|\operatorname{grad} f\|^2\,dv,$$

where $E = 4(M_n(\varepsilon)c(n,\varepsilon))^2$.

The function f was constructed as a combination $f = c_0 + \sum_{j=1}^{i} c_j f_j$, where one of the c_j, $j > 0$, is different from zero and where the eigenfunctions f_i can be assumed to be orthonormal. So, we have

$$\int_V f^2\,dv = \sum_{j=0}^{i} c_j^2,$$

$$\int_V \|\operatorname{grad} f\|^2\,dv = \sum_{j=1}^{i}\lambda_j c_j^2,$$

and so

$$\sum_{j=0}^{i} c_j^2 \leq E\sum_{j=1}^{i}\lambda_j c_j^2.$$

Because one of the c_j, $j > 0$, is nonzero, one of the numbers λ_j must be at least E^{-1}, and hence, the largest eigenvalue λ_i satisfies

$$\lambda_i \geq E^{-1}.$$

We recall that $E = 4(M_n(\varepsilon)c(n,\varepsilon))^2$, where

$$M_n(\varepsilon) = b_{n+1}\left(\frac{13}{2}\varepsilon\right)(b_{n+1}(\varepsilon/2))^1,$$

$$c(n,\varepsilon) = 2\frac{b_{n+1}(2\varepsilon) - b_{n+1}(\varepsilon)}{a_n(\varepsilon)},$$

and $b_{n+1}(\varepsilon)$ denotes the volume of the $(n+1)$-dimensional hyperbolic ball and $a_n(\varepsilon)$ denotes the volume of its boundary. So, for a sufficiently small $C_2 = C_2(n)$, we have

$$\lambda_i \geq \varepsilon^{-2} C_2^{1+\varepsilon}.$$

Proof of the inequality $\lambda_i \leq \varepsilon^{-2}C_1^{1+\varepsilon}$. Let V be any manifold. In order to show that $\lambda_i \leq \lambda$, it is sufficient, according to the variational principle, to find $2i$ nonzero functions $\psi_1, \ldots, \psi_{2i}$ on V with the following properties.

(a) The supports of the ψ_i are pairwise disjoint.

(b) $\int_V \psi_j^2\, dv \geq \lambda \int \|\operatorname{grad}\psi_j\|^2$, $j = 1, \ldots, 2i$.

When $i = N(\varepsilon)$, one can always find $2i$ disjoint balls B_j in V of radius $\varepsilon/5$. With each ball B_j with center $v_j \in B_j$, we associate a function ψ_i defined as follows.

$$\psi_j = \psi_j(v) = \{\operatorname{dist}(v_j, v) - 1/5, v \in B_j, 0, v \notin B_j.\}$$

One obviously has

$$\int_V \|\operatorname{grad}\psi_j\|^2\, dv = \operatorname{vol}(B_j).$$

Denote by $\tilde{B}_j \subset B_j$ the concentric ball of radius $\varepsilon/10$. One has

$$\int_V \psi_j^2\, dv \geq \frac{\varepsilon^2}{100}\operatorname{vol}\tilde{B}_j.$$

When $R(V) \geq -n$, we know that $\operatorname{vol}(B_j)/\operatorname{vol}(\tilde{B}_j) \leq b_{n+1}(\varepsilon/5)/b_{n+1}(\varepsilon/10)$, and so we get

$$\lambda_i \leq 100\varepsilon^{-2}\frac{b_{n+1})(\varepsilon/5)}{b_{n+1}(\varepsilon/10)} \leq \varepsilon^{-2}C_1^{1+\varepsilon}.$$

Notice that the argument we used above is close to the considerations of Cheeger (see $[\text{Che}]_{\text{RBLD}}$) and Yau (see $[\text{Yau}]_{\text{ICFE}}$).

Final remarks. Our lower extimate for λ_i can also be stated as follows.

There exists a constant $C = C_n$ such that for any closed $(n+1)$ dimensional manifold V with $R(V) \geq -n\delta$, $\delta \geq 0$, and $\operatorname{diam} V \leq D$, the i-th eigenvalue λ_i satisfies

$$\lambda_i \geq D^{-2} C^{1+D\sqrt{\delta}}\, i^{2/(n+1)}.$$

On the other hand, the asymptotics of λ_i when $i \to \infty$ depends only on the volume $\operatorname{vol}(V)$. It would be interesting to find a theorem interpolating between these two facts. The following remark indicates one possibility.

Let V be as above and suppose that the fundamental group of V is infinite and has no torsion. Denote by ℓ the length of the shortest noncontractible loop. Then one has

If $i \geq (D/\ell)^n \exp(2nD + 2n)$, then $\lambda_i \geq (D^n\ell)^{-2/(n+1)} C_n^{1+D\sqrt{\delta}} i^{2/(n+1)}$.

Proof. We must estimate from above the number $N(\varepsilon)$, $\varepsilon \leq \ell$. Take the universal covering $\pi : \tilde{V} \to V$ and a point $\tilde{v} \in \tilde{V}$ such that the shortest loop in V passes through $\pi(\tilde{v}) \in \tilde{V}$. Consider the ball $\tilde{B} \subset \tilde{V}$ of radius $2D$ centered at $\tilde{v}$. One can easily see that for each $v \in V$, the set $f^{-1}(v) \cap \tilde{B}$ contains at least $q \geq D\ell^{-1}$ points, and for any two distinct points $\tilde{v}_1, \tilde{v}_2 \in f^{-1}(v)$, one has $\operatorname{dist}(\tilde{v}_1, \tilde{v}_2) \geq \ell$.

When $R(V) \geq -n$, each ε-neighborhood of $\tilde{B}$ contains at most $N \leq b_{n+1}(2D+\varepsilon)/b_{n+1}(\varepsilon)$ disjoint balls of radius ε, and when $\varepsilon < \ell$, we conclude that V itself has at most N/ℓ such balls. It follows that V can be covered by $N_1 \leq N/\ell$ balls of radius 2ε. This is exactly the estimate we need for the case $\delta = 1$, and by scaling the metric in V by $\sqrt{\delta}$, we reduce the general case to $\delta = 1$.

Notice also that for manifolds with pinched sectional curvature, P. Buser (see [Buser]) obtained much sharper results by dividing V into the pieces which are not necessarily balls, but still have a sufficiently simple geometry.

Acknowledgements. I am thankful to V. Milman, who acquainted me with Levy's results and their applications (see [Fie–Lin–Mil], [Gro–Mil]$_{\mathrm{TAI}}$). I had a very helpful conversation with Professor W. Allard, who, in particular, pointed out to me that the lemma of §4 followed from [Alm]$_{\mathrm{Exi}}$. The interest in the subject expressed by J. Cheeger encouraged me to write this paper.

Appendix D

Systolically Free Manifolds

by Mikhail Katz[1]

This appendix fits in with Chapter 4, Section E_+, paragraph 4.45 *"On higher dimensional systoles"* (p. 264). The k-systole of a Riemannian manifold X is, roughly, the least volume of a k-dimensional non-nullhomologous submanifold of X. See paragraph 4.40 of the same section (p. 260) for a definition. A more detailed updated account of systolic freedom appears in [Katz-Suc].

The existence of a systolic inequality will mean that the volume of X imposes a constraint on systoles, as in Loewner's Theorem 4.1 and Gromov's Theorem 4.41.

We will refer to the absence of systolic inequalities as "systolic freedom." To exhibit the freedom of a manifold X, we will construct families of metrics on X whose systoles are not constrained by the volume.

There are two rather different systolic problems: one involving a pair of systoles of distinct complementary dimensions, and one involving the middle-dimensional systole.

The idea for the pair of complementary dimensions is as follows. Let $k < \frac{n}{2}$ where $n = \dim(X)$. Following M. Gromov (see $[\text{Berger}]_{\text{Sys}}$, p. 306 and $[\text{Gro}]_{\text{SII}}$ section $4.A_5$, p. 354), we use Thom's theorem (see [Thom]) to choose a k-dimensional submanifold with the following two properties: (a) its normal bundle is trivial; (b) its connected components generate (rationally) the homology group $H_k(X)$.

[1]address: UMR 9973, Département de Mathématiques, Université de Nancy 1, B.P. 239, 54506 Vandoeuvre FRANCE. email: katz@iecn.u-nancy.fr

The construction takes place within a fixed trivialized tubular neighborhood of the submanifold. Gromov described an expansion-contraction procedure (see [Gro]$_{\text{MIKM}}$, section 2) with the desired effect on the systole of the complementary dimension $n-k$. We combine the expansion-contraction with a volume-saving twist.

In more detail, Thom's theorem allows us to carry out the expansion-contraction procedure in a neighborhood of a suitable k-dimensional submanifold described above. The resulting metrics have large $(n-k)$-dimensional systole (compared to the volume). So far the $(k, n-k)$ systolic inequality is not violated. This is because the metric is essentially a product and the volume is too large. To decrease the volume, we introduce a nondiagonal term in the Riemannian metric. The key idea is due to Gromov and is described in paragraph 4.45 for the product of a circle by a simply connected manifold admitting a free circle action. Generalizing the construction requires a splitting of both the k-dimensional class and its tubular neighborhood (see below). It is at this point that the modulo 4 condition on k comes in.

D.1. Theorem ([Bab–Katz]): *Let X be an orientable $(k-1)$-connected n-dimensional manifold, where $n \geq 3$ and $k < \frac{n}{2}$. Suppose k is not a multiple of 4. Then*

$$\inf_g \frac{\operatorname{vol}(g)}{\operatorname{sys}_k(g)\operatorname{sys}_{n-k}(g)} = 0, \tag{1}$$

where the infimum is taken over all metrics g on X.

In the middle-dimensional case, we have the following general result.

D.2. Theorem ([Bab–Katz–Suc]): *Let X be a manifold of dimension $n = 2m$, where $m \geq 3$. Suppose the group $H_m(X)$ is torsion free. Then*

$$\inf_g \frac{\operatorname{vol}(g)}{\operatorname{sys}_m^2(g)} = 0, \tag{2}$$

where the infimum is taken over all metrics on X.

One may ask if systolic freedom persists if one restricts the class of competing metrics to homogeneous ones. We have the following partial result in this direction.

D.3. Theorem: *The manifold $S^3 \times S^3$ satisfies*

$$\inf_g \frac{\operatorname{vol}_6(g)}{\operatorname{sys}_3^2(g)} = 0, \tag{3}$$

where the infimum is taken over homogeneous metrics g on $S^3 \times S^3$.

A key technical result in the proof of Theorem 1 is the following lemma, inspired by Gromov's first example of free metrics on $S^1 \times S^3$ (see 4.45 and $[\text{Berger}]_{\text{Sys}}$).

D.4. Lemma ($[\text{Ber–Katz}]_{\text{II}}$,[Bab–Katz]): *Consider the cylinder $M = T^1 \times I$, circle C, and the manifold $Y = C \times M = T^2 \times I$. We view M as a relative cycle in $H_2(Y, \partial Y)$. Then*

$$\inf_g \frac{\operatorname{vol}(g)}{\operatorname{sys}_1(g)\operatorname{mass}_2([M])} = 0. \tag{4}$$

Here the infimum is taken over all metrics whose restriction to each component of the boundary $\partial Y = T^2 \times \partial I$ is the standard "unit square" torus satisfying $\operatorname{length}(C) = \operatorname{length}(T^1) = 1$. *The 2-mass of the class $[M]$ is the infimum of areas of* rational *cycles representing it.*

Proof. The metric of the universal cover $p : \mathbb{R}^3 \to N$ of the standard nilmanifold N of the Heisenberg group can be written as

$$(dz - x dy)^2 + dy^2 + dx^2. \tag{5}$$

We will use it to construct a sequence of metrics (g_j), $j \in \mathbb{N}$, on Y with the required asymptotic behavior. An alternative approach using a solvable group was used in [Pit].

The restriction of the above metric to the parallelepiped defined by the inequalities $0 \le x \le j$, $0 \le y \le 1$, $0 \le z \le 1$ defines a metric on $[0, j] \times T^1 \times C$ once we identify the opposite sides of the unit square in the y, z plane. Here $[0, j]$ is parameterized by x, the circle T^1 by y, and the circle C by z. The covering projection p induces a monomorphism

$$p_* : \pi_1([0, j] \times T^2) \to \pi_1(N) \tag{6}$$

from $\mathbb{Z}^2$ to the group of unipotent integer 3×3 matrices.

The metric on $[0, j] \times T^2$ does not satisfy the boundary condition (specified in the statement of the lemma) at $x = j$, since $\operatorname{length}(\{j\} \times T^1) = \sqrt{1 + j^2}$. We correct this by taking its "double" using a mirror placed at $x = j$. In other words, let $I = [0, 2j]$, let $X = \min(x, 2j - x)$, and consider the sequence of metrics on $Y = T^2 \times I$ defined by

$$g_j = (dz - X dy)^2 + dy^2 + dx^2. \tag{7}$$

Note that the projection p extends over Y, so that we still have the monomorphism p_* of fundamental groups.

Let us show that the growth of $\operatorname{mass}_2[M]$ is quadratic in j. Let $\phi_j(x)$ be a partition of unity-type function with support in $[0, j]$, and such that

$\phi_j(x) = 1$ for $x \in [1, j-1]$. We define a 2-form $\alpha = \alpha_j$ by setting

$$\alpha = \frac{\phi_j(x)}{\sqrt{1+x^2}} * (dz) \tag{8}$$

where $*$ is the Hodge star of the metric g_j. Let us check that α is a calibrating form, i.e., $\|\alpha\| = 1$ and $d\alpha = 0$.

By definition of g_j the 1-forms dx, dy, and $dz - xdy$ form an orthonormal basis. Therefore $\|dz\|_{(x,y,z)} = \|(dz - xdy) + xdy\| = \sqrt{1+x^2}$. Thus $\|\alpha\| = 1$. Furthermore,

$$*(dz) = dx \wedge dy - xdx \wedge (dz - xdy) = dx \wedge ((1+x^2)dy - xdz). \tag{9}$$

Let $a_j = \int_M \alpha = \int_M \phi_j(x)\sqrt{1+x^2}dx \wedge dy = \int_0^j \phi_j(x)\sqrt{1+x^2}dx$. Thus a_j grows as j^2. The calculation (9) shows that the 2-form α is closed. Hence by Stokes,

$$\mathrm{mass}_2[M] \geq \int_M \alpha = a_j, \tag{10}$$

and the growth of $\mathrm{mass}_2[M]$ is quadratic.

Meanwhile, the injectivity of p_* of (6) allows us to conclude that $\mathrm{sys}_1(g_j) \geq \pi\,\mathrm{sys}_1(N)$, where $\pi\,\mathrm{sys}_1$ is the length of the shortest noncontractible curve. Hence the 1-systole is uniformly bounded from below in j. The volume of the Riemannian submersion $(Y, g_j) \to [0, 2j]$ with fiber of unit area is $2j$. Thus

$$\frac{\mathrm{vol}(g_j)}{\mathrm{sys}_1(g_j)\,\mathrm{mass}_2[M]} = O\left(\frac{1}{j}\right), \tag{11}$$

proving the lemma.

Proof of Theorem D.1. The idea was described at the beginning of the appendix. Let $e_1, \ldots, e_b$ be an integral basis for a maximal lattice in $H_k(X, \mathbb{Z}) = \mathbb{Z}^b + \text{torsion}$. We will exhibit a k-dimensional submanifold $A \subset X$ whose connected components $A_1, \ldots, A_b$ have the following properties:

(i) each A_i is diffeomorphic to $S^{k-1} \times C$ where C is a circle;

(ii) the classes $[A_i]$ satisfy $[A_i] = 2e_i$;

(iii) the normal bundle ν of $A_i \subset X$ is trivial.

By Hurewicz's theorem, we find k-spheres $f_i : S^k \to X$, $i = 1, \ldots, b$ satisfying (ii). Since the total Stiefel-Whitney class $w(S^k) = 1$ is trivial, we have

$$w_k[f_i^*(\nu)] = w_k[f_i^*(TX)] = w_k(TX)(f_*[S^k]) = w_k(TX)(2e_i), \tag{12}$$

and so $w_k[f_i^*(\nu)] = 2w_k(TX)(e_i) = 0 \in H^0(X, \mathbb{Z}_2)$.

By real Bott periodicity, there is at most a $\mathbb{Z}_2$'s worth of inequivalent bundles of rank $\geq k+1$ over S^k for $k \neq 0 \pmod 4$, distinguished by the class w_k. Hence ν is trivial. We now perform a surgery inside X by attaching a handle to each imbedded sphere $f_i(S^k)$, turning it into a product $A_i = C \times S^{k-1}$ where C is a circle, still with trivial normal bundle. Thus a tubular neighborhood of A_i splits as a product

$$A_i \times B^{n-k}. \tag{13}$$

Let $R \subset B^{n-k}$ be any submanifold of codimension 2 with trivial normal bundle. The boundary of its tubular neighborhood is diffeomorphic to $R \times T^1$, and we obtain a hypersurface

$$\Sigma_i = A_i \times R \times T^1 = C \times S^{k-1} \times R \times T^1 = T^2 \times L \tag{14}$$

in a neighborhood of $A_i \subset X$, where $T^2 = C \times T^1$ is a 2-torus and $L = S^{k-1} \times R$.

Let $\Sigma \subset X$ be the disjoint union of such hypersurfaces $\Sigma_1, \ldots, \Sigma_b$. A tubular neighborhood of $\Sigma \subset X$ is a disjoint union of open submanifolds $\Sigma_i \times I = C \times T^1 \times L \times I$. What happens metrically can be described as follows. We choose a metric on X which is a direct sum in a neighborhood of each Σ_i, with four summands

$$C \times T^1 \times L \times I \subset X, \tag{15}$$

where C and T^1 are of unit length, L has unit volume, and I is an interval. Denote by X_0 the resulting Riemannian manifold. We now modify the direct sum metric on

$$C \times T^1 \times L \times I = Y \times L \tag{16}$$

by using the metrics g_j on $Y = C \times T^1 \times I$ of Lemma D.4, while L remains a direct summand. Denote by X_j the resulting Riemannian manifold. The boundary condition of Lemma D.4 ensures that the new metric on X varies continuously across the boundary $T^2 \times L \times \partial I$.

Proposition 1. *The systole $\mathrm{sys}_k(X_j)$ is uniformly bounded from below in j.*

Proof. Let z be a k-cycle in X_j, $[z] \neq 0$ in $H_k(X_j)$. If the support of z lies outside $\Sigma \times I$, then whether $[z]$ has finite or infinite order, we have

$$\mathrm{vol}_k(z \subset X_j) = \mathrm{vol}_k(z \subset X_0) \geq \mathrm{sys}_k(X_0), \tag{17}$$

yielding a lower bound independent of j. If z lies in $\Sigma \times I$, then its class $[z]$ lies in the span of $[A_1], \ldots, [A_b]$, so that $[z] = \lambda_1[A_1] + \ldots + \lambda_b[A_b]$. Here one of the coefficients is nonzero, say λ_l. Let

$$\pi : Y \times L \to S^{k-1} \tag{18}$$

be the projection to the first factor of $L = S^{k-1} \times R$. By the coarea inequality,

$$\mathrm{vol}_k(z) \geq \int_{S^{k-1}} \mathrm{length}(z \cap \pi^{-1}(t))dt \geq \int_{S^{k-1}} \mathrm{sys}_1(g_j)dt \geq \pi sys_1(N), \tag{19}$$

also yielding a lower bound $\mathrm{sys}_k(\Sigma \times I) \geq \pi sys_1(N)$ independent of j.

Here we must verify that the 1-cycle $z \cap \pi^{-1}(t)$ represents a nonzero class in $\pi^{-1}(t) = (Y, g_j) \times R$. This is so because the set-theoretic formula

$$(C \times S^{k-1}) \cap (Y \times R) = C \tag{20}$$

makes sense at the homological level. Indeed, inside $\Sigma_l \times I = Y \times L$ we have

$$\begin{aligned} [z] \smile [Y \times R] &= \lambda_l[A_l] \smile [Y \times R] = \lambda_l[C \times S^{k-1}] \smile [Y \times R] \\ &= \lambda_l[(C \times S^{k-1}) \cap (C \times T^1 \times I \times R)] = \lambda_l[C] \neq 0, \end{aligned}$$

where $\smile$ is the homological operation (dual to cup product in cohomology) pairing the groups $H_k(\Sigma_l \times I)$ and $H_{n-k+1}(\Sigma_l \times I, \Sigma_l \times \partial I)$. This operation is realized by transverse intersection of representative cycles, and hence $[z \cap \pi^{-1}(t)] \neq 0$.

For general z we argue as follows. Assume for simplicity that $b_k(X) = 1$. We will cut z at a narrow place and replace it by the sum of two cycles, each of which can be handled as above. If $b_k > 1$, we have to make several cuts, but the rest of the argument is the same.

We argue by contradiction. Suppose X_j contains a k-cycle z_j such that

$$\lim_{j \to \infty} \mathrm{vol}_k(z_j) = 0. \tag{22}$$

Let $d\colon X_j \to [0, 2j]$ be the map extending the x-coordinate of Lemma D.4 by collapsing the components of the complement

$$X \setminus (\Sigma \times I) \tag{23}$$

to the respective endpoints. By the coarea inequality, we find points $x_- \in [0,1] \subset I$ and $x_+ \in [2j-1, 2j] \subset I$ such that

$$\mathrm{vol}_{k-1}(z \cap d^{-1}(x_\pm)) \to 0 \tag{24}$$

as $j \to \infty$. Let $\gamma_\pm = z \cap d^{-1}(x_\pm)$. Since $\mathrm{vol}_{k-1}(\gamma_\pm) = o(1)$, $\gamma_\pm$ can be filled in by a k-chain $D_\pm$ with $\mathrm{vol}_k(D_\pm) = o(1)$. This follows from the isoperimetric inequality for small cycles proved in [Gro]$_{\mathrm{FRM}}$, Sublemma 3.4.B′. Since $\partial D_\pm = \gamma_\pm$, we can write $z_j = z' + z''$ where $z' = z \cap d^{-1}([x_-, x_+]) + D_+ - D_-$ is a k-cycle in $\Sigma \times I$. Now

$$\begin{aligned}
\mathrm{vol}_k(z_j) &\geq \min(\mathrm{vol}_k(z'), \mathrm{vol}_k(z'')) - \mathrm{vol}_k(D_+ + D_-) \\
&\geq \min(\mathrm{sys}_k(X_1), \mathrm{sys}_k(\Sigma \times I)) - o(1) \\
&\geq \min(\mathrm{sys}_k(X_1), \pi\, \mathrm{sys}_1(N)) - o(1)
\end{aligned}$$

providing a lower bound independent of j.

D.5. Lemma: *Let X_j be the Riemannian manifold constructed above. Then the $(n-k)$ systole $\mathrm{sys}_{n-k}(X_j)$ grows quadratically in j.*

Proof. Let α_i be the 2-form of Lemma D.4 with support in a neighborhood of A_i. We use the calibrating form

$$\beta_i = \alpha_i \wedge \mathrm{vol}_R \tag{26}$$

where vol_R is the volume form of the codimension 2 submanifold $R \subset B^{n-k}$. Note that $H_{n-k}(X)$ is torsion free as $H_{k-1}(X) = 0$ by hypothesis of Theorem D.1. Using Poincaré duality, we choose integer $(n-k)$-cycles $M_1, \dots, M_b$ which define a basis of $H_{n-k}(X, \mathbb{Z})$ "dual" to the basis $e_1, \dots, e_b$ of $H_k(X, \mathbb{Z})$, so that the *algebraic* intersections satisfy

$$[M_i] \cap e_h = \delta_{ih} \tag{27}$$

and therefore $M_i \cap A_h = 2\delta_{ih}$. We may assume that all intersections of M_i and A_h are transverse and moreover standard, so that $M_i \cap (\Sigma_h \times I)$ is a disjoint union of copies of $M \times R$ where $M = T^1 \times I$ is the cylinder of Lemma D.4. Hence $\int_{M \times R} \beta_i = \pm a_j$, where $a_j = \int_M \alpha_j$ (cf. Lemma D.4), for each connected component $M \times R$ of $M_i \cap (\Sigma_h \times I)$. In particular

$$\int_{M_i \cap (\Sigma_h \times I)} \beta_i = 0 \text{ if } \ i \neq h. \tag{28}$$

Finally, let $m \in H_{n-k}(X)$ be a nonzero class. Let $m = \epsilon_1 d_1 [M_1] + \ldots + \epsilon_b d_b [M_b]$ where $\epsilon_i = \pm 1$ and $d_i \geq 0$. We use the signs ϵ_i to specify a suitable calibrating form β by setting

$$\beta = \epsilon_1 \beta_1 + \ldots + \epsilon_b \beta_b. \tag{29}$$

Now

$$\begin{aligned}
\mathrm{mass}_{n-k}(m) &\geq \int_m \beta = \Sigma_{i=1}^b \epsilon_i d_i \int_{M_i} \beta = \Sigma_i \epsilon_i d_i \int_{M_i} (\Sigma_h \epsilon_h \beta_h) \\
&= \Sigma_i \epsilon_i^2 d_i \int_{M_i} \beta_i \text{ (by 53)} \\
&= \Sigma_i d_i \int_{M_i} \beta_i = \Sigma_i d_i 2a_j \geq a_j \qquad (30)
\end{aligned}$$

since $m \neq 0$, where $a_j = \int_M \alpha_j$ grows quadratically in j. Hence the growth of $\mathrm{sys}_{n-k}(X_j)$ is quadratic in j.

Now Propositions 1 and 2 and the linear growth of the volume of X_j yield

$$\frac{\mathrm{vol}(X_j)}{\mathrm{sys}_k(X_j)\,\mathrm{mass}_{n-k}(X_j)} = O\left(\frac{1}{j}\right). \qquad (31)$$

This completes the proof of Theorem D.1.

We now prove some preliminary results for Theorem D.2, in the form of steps 1 through 4.

Step 1. A $(2m-1)$-dimensional complex admits a map to the bouquet of m-spheres which induces monomorphism in m-dimensional homology with rational coefficients.

Proof. Let K^i be the i-skeleton of the complex K. From the exact sequence of the pair (K, K^{m-1}), we see that the homomorphism $H_m(K) \to H_m(K/K^{m-1})$ is injective. This reduces the problem to the case when K is $(m-1)$-connected. Now let $b_m = b_m(K)$ and let $H_m(K) \to \mathbb{Z}^{b_m}$ be the quotient by the torsion subgroup. Let $f : X \to K(\mathbb{Z}^{b_m}, m)$ be the associated map to the Eilenberg-Maclane space. Now $K(\mathbb{Z}^{b_m}, m) = (K(\mathbb{Z}, m))^{b_m}$ and $K(\mathbb{Z}, m)$ admits a cell structure with no cells in dimension $m+1$. Hence by the cellular approximation theorem, we obtain a map from the $(m+1)$-skeleton of K to the bouquet $\vee S^m$ of b_m spheres S^m, injective on nontorsion m-dimensional homology classes. We now proceed as in [Serre]$_{\mathrm{GH}}$ (page 278 and 287–288). Here an induction on skeleta uses self-maps of the bouquet $\vee S^m$ which annihilate all relevant homotopy. By Eckmann's theorem [Eck], a self map f_q of degree q of the sphere S^m induces multiplication by q in the (stable) finite group $\pi_i(S^m)$ for $m+1 \leq i \leq 2m-2$. By Hilton's theorem, we have a splitting of the stable homotopy groups for a bouquet. Let $g : K^i \to \vee S^m$ be a map defined on the i-skeleton. Then $f_q \circ g$ extends to the $(i+1)$ skeleton if q is a multiple of the order of the finite group $\pi_i(S^m)$.

Step 2. A bouquet of m-spheres admits a self-map which induces a homomorphism of the first unstable group whose image is contained in the subgroup generated by Whitehead products.

Proof. For a single m-sphere, we argue as follows. Let $G \subset \pi_{2m-1}(S^m)$ be the subgroup generated by the Whitehead product $[e,e]$ where e is the fundamental class of the sphere. Then G is precisely the kernel of the suspension homomorphism

$$E\colon \pi_{2m-1}(S^m) \to \pi_{2m}(S^{m+1}). \tag{32}$$

Since E commutes with f_q, it suffices to choose q to be the order of the group $\pi_{2m}(S^{m+1})$. For a bouquet $\vee_i S_i^m$, let e_i be the fundamental class of S_i^m. By Hilton's theorem, the group $\pi_{2m-1}(\vee_i S_i^m)$ modulo mixed Whitehead products $[e_i, e_j]$ for $i \neq j$ is just

$$\oplus_i \pi_{2m-1}(S_i^m). \tag{33}$$

Hence the same argument applies.

If we start with a $2m$-manifold represented as a CW complex with a single $2m$-cell, we can apply Steps 1 and 2 to its $(2m-1)$ skeleton K^{2m-1} to construct a suitable map to the bouquet $\vee S^m$. We can now proceed in two different ways: (a) either by eliminating all homotopy of the bouquet (at least rationally) in dimension $2m-1$, or, more economically, (b) eliminating just enough to allow the map to extend across the $2m$-cell. Described below is the first approach.

D.6. Definition: A $2m$-dimensional complex is *"free"* if

$$\inf_g \frac{\mathrm{vol}_{2m}(g)}{\mathrm{sys}_m^2(g)} = 0, \tag{34}$$

where the infimum is taken over all metrics g.

Step 3. Consider a $2m$-dimensional complex X obtained from a bouquet of m-spheres by attaching b copies of a $2m$-cell. If the closure of each $2m$-cell is free, then X is free.

Proof. One chooses the following representative X_0 in the homotopy class of X. We take the Cartesian product of the bouquet $\vee S^m$ with the bouquet of b intervals of sufficiently big length l. Here each interval corresponds to a $2m$-cell. Each $2m$-cell is then attached at the end of its interval. We choose free metrics of the same volume on each cell closure. The freedom of X reduces to obtaining a lower bound on the m-volume of a cycle z in

X_0. This is done as follows. Suppose for simplicity that there are only two $2m$-cells, attached at either end of the cylinder

$$\vee S^m \times [0,l] \tag{35}$$

(for l to be chosen later). We use the coarea inequality to find a narrow place and cut z into two pieces along an $(m-1)$-cycle $\gamma = z \cap p^{-1}(x)$. Here $p : \vee S^m \times [0,l] \to [0,l]$ is the projection to the second factor. The point $x \in [0,l]$ is chosen so that

$$\mathrm{vol}_{m-1}(\gamma) \leq \frac{1}{l}\,\mathrm{vol}_m(z). \tag{36}$$

It remains to "fill in" the cycle γ. This can be done by the isoperimetric inequality for cycles of small volume ([Gro]$_{\mathrm{FRM}}$, Sublemma 3.4.B$'$). Here l is chosen sufficiently big so that the inequality would apply to γ.

D.7. Definition: Let X and Y be $2m$-dimensional manifolds. A *"meromorphic map"* from X to Y is a continuous map from X to a CW complex obtained from Y by attaching cells of dimension at most $2m-1$, inducing monomorphism in middle-dimensional homology.

Example: Let X be a complex surface and $\widehat{X} \to X$ its blow-up at a point $p \in X$. Then the classical meromorphic map $X \to \widehat{X}$ can be modified in a neighborhood of p and extended to a continuous map from X to $\widehat{X} \cup_f B^3$ where the 3-ball is attached along the exceptional curve.

Step 4. The complex obtained by 'glueing in' a single Whitehead product is free if $m \geq 3$.

Proof. For a "mixed" Whitehead product $[e_i, e_j]$, we obtain that $(S^m \vee S^m) \cup_{[e_i,e_j]} B^{2m} = S^m \times S^m$, which is free by [Katz].

Consider the complex P defined by

$$P = S^m \cup_{[e,e]} B^{2m}, \tag{37}$$

where e is the fundamental class of S^m. The idea is to construct a "meromorphic map" from P to the product of spheres, using multiplicative structure in the algebra of homotopy groups. Let a and c be the generators of $H_m(S^m \times S^m)$, so that $[a,c] = 0$. We now attach an $(m+1)$-cell along the diagonal $a + c$. Let

$$W = S^m \times S^m \cup_{a+c} B^{m+1}, \tag{38}$$

so that $a = -c$ in W. Let $f : S^m \to W$, $e \mapsto a$. Then

$$f([e,e]) = [a,a] = [a,-c] = 0, \tag{39}$$

and so the map extends to $P \to W$. We pull back to P the free metrics on W. Here the cell added to $S^m \times S^m$ does not affect the $(2m)$-dimensional volume (to obtain a positive definite metric on P, we increase slightly the pullback of the quadratic form, using the convexity of the cone of positive quadratic forms). This proves the freedom of P once we know that of $S^m \times S^m$.

Proof of Theorem D.2. Let X be any CW complex of dimension $2m$. We map its $(2m-1)$-skeleton K to a bouquet as in Step 1. By Step 2, we can assume that the elements of $\pi_{2m-1}(K)$ are mapped to the subgroup G generated by (finitely many) Whitehead products in the bouquet. We attach a $2m$-cell to this bouquet for each generator of G. The resulting space is free by Steps 3 and 4. Now the map from K extends to X. It follows that X is free as well.

Remark: The proof can be simplified in the case of projective planes $\mathbb{H}P^2$ and CaP^2. It relies on the following proposition.

D.8. Proposition: *Let F be $\mathbb{C}$, $\mathbb{H}$ or Ca, and let $m = \dim(F)$. Consider the map of degree q from FP^1 to the first component of $S^m \times S^m$, to which we glue an $(m+1)$-ball along the diagonal sphere. Then the map extends to the projective plane FP^2 if and only if q is a multiple of*

$$2\, tor(\pi_{2m-1}(FP^1)), \tag{40}$$

where tor is the order of the torsion subgroup, equal to 1, 12, or 120 respectively when F is $\mathbb{C}$, $\mathbb{H}$, or Ca.

Proof of Theorem D.3. The metrics we use are Riemannian submersions over a round 2-sphere, where the fiber is Gromov's left-invariant metric on $S^1 \times S^3$ which was the first example of systolic freedom. To describe free left-invariant metrics on $S^3 \times S^3$, it is convenient first to describe some useful coordinates on flat tori of unit area and unit 1-systole. Let $r \geq 0$, and consider the metric

$$(dz - r\,dy)^2 + dy^2 \quad \text{on} \quad T^2 = \mathbb{R}^2/\mathbb{Z}^2 = T^1 \times C, \tag{41}$$

and let $*$ be its Hodge star operator. Here T^1 and C are parameterized respectively by the y and z axes. Note that if r is an integer, this torus is isometric to the standard "unit square" torus, one of whose sides is the loop C (what will appear in the sequel is the homothetic metric "square of side 2π"). Meanwhile $\text{length}^2(T^1) = \|\partial/\partial y\|^2 = 1 + r^2$. We set $R^2 = 1 + r^2$. We have

$$*(dz) = *(dz - r\,dy) + r*(dy) = dy - r(dz - r\,dy) = (1+r^2)dy - r\,dz = R^2 dy - r\,dz. \tag{42}$$

Now we describe Gromov's metric on $S^1 \times S^3$ in a way that lends itself to generalization. Let $S^1 = C$ be the unit circle with the standard 1-form dz, and let b be the standard contact 1-form on S^3. Gromov's metric is then obtained simply by modifying the product metric of a circle of length 2π and 3-sphere of radius $R = \sqrt{1+r^2}$, by adding a nondiagonal term $-2rb\,dz$ (symmetric tensor product). Here the Hopf great circles of length $2\pi R$ play the role of T^1 above. More explicitly, let us complete b to a basis (b, b', b'') of 1-forms which is orthonormal with respect to the metric of unit radius, so that $db = b' \wedge b''$. Then Gromov's metric is a modification of formula (41):

$$(dz - rb)^2 + b^2 + R^2((b')^2 + (b'')^2). \tag{43}$$

Note that

$$*(dz) = *(dz - rb) + r * b = (b - r(dz - rb)) \wedge R^2 db = R^4 b \wedge db - rR^2 dz \wedge db. \tag{44}$$

The form $*dz$ is closed since $d * dz = R^4 db \wedge db = 0$. We normalize $*dz$ to obtain a calibrating form $\frac{1}{R} * dz$, whose restriction to the 3-sphere $S^3 \subset S^1 \times S^3$ coincides with the volume form $R^3 b \wedge db$ of this round sphere of radius R.

Now take two 3-spheres, denoted respectively (A, a) and (B, b), with their contact forms a and b. Choose metrics on A and B as follows. The sphere A is endowed with the metric of a Riemannian submersion, whose fiber is of length 2π, over a round 2-sphere of radius $\sqrt{r}R$ (the reason for choosing $\sqrt{r}R$ instead of $R^{3/2}$ will become clear later). The sphere B is a round 3-sphere of radius R. We form the direct sum metric on $A \times B$, of volume growth on the order of R^6 when $R \to \infty$. Next we modify it by adding the nondiagonal term

$$-2rab \quad \text{where } r = \sqrt{R^2 - 1}, \tag{45}$$

and a and b are the contact forms. The new metric on $A \times B$ has volume growth R^5 instead of R^6, due to the appearance in the determinant of the unimodular matrix

$$\begin{bmatrix} 1 & r \\ r & R^2 \end{bmatrix} \tag{46}$$

with respect to the pair of vectors dual to (a, b). More precisely, let (a, a', a'') be a standard basis completing a, and similarly for b, as before. The metric described above can be written down as follows:

$$rR^2((a')^2 + (a'')^2) + (a - rb)^2 + b^2 + R^2((b')^2 + (b'')^2). \tag{47}$$

Here we inflate the metric of the contact plane of the first factor A to endow it with volume growth R^3, that of the second factor. If $S^1 \subset A$ is a Hopf fiber, the restriction of the metric to $S^1 \times B \subset A \times B$ coincides with Gromov's metric. Then $\mathrm{sys}_3([A]) = \mathrm{vol}(A) = 8\pi^2 r R^2$ by a standard argument using the coarea inequality. As far as B is concerned, we show that $\mathrm{sys}_3([B])$ has the growth of $\mathrm{vol}(B) = 2\pi^2 R^3$. We show this by using a calibrating form described below.

By analogy with the case of $S^1 \times S^3$, one would like to replace $*(dz)$ by $*(a \wedge da)$. The difficulty is that this form is no longer closed. There are two natural ways to introduce a corrective term, only one of which works, as we now show.

Lemma D.6: *With respect to the metric* (47) *we have:*

(i) $\|da\| = r^{-1}R^{-2}$ and $\|db\| = R^{-2}$;

(ii) $\|a \wedge da\| = r^{-1}R^{-1}$;

(iii) $d * (a \wedge da) = -da \wedge db$.

Proof. (i) Indeed, $\|\sqrt{r}Ra'\| = 1$ and $\|Rb'\| = 1$. (ii) We have $\|a \wedge da\| = \|a\|\|a'\|\|a''\| = \sqrt{1+r^2}r^{-1}R^{-2} = r^{-1}R^{-1}$. (iii) We have

$$\begin{aligned} *(a \wedge da) &= *(((a-rb)+rb) \wedge da) = (b - r(a-rb))\|da\|\|db\|^{-1}db \\ &= r^{-1}((1+r^2)b - ra)db. \end{aligned}$$

Note that the restriction of the form $rR*(a \wedge da)$ to a coordinate sphere $B \subset A \times B$ is

$$R(1+r^2)b \wedge db = R^3 b \wedge db, \tag{49}$$

the volume form of B induced by the metric (47). We have $d * (a \wedge da) = r^{-1}(R^2 db - r da) \wedge db = -da \wedge db$. Note that we chose the metric on A so as to render the formula for $d * (a \wedge da)$ free of r and R. Thus we need to correct $*(a \wedge da)$ by $(1-t)da \wedge b + ta \wedge db$, for some $t \in \mathbb{R}$, to make it closed.

Now $\|a \wedge db\| = \|a\|\|b'\|\|b''\| = \sqrt{1+r^2}R^{-2} = R^{-1}$ while $\|da \wedge b\| = \|a'\|\|a''\|\|b\| = r^{-1}R^{-2}$, which is much smaller. In view of (ii) above, the closed form

$$r^2(*(a \wedge da) + da \wedge b) \tag{50}$$

is of norm which tends to 1 as $r \to \infty$. Using this form, we obtain volume estimates on the order of R^3 for $[B]$ and also any class $\alpha[A] + \beta[B]$ with β in $\mathbb{Z}^*$. This completes the proof of Theorem D.3.

Acknowledgments. The author is grateful to the Geometry & Dynamics Program at Tel Aviv University for hospitality during the preparation of a part of this work.

Bibliography

[Abr] Abraham, R., Bumpy Metrics, in *Global Analysis*, S.S. Chern and S. Smale, eds., Symposium in Pure Math., American Math. Soc., Providence, 1970.

[Abre–Grom] Abresch, U. and Gromoll, D., On complete manifolds with nonnegative Ricci curvature, *J. Amer. Math. Soc.* **3** (1990), 355–374.

[Acc] Accola, R.D.M., Differentials and extremal lengths on Riemann surfaces, *Proc. Natl. Acad. Sci.* **46** (1960), 540–543.

[Ad–Hil] Adams, J. F. and Hilton, P. J., On the chain algebra of a loop space, *Comment. Math. Helv.* **30** (1956), 305–330.

[Agard] Agard, S., Mostow rigidity on the line: a survey, in *Holomorphic functions and moduli II*, Mathematical Sciences Research Institute publications **11** (1988), Springer-Verlag, 1–12.

[Ahl] Ahlfors, L., Zur Theorie des Überlagerungs-flächen, *Acta Math.* **65** (1936), 157–194.

[Alex]$_{\mathrm{Tri}}$ Alexandrov, A. D., A theorem on triangles in metric spaces and some of its applications, *Trudy Math. Inst. Steklov* **38** (1951), 5–23.

[Alex]$_{\mathrm{VRG}}$ Alexandrov, A. D., Über eine Verallgemeinerung der riemannschen Geometrie, *Schr. Forschungsinst. Math.* **1** (1957), 33–84.

[Alex–Zalg] Alexandrov, A. D. and Zalgaller, V. A., Two-dimensional manifolds of bounded curvature, *Trudy Math. Inst. Steklov* **63** (1962).

[Alm]$_{\mathrm{Exi}}$ Almgren, F., Existence and regularity almost everywhere of solutions to elliptic variational problems with constraints, *Mem. Am. Math. Soc.*, **4** (1976).

[Alm]$_{\mathrm{Iso}}$ Almgren, F., An isoperimetric inequality, *Proc. Am. Math. Soc.* **15** (1975), 284–285.

[Alm]$_{\mathrm{Opt}}$ Almgren, F., Optimal isoperimetric inequalities, *Indiana Univ. Math. J.*, **35** (1986), 451–547.

[Alon] Alon, N., Eigenvalues, geometric expanders, sorting in rounds, and Ramsey theory, *Combinatorica* **6** (1986), 207–219.

[Alon–Milm] Alon, N. and Milman, V. D., Concentration of measure phenomena in the discrete case and the Laplace operator of a graph, in *Israel seminar on geometrical aspects of functional analysis* (1983/84), XIII-3, Tel Aviv Univ., Tel Aviv, 1984.

[Al–Pr–Wa] Alonso, J. M., Pride, S. J., Wang, X., Higher dimensional isoperimetric (or Dehn) functions of groups, preprint.

[And] Anderson, M., Hausdorff perturbations of Ricci-flat metrics and the splitting theorem, *Duke J.* **68**, 1992.

[Assou]$_{\mathrm{DM}}$ Assouad, P., Étude d'une dimension métrique liée à la possibilité de plongement dans $\mathbb{R}^n$, *C. R. Acad. Sci. Paris* **288** (1979), 731–734.

[Assou]$_{\mathrm{EM}}$ Assouad, P., *Espaces Métriques, Plongements, Facteurs*, Thèse de Doctorat (January, 1977), Université de Paris XI, Orsay, France.

[Assou]$_{\mathrm{PL}}$ Assouad, P., Plongements Lipschitziens dans $\mathbb{R}^n$, *Bull. Soc. Math. France* **111** (1983), 429–448.

[Avez] Avez, A., Variétés sans points focaux, *C.R.A.S.* **270** (1970), 188–191.

[Bab–Katz] Babenko, I., Katz, M., Systolic freedom of orientable manifolds, preprint.

[Bab–Katz–Suc] Babenko, I., Katz, M., Suciu A., (in preparation).

[Ball] Ball, K., Markov chains, Riesz transforms, and Lipschitz maps, *Geom. and Funct. Analysis* **2** (1992).

[Ba–Gr–Sch] Ballmann, W., Gromov, M., Schroeder, V., *Manifolds of nonpositive curvature*, Progress in Math. vol. 61, Birkhäuser, Boston, 1985.

[Bang] Bang, T., A solution of the "Planck" problem, *Proc. Amer. Math. Soc.* **2** (1951), 990.

[Barlow] Barlow, M., Harmonic analysis on fractal spaces, Séminaire Bourbaki 1991/92, *Astérisque* **206**, (1992), Exp. No. 755, 5, 345–368.

[Bar–Bass] Barlow, M., and Bass, R., Coupling and Harnak inequalities for Sierpinski carpets, *Bull. Amer. Math. Soc.* (N. S.) **29** (1993), 208–212.

[Barv] Barvinok, A., Measure concentration in optimization, to appear in *Mathematical Programming: State of the Art*, 1997.

[Bass] Bass, H., The degree of polynomial growth of finitely generated nilpotent groups, *Proc. London Math. Soc.* **25** (1972), 603–614.

[B–J–L–P–S] Bates, S., Johnson, W. B., Lindenstrauss, J., Preiss, D., Schechtman, G., Nonlinear quotients, preprint.

[Bau–Pri] Baumslag, G. and Pride, S., Groups with one more generator than relators, *Math. Z.* **167** (1979), 279–281.

[Ber–Gal] Bérard, P. and Gallot, S., Inégalités isopérimétriques pour l'équation de la chaleur et application à l'estimation de quelques invariants, in *Goulaouic, Meyer, Schwartz seminar*, 1983–1984, Exp. No. 15, 35 pp., Ecole Polytech., Palaiseau, 1984.

[Ber–Katz]$_{\mathrm{II}}$ Berard Bergery, L. and Katz, M., Intersystolic inequalities in dimension 3, *Geometric and Functional Analysis* **4**, No. 6 (1994), 621–632.

[Berger]$_{\mathrm{Cours}}$ Berger, M., *Cours de Géométrie*, Cedic-Fernand Nathan, Paris (1977).

[Berger]$_{\mathrm{LGRG}}$ Berger, M., *Lectures on Geodesics in Riemannian Geometry*, Tata Institute, Bombay, 1965.

[Berger]$_{\text{ombre}}$ Berger, M., A l'ombre de Loewner, *Ann. Sc. Ec. Norm. Sup.* **5** (1972), 241–260.

[Berger]$_{\text{Pu}}$ Berger, M., Du côté de chez Pu, *Ann. Sc. Ec. Norm. Sup.* **5** (1972), 1–44.

[Berger]$_{\text{QPGR}}$ Berger, M., Quelques problèmes de géométrie riemannienne, *l'Enseignement Math.*, **16** (1970), 73–96.

[Berger]$_{\text{RGSHXX}}$ Berger, M., Riemannian geometry during the second half of the twentieth century, *Jahrbericht der Deutsche math. Vereinigung*, v. 100 (1998), 45–208.

[Berger]$_{\text{Sys}}$ Berger, M., Systoles et applications selon Gromov, *Sém. Bourbaki Exp. 771 Astérisque* (1993), 274–310.

[Berger]$_{\text{VCN}}$ Berger, M., Variétés à courbure négative, *Séminaire de Géométrie Riemannienne*, Publ. Math. de l'Univ. Paris (1971).

[Berger]$_{\text{VIR}}$ Berger, M., Some relations between volume, injectivity radius, and convexity radius in riemannian manifolds, in *Differential Geometry and Relativity*, Cahen and M. Flato, eds, D. Reidel 1976, 33–42.

[Ber–Gau–Maz] Berger, M., Gauduchon, P., Mazet, E., *Le Spectre d'une Variété Riemannienne*, Springer Lecture Notes **194** (1971).

[Ber–Gost] Berger, M., Gostiaux, *Géométrie Différentielle*, Armand Colin (1971).

[Bers] Bers, L., Quasiconformal mappings with applications to differential equations, function theory, and topology, *Bull. Amer. Math. Soc.* **83** (1977), 1083–1100.

[Besi] Besicovitch, A., On two problems of Loewner, *J. London Math. Soc.*, **27** (1952), 141–144.

[Besse] Besse, M., *Manifolds all of whose geodesics are closed*, Springer, New York, 1976.

[B–C–G] Besson, G., Courtois, G., and Gallot, S., Minimal Entropy and Mostow's Rigidity Theorems, *Ergod. Th. & Dynam. Sys.* **16** (1996), 623–649.

[Beur–Ahl] Beurling, A. and Ahlfors, L., The boundary correspondence under quasiconformal mappings, *Acta Math.* **96** (1956), 125–142.

[Bing] Bing, R.H., A convex metric with unique segments, *Proc. Amer. Math. Soc.* **4** (1953), 167–174.

[Bish–Crit] Bishop, R., Crittenden, *Geometry of Manifolds*, Academic Press, New York, 1964.

[Bish–Steg] Bishop, R., Steger, T., Three rigidity criteria for $PSL(2,\mathbb{R})$, *Bull. Amer. Math. Soc.* **24** (1991), 117–123.

[Blat] Blatter, C., Über Extremallängen auf geschlossenen Flächen, *Comm. Math. Helv.* **35** (1961) 151–168.

[Blum] Blumenthal, *Theory and Applications of Distance Geometry*, Cambridge Univ. Press, 1953, reprinted Chelsea Publ., New York, 1979.

[Boch–Yano] Bochner, S. and Yano, K., *Curvature and Betti numbers*, Princeton University Press, Princeton, 1953.

[Bob] Bobkov, S.G., An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space, *Ann. Prob.* **25** (1997), 206–214.

[Bob–Houd] Bobkov, S.G. and Houdré C., Isoperimetric constants for product probability measures, *Ann. Prob.* **25** (1997), 184–205.

[Bob–Led] Bobkov, S.G. and Ledoux, M., Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution *Prob. Theory Relat. Fields* **107** (1997), 383–400.

[Bog] Bogopolski, O., Commensurable hyperbolic groups are bilipschitz equivalent, preprint (Russian).

[Bol] Bollobás, B. and Varopoulos, N. Th., Representation of systems of Measurable sets, *Math. Proc. Cambridge Philos. Soc.* **78** (1975), no. 2, 323–325.

[Bour] Bourbaki, N., *Eléments de Mathématiques, Topologie Générale*, tome 1, exercise 19, Hermann.

[Bourg] Bourgain, J., On the distribution of polynomials on high-dimensional convex sets, Springer Lecture Notes 1669, 1991, 127–137.

[Brez–Nir] Brezis, H. and Nirenberg, L., Degree theory and BMO; Part I: Compact manifolds without boundaries, *Selecta Math.* (New Series) **1** (1995), 197–263.

[Brid] Bridson, M., Fractional isoperimetric inequalities and subgroup distortion, preprint.

[Bri–Vog] Bridson, M. and Vogtmann, K., On the geometry of the automorphism group of a free group, *Bull. London Math. Soc.* **27** (1995), 544–552.

[Brody] Brody, R., Compact manifolds and hyperbolicity, *Trans. Amer. Math. Soc.* **235** (1978), 213–219.

[Bro–Gre] Brody, R. and Green, M., A family of smooth hyperbolic hypersurfaces in P_3, *Duke Math. J.* **44** (1977), 873–874.

[Bur]$_{\mathrm{PM}}$ Burago, D., Periodic metrics, *Adv. in Sov. Math.* **9** (1992), 205–210.

[Bur–Klei] Burago, D. and Kleiner, B., Separated nets in euclidean space and jacobian of bi-Lipschitz maps, preprint.

[Bur–Iv]$_{\mathrm{AV}}$ Burago, D. and Ivanov, S., On asymptotic volume of tori, *Geometric and Functional Analysis* **5** (1995), 800–808.

[Bur–Iv]$_{\mathrm{RT}}$ Burago, D. and Ivanov, S., Riemannian tori without conjugate points are flat, *Geometric and Functional Analysis* **4** (1994), 259–269.

[Bur–Maz] Burago, Yu. and Maz'ya, V., Some problems of the potential theory and the function theory for domains with non-regular boundary, *Zapiski*, AOMI, Leningrad, (1967) (Russian).

[Bur–Zalg] Burago, Yu. and Zalgaller, V., *Geometric Inequalities*, Springer, New York, 1988.

[Bu–Gr–Per] Burago, Yu., Gromov, M., and Perelman, G., A. D. Alexandrov spaces with curvatures bounded below, *Russian Math. Surveys* **47** (1992), 1–58.

[Bur–Pat] Burns, K. and Paternain, G., Counting geodesics on a riemannian manifold and topological entropy of geodesic flows, preprint 1996.

[Busemann] Busemann, H., *The Geometry of Geodesics*, Academic Press, New York (1955).

[Buser] Buser, P., On Cheeger's inequality $\lambda_1 \geq \frac{h^2}{4}$, in *Geometry of the Laplace operator*, Proc. Symp. Pure Math. **36** Amer. Math. Soc., pp. 29–77.

[Bus–Kar] Buser, P. and Karcher, H., Gromov's almost flat manifolds, *Astérisque*, 1981.

[Buy] Buyalo, S., An extremal problem of riemannian geometry, *Math. Notes Ak. Sci.* USSR, **19** (1976), 795–804 (English translation, 468–472).

[Cal] Calabi, E., Extremal isosystolic metrics for compact surfaces, *Actes de la table ronde de géometrie diff. en l'honneur de Marcel Berger*, Collection SMF **1**.

[Cal]$_{\mathrm{MK}}$ Calabi, E., Métriques kähleriennes et fibrés holomorphes, *Ann. Sci. Ecole Norm. Sup.* **12** (1979), 269–294.

[Cald] Calderón, A., *Commutators, singular integrals on Lipschitz curves and applications*, Proc. I. C. M. Helsinki (1978), Acad. Sci. Fennica, Helsinki, 1980, 85–96.

[Cass] Cassels, J.W.S., *An Introduction to the Geometry of Numbers*, Springer, Berlin, 1959.

[Chab] Chabauty, C., Limites d'ensembles et géométrie des nombres, *Bull. Soc. Math. France* **78** (1950), 143–151.

[Che]$_{\mathrm{CFT}}$ Cheeger, J., Comparison and finiteness theorems for riemannian manifolds, Ph.D. thesis, Princeton Univ., 1967.

[Che]$_{\mathrm{Fin}}$ Cheeger, J., Finiteness theorems for riemannian manifolds, *Amer. J. Math.* **92** (1970), 61–74.

[Che]$_{\mathrm{lbse}}$ Cheeger, J., A lower bound for the smallest eigenvalue of the Laplacian, *Problems in Analysis*, Princeton, (1970), 195–199.

[Che]$_{\mathrm{RBLD}}$ Cheeger, J., The relation between the Laplacian and the diameter for manifolds of non-negative curvature, *Archiv. der Math.* **XIX** (1968), 558–560.

[Che–Col]$_{\mathrm{AR}}$ Cheeger, J. and Colding, T., Almost rigidity of warped products and the structure of spaces with Ricci curvature bounded below, *C.R. Acad. Sci. Paris*, Ser 1 Math **320** 3(1995), 355–357.

[Che–Col]$_{\mathrm{LBRC}}$ Cheeger, J. and Colding, T., Lower bounds on Ricci curvature and the almost rigidity of warped products, *Ann. of Math.* (2) 144, 1 (1996), 189–223.

[Che–Col]$_{\mathrm{SSRC}}$ Cheeger, J. and Colding, T., On the structure of spaces with Ricci curvature bounded below, I., to appear *Ann. Math.*

[Che–Col]$_{\mathrm{SSRC}}$ Cheeger, J. and Colding, T., On the structure of spaces with Ricci curvature bounded below, II., preprint.

[Che–Col]$_{\mathrm{SSRC}}$ Cheeger, J. and Colding, T., On the structure of spaces with Ricci curvature bounded below, III., preprint.

[Che–Ebin] Cheeger, J. and Ebin, D., *Comparison Theorems in Riemannian Geometry*, North Holland, New York, 1975.

[Che–Fu–Gr] Cheeger, J., Fukaya, K., and Gromov, M., Nilpotent structures and invariant metrics on collapsed manifolds, *J. Amer. Math. Soc.*, **5** (1992), 327–372.

[Che–Gro] Cheeger, J. and Gromov, M., Collapsing riemannian manifolds, in *Differential Geometry: A Symposium in honour of Manfredo do Carmo*, Longman Scientific, 1989, 85–94.

[Che–Rong] Cheeger, J. and Rong, X., Existence of polarized F-structures on collapsed manifolds with bounded curvature and diameter, *Geo. and Funct. Analysis* **6** (1996), 411–429.

[Cheng] Cheng, S., Eigenvalue comparison theorem and its geometric applications, *Math. Z.*, **143** (1975), 298–297.

[Choq] Choquet, G., *Topologie*, Masson, 1964.

[Christ]$_{\mathrm{LSIO}}$ Christ, M., *Lectures on Singular Integral Operators*, CBMS Regional Conference Series in Math. **77**, American Math. Soc., 1990.

[Christ]$_{T(b)}$ Christ, M., *A* $T(b)$ *theorem with remarks on analytic capacity and the Cauchy integral*, Colloq. Math. 60/61 (1990), 601–628.

[Coif–Dav–Mey] Coifman, R., David, G. and Meyer, Y. La solution des conjectures de Calderón, *Adv. in Math.* **48** (1983), 144–148.

[Coif–Mey]$_{\rm GTC}$ Coifman, R. and Meyer, Y., Une généralisation du théorème de Calderón sur l'intégrale de Cauchy, in *Fourier Analysis*, Proc. Sem. (El Escorial, 1979), ed. M. de Guzmán and I. Peral, Asoc. Mat. Española, Madrid, 1980.

[Coif–Mey]$_{\rm LC}$ Coifman, R. and Meyer, Y., Lavrentiev's curves and conformal mapping, Institut Mittag-Leffler, Report No. 5, 1983.

[Coif–Mey]$_{\rm TC}$ Coifman, R. and Meyer, Y. Le théorème de Calderón par les méthodes de variable réelle, *C. R. Acad. Sci. Paris*, Series A **289** (1979), 425–428.

[Coif–Weis]$_{\rm AH}$ Coifman, R. and Weiss, G. *Analyse Harmonique Noncommutative sur Certains Espaces Homogènes*, Lecture Notes in Math. **242** (1971), Springer-Verlag.

[Coif–Weis]$_{\rm EHS}$ Coifman, R. and Weiss, G., Extensions of Hardy spaces and their use in analysis, *Bull. Amer. Math. Soc.* **83** (1977), 569–645.

[Col]$_{\rm ARC}$ Colding, T., Aspects of Ricci curvature, *Comparison Geometry*, Math. Sci. Res. Inst. Publ., Cambridge Univ. Press, 1997, 83–98.

[Col]$_{\rm LMPR}$ Colding, T., Large manifolds with positive Ricci curvature, *Inv. Math.* **124** (1996) 193–214.

[Col]$_{\rm RCVC}$ Colding, T., Ricci curvature and volume convergence, *Ann. of Math* (2) 145, 3 (1997), 477-501.

[Col]$_{\rm SRC}$ Colding, T., Stability and ricci curvature, *C.R. Acad. Sci. Paris*, Ser 1 Math 320, 11 (1995), 1343–1347.

[Connes] Connes, A., *Noncommutative Geometry*, Academic Press, 1994.

[Co–Pign] Cooper, D. and Pignataro, T., On the shape of Cantor sets, *J. Diff. Geom.* **28** (1988), 203–221.

[Coul] Coulhon, T., Dimensions at infinity for riemannian manifolds, preprint.

[Cou–Sal] Coulhon, T. and Saloff–Coste, L., Isopérimétrie pour les groupes et les variétés, *Rev. Math. Iber.* **9** (1993), 293–314.

[Croke]$_{4DII}$ Croke, C., The sharp four dimensional isoperimetric inequality, *Comment. Math. Helv.* **59** (1984), 187–192.

[Croke]$_{IIEE}$ Croke, C., Some isoperimetric inequalities and eigenvalue estimates, *Ann. Sci. École Norm. Sup.* **13** (1980), 419–435.

[D'Amb–Gr] D'Ambra, G. and Gromov, M., Lectures on transformation groups: Geometry and dynamics, in *Surveys in Differential Geometry* **1** (1991), 19–111.

[David]$_{CL}$ David, G., *Courbes de Lavrentiev et intégrales singulières,* Thèse de troisième cycle (1981), Université de Paris-Sud, Orsay, France.

[David]$_{MGL}$ David, G., Morceaux de graphes lipschitziens et intégrales singulières sur un surface, *Rev. Mat. Iberoamericana* **4** (1988), 73–114.

[David]$_{OIC}$ David, G., Opérateurs intégraux singuliers sur certaines courbes du plan complexe, *Ann. Sci. École. Norm. Sup.* (4) **17** (1984), 157–189.

[David]$_{OIS}$ David, G., Opérateurs d'intégrale singulière sur les surfaces régulières, *Ann. Sci. École. Norm. Sup.* (4) **21** (1988), 225–258.

[David]$_{WSI}$ David, G., *Wavelets and Singular Integrals on Curves and Surfaces*, Lecture Notes in Math. **1465** (1991), Springer-Verlag.

[Dav–Jer] David, G. and Jerison, D., Lipschitz approximations to hypersurfaces, harmonic measure, and singular integrals, *Indiana Univ. Math. J.* **39** (1990), 831–845.

[Dav–Sem]$_{A_\infty}$ David, G. and Semmes, S., Strong A_∞-weights, Sobolev inequalities, and quasiconformal mappings, in *Analysis and Partial Differential Equations*, edited by C. Sadosky, Lecture Notes in Pure and Applied Mathematics **122**, Marcel Dekker 1990.

[Dav–Sem]$_{\mathrm{AURS}}$ David, G. and Semmes, S., *Analysis of and on Uniformly Rectifiable Sets*, Mathematical Surveys and Monographs **38**, 1993, American Mathematical Society.

[Dav–Sem]$_{\mathrm{QR}}$ David, G. and Semmes, S., Quantitative rectifiability and Lipschitz mappings, *Trans. Amer.Math. Soc.* **337** (1993), 855–889.

[Dav–Sem]$_{\mathrm{QSJD}}$ David, G. and Semmes S., *Quasiminimal surfaces of codimension* 1 *and John domains*, manuscript.

[Dav–Sem]$_{\mathrm{SI}}$ David, G. and Semmes, S., *Singular Integrals and Rectifiable Sets in* $\mathbf{R}^n$*: au-delà des graphes lipschitziens*, *Astérisque* **193**, Société Mathématique de France, 1991.

[Der] Derrik, W.R., A volume-diameter inequality for n-cubes, *J. Analyse Math.* **22** (1969), 1–36.

[Dieu] Dieudonne, J., *Elements d'Analyse*, tome 3, Gauthiers Villars (1970). New York, 1992.

[Din] Dinaburg, E., On the relations among various entropy characteristics of dynamical systems, *Math. USSR Izv.* **5** (1971), 337–378.

[Dorr] Dorronsoro, J., A characterization of potential spaces, *Proc. Amer. Math. Soc.* **95** (1985), 21–31.

[Dougl] Douglas, R., *Banach Algebra Techniques in Operator Theory*, Academic Press, Pure and Applied Mathematics **49**, 1972.

[Dra–Fer] Dranishnikov, A. and Ferry, S., Cell-like images of topological manifolds and limits of manifolds in Gromov–Hausdorff space, preprint.

[Dri–Wil] Van den Dries, L. and Wilkie, A., On Gromov's theorem concerning groups of polynomial growth and elementary logic, *J. Algebra* **89** (1984), 349–374.

[E–H–S] Eberlein, P., Hamenstädt, U., and Schroeder, V., Manifolds of nonpositive curvature, Proc. Symp. Pure Math. **54**, Amer. Math. Soc., 179–228.

[Eck] Eckmann, B., Über die homotopie Gruppen von Gruppenraumen, *Comment. Math. Helv.* **14** (1941), 234–256.

[Ellis] Ellis, R., *Entropy, Large Deviations, and Statistical Mechanics*, Springer, New York, 1985.

[Esk–Farb] Eskin, A. and Farb, B., Quasi-flats and rigidity in higher rank symmetric spaces, to appear, *J. Amer. Math. Soc.*

[Ev–Gar] Evans, L.C. and Gariepy, R., *Measure theory and fine properties of functions*, CRC Press, 1992.

[Fal] Falconer, K., *The Geometry of Fractal Sets*, Cambridge Univ. Press, 1984.

[Fal–Mar] Falconer, K. and Marsh, D., On the Lipschitz equivalence of Cantor sets, *Mathematika* **39** (1992), 223–233.

[Fang–Rong] Fang, F. and Rong X., Positive pinching, volume and homotopy groups, to appear *GAFA*.

[Fed]$_{\mathrm{GMT}}$ Federer, H., *Geometric Measure Theory*, Springer, 1969.

[Fed]$_{\mathrm{VP}}$ Federer, H., Real flat chains, cochains, and variational problems, *Indiana U. Math. J.*, **24** (1974) 351–407.

[Fed–Flem] Federer, H. and Fleming, W.H., Normal and integral currents, *Ann. Math.*, **72** (1960) 458–520.

[Fer]$_{\mathrm{CSHT}}$ Ferry, S., Counting simple homotopy types in Gromov–Hausdorff space, preprint.

[Fer]$_{\mathrm{TFT}}$ Ferry, S., Topological finiteness theorems for manifolds in Gromov- Hausdorff space, *Duke Math. J.* **74** (1994) 95–106.

[Fer]$_{\mathrm{UV}^k}$ Ferry, S., Constructing UV^k maps between spheres, *Proc. Amer. Math. Soc.* 120:1 (1994), 329–332.

[Fer–Ok]$_{\mathrm{ATM}}$ Ferry, S. and Okun, B., Approximating topological metrics by riemannian metrics, *Proc. Amer. Math. Soc.* **123** (1995), 1865–1872.

[Fer–Kar–Munz] Ferus, D., Karcher, H., Munzer, H.F., Clifford algebren und neue isoparametrische Hyperflächen, *Math. Zeit.* **177** (1981), 479–502.

[Fie–Lin–Mil] Fiegel, T., Lindenstrauss, J., Milman, V., The dimension of almost spherical sections of convex bodies, *Acta Math.*, **139** (1977), 53–94.

[Fol–Ste]$_{\mathrm{Est}}$ Folland, G. and Stein, E., Estimates for the $\bar{\partial}_b$-complex and analysis on the Heisenberg group, *Comm. Pure Appl. Math.* **27** (1974), 429–522.

[Fol–Ste]$_{\mathrm{HS}}$ Folland, G. and Stein, E., *Hardy Spaces on Homogeneous Groups*, Math. Notes **28** (1982), Princeton University Press.

[Fr–Gr–Morg] Friedlander, E., Griffiths, P.A., Morgan, J., *Homotopy theory and differential forms* Seminario di Geomertia, Florence, 1972.

[Fuk] Fukaya, K., Hausdorff convergence of Riemannian manifold and its applications, in *Advanced studies in Pure Mathematics* 18-I 143–238, (1990) Recent Topics in Differential and Analytic Geometry, Academic Press.

[Fu]$_{\mathrm{CMCC}}$ Fu, J., Curvature measures and Chern classes of singular varieties, *J. Diff. Geo.* **39** (1994).

[Fu]$_{\mathrm{MAF}}$ Fu, J., Monge–Ampère functions I, II, *Indiana Univ. Math. J.* **38** (1989).

[Ga–La] Gaffney, T. and Lazerfeld, R., On the ramification of branched coverings of P^n, *Inv. Math.* **52:2** (1979), 127–130.

[Gal]$_{\mathrm{ES}}$ Gallot, S., Éstimées de Sobolev quantitatives sur les variétes riemanniennes et applications, *C.R.A.S.* **292** (1981) 375–377.

[Gal]$_{\mathrm{IIA}}$ Gallot, S., Inégalités isopérimétriques et analytiques sur les variétés riemanniennes, *Astérisque*, 1986.

[Gal]$_{\mathrm{Iso}}$ Gallot, S., Isoperimetric inequalities based on integral norms of Ricci curvature, in *Colloque P. Lévy sur les Processus Stochastiques*, Astérisque 157–158 (1988), pp. 191–216.

[Gan–McCann] Gangbo, W. and McCann, R., The geometry of optimal transportation, preprint.

[Garn] Garnett, J., *Bounded Analytic Functions*, Academic Press, 1981.

[Gehring] Gehring, F., The L^p integrability of the partial derivatives of a quasiconformal mapping, *Acta Math.* **130** (1973), 265–277.

[Gersten]$_{\mathrm{II}}$ Gersten, S. M., Isoperimetric and isodiametric functions of finite presentation, *Geometric Group Theory, vol. 1*, G. Niblo and M. Roller, eds. LMS lecture note series.

[Gersten]$_{\mathrm{DF}}$ Gersten, S. M., Dehn functions and l_1-norms of finite presentations, in *Algorithms and classification in combinatorial group theory* (Berkeley, CA, 1989), 195–224, Math. Sci. Res. Inst. Publ., 23, Springer, New York, 1992.

[Godb] Godbillion, G., *Eléments de Topologie Algébrique*, Hermann, 1970.

[Greene] Greene, R., Some concepts and methods in riemannian geometry, Proc. Symp. Pure Math. **54**, Amer. Math. Soc., part 3, 1–22.

[Gr–Wu] Greene, R., and Wu, H.H., Lipschitz convergence of riemannian manifolds, *Pac. J. of Math.* **131** (1988) 119–141.

[Gre] Greenleaf, F., *Invariant means on topological groups and their applications,* Van Nostrand, New York, 1969.

[Grom] Gromoll, D., Spaces of nonnegative curvature, Proc. Symp. Pure Math. **54** Amer. Math. Soc., 337–356.

[Grl–Kl–Mey] Gromoll, D., Klingenberg, W., Meyer, W., *Riemannsche Geometrie im Grossen*, Springer Lecture Notes **55** (1968).

[Gro–Mey] Gromoll, D. and Meyer, W., Periodic geodesics on compact riemannian manifolds, *J. Diff. Geometry* **3** (1969), 493–510.

[Gro]$_{\mathrm{AFM}}$ Gromov, M., Almost flat manifolds, *J. Diff. Geometry* **13** (1978), 231–241.

[Gro]$_{\mathrm{AI}}$ Gromov, M., Asymptotic invariants of infinite groups, in *Geometric Group Theory*, G. Niblo and M. Roller, eds, vol. 2, London Math. Soc. Lecture Notes **182**, Cambridge University Press, Cambridge, 1993.

[Gro]$_{\mathrm{CC}}$ Gromov, M., Carnot–Caratheodory spaces seen from within, preprint IHES M/94/06, 1994.

[Gro]$_{\mathrm{DN}\ell\mathrm{S}}$ Gromov, M., Dimension, nonlinear spectra, and width, in *Geometric Aspects of Functional Analysis*, Lecture Notes in Math. 1317, Springer, New York, 1988. *Séminaire Bourbaki* **546** (1980).

[Gro]$_{\mathrm{FRM}}$ Gromov, M., Filling riemannian manifolds, *J. Diff. Geometry* **18** (1983), 1–147.

[Gro]$_{\mathrm{GPG}}$ Gromov, M., Groups of polynomial growth and expanding maps, *Publ. Math. de l'IHES*, 1980.

[Gro]$_{\mathrm{HED}}$ Gromov, M., Homotopical effects of dilatation, *J. Diff. Geometry* **13** (1978), 303–310.

[Gro]$_{\mathrm{HG}}$ Gromov, M., Hyperbolic groups, in *Essays in Group Theory*, Mathematical Sciences Research Institute Publications **8** (1978), 75–263, Springer-Verlag.

[Gro]$_{\mathrm{HMGA}}$ Gromov, M., Hyperbolic manifolds, groups and actions, in *Riemannian surfaces and related topics, Ann. Math. Studies* **97** (1981), 183–215.

[Gro]$_{\mathrm{HMTJ}}$ Gromov, M., Hyperbolic manifolds according to Thurston and Jorgensen, *Séminaire Bourbaki*, **546** (1980).

[Gro]$_{\mathrm{MIKM}}$ Gromov, M., Metric invariants of Kähler manifolds, in *Diff. Geometry and Topology*, Algero, Italy 20-26 June, ed. Caddeo, Tricerri, World Sci. Publ., (1993), 90–117.

[Gro]$_{\mathrm{MV}}$ Gromov, M., Monotonicity of the volume of intersections of balls, *Geometric Aspects of Functional Analysis*, Lecture Notes in Math. 1267, Springer, New York, 1987.

[Gro]$_{\mathrm{PLII}}$ Gromov, M., Paul Levy's isoperimetric inequality, *Publ. Math. de l'IHES*, (1980).

[Gro]$_{\mathrm{PCMD}}$ Gromov, M., Positive curvature, macroscopic dimension, spectral gaps and higher signatures, in *Functional Analysis on the Eve of the 21st Century*, S. Gindikin, J. Lepowski, and R. Wilson, eds., 1–213.

[Gro]$_{\mathrm{PDR}}$ Gromov, M., *Partial Differential Relations*, Springer, New York, 1986.

[Gro]$_{\mathrm{SGSS}}$ Gromov, M., Spectral geometry of semialgebraic sets, *Ann. Inst. Fourier* **42** (1992), 249–274.

[Gro]$_{\mathrm{sign}}$ Gromov, M., Sign and geometric meaning of curvature, *Rend. Sem. Math. Fis. Milano* **61** (1991), 9–123.

[Gro]$_{\mathrm{SII}}$ Gromov, M., Systoles and intersystolic inequalities, *Actes de la Table Ronde de Géométrie Différentielle en l'honneur de Marcel Berger* (Arthur L. Besse, ed.), Séminaires and Congres of the SMF, No. 1, 1996.

[Gro]$_{\mathrm{Stab}}$ Gromov, M., Stability and pinching, in "Geometry Seminars. Sessions in Topology and Geometry of Manifolds" (Bologna, 1990), 55-97, Univ. Stud. Bologna (1992).

[Gro]$_{\mathrm{VBC}}$ Gromov, M., Volume and bounded cohomology, *Publ. Math. IHES*, **56** (1982), 5–100.

[Gro]$_{\mathrm{Yom}}$ Gromov, M., Entropy, homology, and semialgebraic geometry (after Y. Yomdin), *Sém. Bourbaki* Juin 1986, Astérisque (1987), 225–241.

[Gro–Mil]$_{\mathrm{Brunn}}$ Gromov, M. and Milman, V., Brunn theorem and a concentration of volume phenomenon for symmetric convex bodies, in *Israel seminar on geometrical aspects of functional analysis* (1983/4) Tel Aviv.

[Gro–Mil]$_{\mathrm{GSII}}$ Gromov, M. and Milman, V., Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces, *Compositio Math.* **62** (1987), 263–282.

[Gro–Mil]$_{\mathrm{TAI}}$ Gromov, M. and Milman, V., A topological application of the isoperimetric inequality, *Amer. J. Math.* **105** (1983), 843–854.

[Groups] Duncan, A., Gilbert, N.D., Howie, J., eds, *Combinatorial and Geometric Group Theory*, London Math. Soc. Lecture Notes 204, Edinburgh, 1993.

[Grov]$_{CPT}$ Grove, K., Critical point theory for distance functions, Proc. Symp. Pure Math. **54** Amer. Math. Soc., 357–386.

[Grov]$_{MTM}$ Grove, K., Metric and topological measurements of manifolds, *Proc. Internat. Congress of Math.* (1990).

[Grov–Halp] Grove, K. and Halperin, S., Contributions of rational homotopy theory to global problems in geometry, *Publ. Math. IHES* **56** (1983), 379–385.

[G–K–R] Grove, K., Karcher, H., and Ruh, E., Group actions and curvature, *Inv. Math.* **23** (1974), 31–48.

[Grov–Pet] Grove, K. and Petersen, P., A radius sphere theorem, *Inv. Math.* **112** (1993), 577–583.

[Heb] Hebda, J., The collars of a riemannian manifold and stable isosystolic inequalities, *Pacific J. Math.* **121** (1986), 339–356.

[Heft] Hefter, H., Über die Dehnung von Sphärenabbildungen, *Bonner Math. Schriften* (1980).

[Hein–Kosk]$_{DQC}$ Heinonen, J. and Koskela, P., Definitions of quasiconformality, *Inv. Math.* **120** (1995), 61–79.

[Hein–Kosk]$_{QCG}$ Heinonen, J. and Koskela, P., Quasiconformal maps in metric spaces with controlled geometry, preprint.

[Heint–Kar] Heintze, E. and Karcher, H., A general comparison theorem with applications to volume estimates for submanifolds, *Ann. Ec. Norm. Sup.* **11** (1978), 451-470.

[Helg] Helgason, S., *Differential Geometry and Symmetric Spaces*, Academic Press, New York, 1962.

[Hof–Zehn] Hofer, H. and Zehnder, E., *Symplectic Invariants and Hamiltonian Dynamics*, Birkhäuser Advanced Texts, Basel, 1994.

[Hoff–Spru] Hoffman, D. and Spruck, J., Sobolev and isoperimetric inequalities for riemannian submanifolds, *Comm. Pure Appl. Math.* **27** (1974), 115–204.

[Hur–Wall] Hurewicz, W. and Wallman, H., *Dimension Theory*, Princeton University Press, Princeton, 1948.

[Inc] Incitti, R., An example of limit space of a finitely generated nilpotent group in the sense of Gromov, in *Geometry Seminars*, Univ. Bologna, 1994-95, 155–166.

[Iva] Ivanov, L.D., *Variations of sets and functions,* Nauka, Moscow, 1975.

[Jer] Jerison, D., The Poincaré inequality for vector fields satisfying Hörmander's condition, Duke Math. J. **53** (1986), 503–523.

[John] John, F., Rotation and strain, *Comm. Pure Appl. Math.* **14** (1961), 393–413.

[Johns] Johnson, B.E., Cohomology in Banach Algebras, *Mem. AMS* **127**, 1972.

[Jones]$_{\mathrm{Squ}}$ Jones, P., *Square functions, Cauchy integrals, analytic capacity, and harmonic measure*, Lecture Notes in Math. **1384** (1989), 24–68, Springer-Verlag.

[Jones]$_{\mathrm{Lip}}$ Jones, P., Lipschitz and bi-Lipschitz functions, *Rev. Mat. Iberoamericana* **4** (1988), 115–122.

[Jones]$_{\mathrm{Rec}}$ Jones, P., Rectifiable sets and the travelling salesman problem, *Inv. Math.* **102** (1990), 1–15.

[Jon–Ka–Var] Jones, P., Katz, N., and Vargas, A., *Checkerboards, Lipschitz functions, and uniform rectifiability*, preprint, 1995.

[Jordan] Jordan, C., Mémoire sur les équations différentielles linéaires à intégrales algébriques, *J. Reine Angew. Math.* **84** (1878), 89–215.

[Jour] Journé, J.L., *Calderón-Zygmund Operators, Pseudodifferential Operators, and the Cauchy Integral of Calderón*, Lecture Notes. in Math. **994** (1983), Springer-Verlag.

[Kac] Kac, M., *Probability and related topics in physical sciences,* Lectures in Applied Mathematics, Interscience Publishers, New York, 1957.

[Kan] Kan, D., A combinatorial definition of homotopy groups, *Ann. Math.* **65** (1958), 282–312.

[Ka–Lo–Si] Kannan, R., Lovasz, L., and Simonovits, M., Isoperimetric problems for convex bodies and a localization lemma, *Discrete Comput. Geom.* **13** (1995), 541–559.

[Kantor]$_{\text{trans}}$ Kantorovich, L., On the translocation of masses, *C.R. (Doklady) Acad. Sci. URSS* **37** (1942), 199–201.

[Kantor]$_{\text{Monge}}$ Kantorovich, L., On a problem of Monge, *Uspekhi Math. Nauk.* **3** (1948), 225–226

[Katok] Katok, A., Entropy and closed geodesics, *Ergod. Thy. Dyn. Sys.* **2** (1982), 339–365.

[Katz] Katz, M., Counterexamples to isosystolic inequalities, *Geometriae Dedicata* **57** (1995), 195–206.

[Katz]$_{\text{RFR}}$ Katz, M., The rational filling radius of complex projective space, *Topology and Its Applications* 42 (1991), 201–215.

[Katsu] Katsuda, A., Gromov's convergence theorem and its applications, *Nagoya Math. J.*, **100** (1985), 11–48.

[Katzn] Katznelson, Y., *An Introduction to Harmonic Analysis*, Dover Publications, 1976.

[Kaz–Marg] Kazdan, D.A., and Margulis, G.A., A proof of Selberg's hypothesis, *Mat. Sb.* (N.S.), **75** (117) (1968), 163–168.

[Katz-Suc] Mikhail Katz and Alexander Suciu, Volume of Riemannian Manifolds, Geometric Inequalities, and homotopy theory, in: *Rothenberg Festschrift*, (M. Farber, W. Lueck, S. Weinberger, eds.), Contemporary Mathematics, AMS, 1999. Available at http://xxx.lanl.gov/abs/math.DG/9810172.

[Keen] Keen, L., An extremal length on a torus, *J. d'Analyse Math.*, **19** (1967), 203–206.

[Kier–Kobay] Kiernan, P., and Kobayashi, S., Holomorphic mappings into projective space with lacunary hyperplanes, *Nagoya J. Math.*, **50** (1973), 199–216.

[Kirsz] Kirszbraun, M.D., Über die zussamenziehende und lipschitzsche Transformationen, *Fund. Math.* **22** (1934), 77–108.

[Klein] Klein, P., Über die Kohomologie des freien Schleifenraums, *Bonn. Math. Schriften* **55** (1972).

[Kle] Kleiner, B., An isoperimetric comparison theorem, *Inv. Math.*, **108** (1992), 37–47.

[Kle–Le] Kleiner, B. and Leeb, B., Rigidity of quasi-isometries for symmetric spaces and euclidean buildings, preprint.

[Klin] Klingenberg, W., *Lectures on Closed Geodesics*, Springer, 1977.

[Kobay] Kobayashi, S., *Hyperbolic Manifolds and Holomorphic Mappings*, M. Dekker, New York, 1970.

[Kor–Rei] Korányi, A. and Reimann, M., *Quasiconformal mappings on the Heisenberg group*, Inv. Math. **80** (1985), 309–338.

[Krug–Over] Kruglikov, B. and Overholt, M., The Kobayashi pseudodistance on almost complex manifolds, preprint.

[Kulk] Kulkarni, S., On the Bianchi identities, *Math. Ann.* **199** (1972), 175–204.

[Kur] Kuratowski, C., Quelques problèmes concernant les espaces métriques non-séparables, *Fund. Math.* **25** (1935), 534–545.

[Lan–Sch] Lang, U. and Schroeder, V., Kirszbraun's theorem and metric spaces of bounded curvature, *Geometric and Functional Analysis*, 7:3 (1997), 535–560.

[Laur–Belin] Laurentiev, M. and Belinskii, P., On locally quasiconformal mappings in space, *Contributions to Analysis* Academic Press, 1974.

[Laws] Lawson, B., The stable homology of a flat torus, *Math. Scand.*, **36** (1975), 49–73.

[Ledoux] Ledoux, M., A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold, *Geom. Funct. Anal.* **2** (1992), 221–224.

[Leh] Lehmann, D., *Théorie Homotopique des Formes Différentielles*, Astérisque **45**, 1977.

[Lekerk] Lekerkerker C., *Geometry of Numbers*, Wolters-Noordhoff, 1969.

[Levin] Levin, V. L., General Monge-Kantorovich problem and its applications in measure theory and mathematical economics, in *Functional Analysis, Optimization, and Mathematical Economics*, L. J. Leifman, eds, Oxford University Press, New York, 1990, p. 141–176.

[Levy] Levy, P., *Problèmes Concretes d'Analyse Fonctionelle*, Gauthier–Villars, Paris, 1951.

[Li] Li, P., A lower bound for the first eigenvalue on a compact riemannian manifold, *Indiana U. Math. J.* 28:4 (1972), 1013–1019.

[Li–Tian] Li, P. and Tian, G., On the heat kernel of the Bergmann metric on algebraic varieties, *J. Amer. Math. Soc.* **8** (1995), 857–877.

[Li–Yau] Li, P. and Yau, S., Estimates of eigenvalues of a compact riemannian manifold, in *Geometry of the Laplace operator* Proc. Sympos. Pure Math., XXXVI, Amer. Math. Soc. 1979, pp. 205–239.

[Lichn] Lichnerowicz, A., Applications harmoniques dans un tore, *C.R. Acad. Sci. Paris*, **269**, 912–916.

[Lieb] Lieb, E. H., The stability of matter: from atoms to stars, *Bull. Amer. Math. Soc.* **22** (1990), 1–49.

[Lov] Lovasz, L. and Plummer, MD., *Matching Theory*, North-Holland, Amsterdam (88), 1986.

[Lub] Lubotzky, A., Cayley graphs: eigenvalues, expanders and random walks, in *Surveys in combinatorics* p. 155–189, London Math. Soc. Lecture Note Ser., 218, Cambridge Univ. Press, Cambridge, 1995.

[Lück] Lück, W., L^2-invariants of regular coverings of compact manifolds and CW-complexes, preprint.

[Lus–Fet] Lusternik, L.A. and Fet, A.I., Variational problems on closed manifolds, *Dokl. Akad. Nauk SSSR* (N.S.) **81** (1951) 17–18.

[Lynd–Sch] Lyndon, R.C. and Schupp, P.E., *Combinatorial Group Theory*, Springer, 1977.

[McD–Sal] McDuff, D. and Salamon, D., *J-holomorphic Curves and Quantum Cohomology*, University Lecture Series, Amer. Math. Soc., Providence, 1994.

[McMl] McMullen, C., Lipschitz maps and nets in euclidean space, preprint.

[Mac–Sego] Macias, R. and Segovia, C., Lipschitz functions on spaces of homogeneous type, *Adv. Math.* **33** (1979), 257–270.

[Mann] Manning, A., Topological entropy for geodesic flows, *Ann. Math.* **110** (1979) 567–573.

[Mar–Ri–Vais] Martion, O., Rickman, S., Vaisala, J., Definitions for quasiregular mappings, *Ann. Ac. Sc. Fennicae* **448** (1969), 1–40.

[Massey] Massey, W.S., *Algebraic Topology: An Introduction*, Harcourt, Brace & World, (1967).

[Mattila] Mattila, P., *Geometry of Sets and Measures in Euclidean Spaces*, Cambridge Studies in Advanced Math. **44**, Cambridge Univ. Press., 1995.

[Maz] Maz'ya, V., Classes of domains and embedding theorems for function spaces, *Dokl. Ak. Sci. USSR*, **133** (1960) 527–530 (English translation, 882–884).

[Mic] Mickle, F.J., On the existence of a transformation, *Bull. Amer. Math. Soc.* **55** (1949), 160–164.

[Milm]$_{\mathrm{APF}}$ Milman, V. D., Asymptotic properties of functions of several variables that are defined on homogeneous spaces, *Soviet Math. Dokl.* **12** (1971), 1277–1281.

[Milm]$_{\mathrm{HPL}}$ Milman, V. D., The heritage of P. Lévy in geometrical functional analysis, in *Colloque P. Lévy sur les Processus Stochastiques*, Astérisque 157–158 (1988), pp. 273–301.

[Milm]$_{71}$ Milman, V. D., A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies, *Funkcional. Anal. i Priložen* **5** (1971), 28–37.

[Mil–Sch] Milman, V. and Schechtman, G., *Asymptotic Theory of Finite Dimensional Normed Spaces*, Lecture Notes in Math. 1200, Springer.

[Milnor]$_{\mathrm{CFG}}$ Milnor, J., A note on curvature and the fundamental group, *J. Diff. Geometry*, **2** (1968), 1–7.

[Milnor]$_{\mathrm{MT}}$ Milnor, J., *Morse Theory*, Ann. Math Studies, Princeton Univ. Press, 1962.

[Milnor]$_{\mathrm{TDV}}$ Milnor, J., *Topology from the Differentiable Viewpoint*, U. Virginia Press, 1965.

[Mont–Zip] Montgomery, D. and Zippin, L., *Topological transformation groups*, Interscience, New York, 1955.

[Most] Mostow, G.D., *Strong Rigidity of Locally Symmetric Spaces*, Ann. of Math. Studies, Princeton University Press, 1973.

[Nab] Nabutovski, A., Fundamental group and contractible closed geodesics, *Comm. Pure and Appl. Math.* **49** (1996), 1257–1270.

[Necr] Necrashevich, V., Quasi-isometric nonamenable groups are bi-lipschitz equivalent, preprint, 1997.

[O'Hara]$_{\mathrm{EK}}$ O'Hara, J., Energy of a Knot, *Topology* **30** (1991), 241–247.

[Oliv] Olivier, R., Über die Dehnung von Sphärenabbildungen, *Inv. Math.*, **1** (1966), 380–390.

[Osser] Osserman, R., Isoperimetric inequalities, *Bull. Am. Math. Soc.*, **84** (1978).

[Pan]$_{\mathrm{CBG}}$ Pansu, P., Croissance des boules et des géodesiques fermés dans les nilvariétés, *Ergod. Thy. Dynam. Sys.* **3** (1983), 415–445.

[Pan]$_{\mathrm{CC}}$ Pansu, P., Métriques de Carnot-Carathéodory et quasi-isométries des espaces symétriques de rang un, *Ann. Math.* **129** (1989), 1–60.

[Pan]$_{\mathrm{IIGH}}$ Pansu, P., Une inégalité isopérimétrique sur le groupe d'Heisenberg, *C. R. Acad. Sci.* **295** (1982), 127–131.

[Pan]$_{\text{VCE}}$ Pansu, P., Volume, courbure et entropie (d'après G. Besson, G. Courtois, and S. Gallot), *Sem. Bourbaki* **823** (1996–97).

[Pap]$_{\text{HTBE}}$ Papasoglu, P., Homogeneous trees are bilipschitz equivalent, *Geometriae Dedicata* **54** (1995), 301–306.

[Pap]$_{\text{SQII}}$ Papasoglu, P., On the sub-quadratic isoperimetric inequality, *Geometric group theory* (R. Charney, M. Davis, M. Shapiro, eds) de Gruyter, Berlin, New York, 1995.

[Parth] Parthasarathy, K., *Probability measures on metric spaces*, Academic Press, New York, 1967.

[Paul] Paulin, F., Sur les automorphismes extérieurs des groupes hyperboliques, *Ann. Sci. Ecole Norm. Sup.* **4** (1997), 147–167.

[Per]$_{\text{Wid}}$ Perelman, G., Widths of nonnegatively curved spaces, *Geometric and Functional Analysis* **5** (1995) 445–463.

[Per]$_{\text{SCBB}}$ Perelman, G., Spaces with curvature bounded below, in *Proc. ICM*, S. Chatterji, ed., Birkhäuser Verlag, Basel, 1995, pp. 517–525.

[Peters] Peters, S., Convergence of riemannian manifolds, *Compositio Math.*, **62** (1987), 3–16.

[Pet] Petersen, P., Gromov–Hausdorff convergence of metric spaces, Proc. Symp. Pure Math. **54** Amer. Math. Soc., 489–594.

[Pet–Tus] Petrunin, A., Tuschmann, W., Diffeomorphism finiteness, positive pinching, and second homotopy, to appear in *GAFA*.

[Pet–Rong–Tus] Petrunin, A., Rong X., Tuschmann, W., Collapsing vs positive pinching, to appear in *GAFA*.

[Pit] Pittet, C., Systoles on $S^1 \times S^n$, *Diff. Geom. Appl.* **7** (1997), 139–142.

[Pit–Sal] Pittet, C. and Saloff–Coste, L., Amenable groups, isoperimetric profiles, and random walks, preprint.

[Pr] Proust, M., *À la Recherche du Temps Perdu*, Pléiade, 1954.

[Pu] Pu, P.M., Some inequalities in certain nonorientable riemannian manifolds, *Pac. J. of Math.*, **2** (1952), 55–71.

[Pugh] Pugh, C.C., The $C^{1,1}$ conclusions in Gromov's theory, *Ergod. Thy. Dynam. Sys.*, **7** (1987), 133–147.

[Ragun] Ragunathan, M.S., *Discrete subgroups of Lie groups*, Springer, 1972.

[Rickm] Rickman, S., *Quasiregular Mappings*, Ergeb. der Math. vol. 26, Springer, New York, 1993.

[Rinow] Rinow, W., *Die Innere Geometrie der Metrischen Räume*, Springer, New York, 1961.

[Rom] Romanovskii, N.S., Free subgroups of finitely presented groups, *Algebra i Logika* **16** (1977), 88–97.

[Rudin] Rudin, W., *Real and Complex Analysis*, McGraw-Hill, 1966.

[Sak] Sakai, T., Comparison and finiteness theorems in riemannian geometry, in *Geometry of Geodesics and Related Topics* (Tokyo 1992), North-Holland, Amsterdam 1984, pp. 125-181.

[Samuel] Samuel, P., *Théorie Algébrique des Nombres*, Hermann, 1967.

[Sar] Sarason, D., *Algebras of functions on the unit circle*, Bull. Amer. Math. **79** (1973), 286–299.

[Sch–Sch] Schechtman, G. and Schmuckenschläger, M., Another remark on the volume of the intersection of two L_p^n balls, *Lecture Notes in Math.*, vol. 1469, Springer, New York, 1991, 174–178.

[Sem]$_{\mathrm{AGRect}}$ Semmes, S., Analysis vs. geometry on a class of rectifiable hypersurfaces in $\mathbb{R}^n$, *Indiana Univ. Math. J.* **39** (1990), 1005–1035.

[Sem]$_{\mathrm{Bil}}$ Semmes, S., Bilipschitz mappings and strong A_∞ weights, *Ann. Acad. Sci. Fenn.* A I **18** (1993), 211–248.

[Sem]$_{\mathrm{BMO}}$ Semmes, S., Hypersurfaces in $\mathbf{R}^n$ whose unit normal has small BMO norm, *Proc. Amer. Math. Soc.* **112** (1991), 403–412.

[Sem]$_{\mathrm{CBSI}}$ Semmes, S., A criterion for the boundedness of singular integrals on hypersurfaces, *Trans. Amer. Math. Soc.* **311** (1989), 501–513.

[Sem]$_{\mathrm{CIQM}}$ Semmes, S., The Cauchy integral, chord-arc curves, and quasiconformal mappings, in *The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof*, edited by A. Baernstein, D. Drasin, P. Duren, and A. Marden, Amer. Math. Soc. Surveys and Monographs **21** (1986), 167–183.

[Sem]$_{\mathrm{Diff}}$ Semmes, S., Differentiable function theory on hypersurfaces in $\mathbb{R}^n$ (without bounds on their smoothness), *Indiana Univ. Math. J.* **39** (1990), 985–1004.

[Sem]$_{\mathrm{E}\bar{\partial}}$ Semmes, S., Estimates for $(\overline{\partial}-\mu\partial)^{-1}$ and Calderón's theorem on the Cauchy integral, *Trans. Amer. Math. Soc.* **306** (1988), 191–232.

[Sem]$_{\mathrm{FCGSQT}}$ Semmes, S., Finding curves on general spaces through quantitative topology with applications for Sobolev and Poincaré inequalities, *Selecta Math*, 2:2 (1996), 155–295.

[Sem]$_{\mathrm{Find}}$ Semmes, S., Finding structure in sets with little smoothness, *Proc. I.C.M.*, S. Chatterji, ed., Birkhäuser Verlag, 1995, pp. 875–885.

[Sem]$_{\mathrm{Good}}$ Semmes, S., Good metric spaces without good parameterizations, *Revista Matemática Iberoamericana* 12:1 (1996), 187–275.

[Sem]$_{\mathrm{Map}}$ Semmes, S., Mappings and spaces, to appear in the volume dedicated to Fred Gehring, Springer-Verlag, 1998, pp 347–368.

[Sem]$_{\mathrm{NBP}}$ Semmes, S., On the nonexistence of bilipschitz parameterizations and geometric problems about A_∞ weights, *Revista Matemática Iberoamericana* 12:2 (1996), 337–410.

[Sem]$_{\mathrm{QM}}$ Semmes, S., Quasiconformal mappings and chord-arc curves, *Trans. Amer. Math. Soc.* **306** (1988), 233–263.

[Sem]$_{\mathrm{Quasi}}$ Semmes, S., Quasisymmetry, measure, and a question of Heinonen, *Revista Matemática Iberoamericana*, 12:3 (1996), 727–781.

[Sem]$_{\mathrm{Rmks}}$ Semmes, S., Some remarks about metric spaces, spherical mappings, functions and their derivatives, *Publ. Math.* 40: 2 (1996), 411–430.

[Sem]$_{SC1}$ Semmes, S., Chord-arc surfaces with small constant I, *Adv. Math.* **85** (1991), 198–223.

[Sem]$_{SC2}$ Semmes, S., Chord-arc surfaces with small constant II: Good parameterizations, *Adv. Math.* **88** (1991), 170–199.

[Serre] Serre, J.-P., Homologie singulière des espaces fibrés, *Ann. Math.*, **54** (1951).

[Serre]$_{GH}$ Serre, J.-P., Groupes d'homotopie et classes de groupes abeliens, Ann. Math. **58** (1953), 258–294.

[Sém Cartan] *Séminaire H. Cartan*, Année 49-50, exposé no. 3.

[Sha–Ya] Sha, J.-P. and Yang, D., Positive Ricci curvature on compact simply connected 4-manifolds, in *Differential Geometry*, R. Greene and S.-T. Yau, eds, Amer. Math. Soc. 1993, 529–538.

[Shik] Shikata, Y., On the differentiable pinching problem, *Osaka Math. J.* **4** (1967), 279–287.

[Shio] Shioya, T., The limit spaces of two-dimensional manifolds with uniformly bounded integral curvature, preprint.

[Simon] Simon, L., *Lectures on Geometric Measure Theory*, Proceedings of the Centre for Mathematical Analysis, Australian National University, Vol. 3, 1983.

[Sing] Singer, I.M., Infinitesimally homogeneous spaces, *Comm. Pure Appl. Math.*, **13** (1960), 685–687.

[Span] Spanier, E., *Algebraic Topology*, McGraw-Hill, 1966.

[Stall] Stallings, J., Homology and central series of groups, *J. of Algebra* **2** (1965), 170–181.

[Stein]$_{HA}$ Stein, E.M. *Harmonic Analysis : Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, 1993.

[Stein]$_{SI}$ Stein, E.M., *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[Stein–Weiss] Stein, E.M. and Weiss, G.L., *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, 1971.

[Sul]$_{\mathrm{Cyc}}$ Sullivan, D., Cycles for the dynamical study of foliated manifolds and complex manifolds, *Inv. Math.*, **36** (1976) 225–255.

[Sul]$_{\mathrm{DF}}$ Sullivan, D., Differential forms and the topology of manifolds, *Proc. Int. Conf. on Manifolds and Related Topics*, Tokyo, 1973.

[Sul]$_{\mathrm{ICT}}$ Sullivan, D., Infinitesimal computations in topology, *Publ. Math. IHES*, (1978), 269–331.

[Sul–Vig] Sullivan, D. and Vigue, M., The homology theory of the closed geodesic problem, *J. Diff. Geometry*, **11** (1976), 633–644.

[Šva] Švarc, A., A volume invariant of coverings, *Dokl. Akad. Nauk. SSSR* **105** (1955), 32–34.

[Tal]$_{\mathrm{CMII}}$ Talagrand, M., Concentration of measure and isoperimetric inequalities in product spaces, *Inst. Hautes Études Sci Publ. Math.* **81** (1995), 73–205.

[Tal]$_{\mathrm{IIA}}$ Talagrand, M., Some isoperimetric inequalities and their applications, in *Proceedings of the International Congress of Mathematicians*, Vol. I, II (Kyoto, 1990), 1011–1024, Math. Soc. Japan, Tokyo, 1991.

[Tal]$_{\mathrm{NII}}$ Talagrand, M., A new isoperimetric inequality and the concentration of measure phenomenon, in *Geometric aspects of functional analysis* (1989- 90), 94–124, Lecture Notes in Math., 1469, Springer, Berlin, 1991.

[Tal]$_{\mathrm{NLI}}$ Talagrand, M., A new look at Independence, *Ann. Prob.* **23** (1996), 1–37.

[Thom] Thom, R., Quelques propriétés globales des variétés différentiables, *Comment. Math. Helv.* 28 (1954), 17–86.

[Thurs] Thurston, W.P., *The Geometry and Topology of 3-Manifolds*, Princeton Lecture Notes 1978.

[Top] Topogonov, V., Riemann spaces with curvature bounded below, *Uspehi Math. Nauk.* **14**.

[Toro]$_{\mathrm{BP}}$ Toro, T., Geometric conditions and existence of bilipschitz parameterizations, *Duke Math. J.* **77** (1995), 193–227.

[Toro]$_{\mathrm{SFF}}$ Toro, T., Surfaces with generalized second fundamental form in L^2 are Lipschitz manifolds, *J. Diff. Geom.* **39** (1994), 65–102.

[Tri–Van] Tricerri, F. and Vanhecke, L., Curvature homogeneous riemannian manifolds, *Ann. Sci. École Norm. Sup.* **22** (1989), 535–554.

[Tukia] Tukia, P., *Hausdorff dimension and quasisymmetric mappings*, Math. Scand. **65** (1989), 152–160.

[Tuk–Vais] Tukia, P. and Väisälä, J., Quasisymmetric embeddings of metric spaces, *Ann. Acad. Sci. Fenn.* Ser. A I Math. 5 (1980), 97–114.

[Ulam] Ulam, S., *A Collection of Mathematical Problems*, New York, 1960.

[Ury] Uryson, P.S., Sur un espace métrique universel, *Bull. de Sciences Math.* **5** (1927), 1–38.

[Var] Varopoulos, N. Th., *Analysis and Geometry on Groups*, preprint, Paris VI.

[Va–Sa–Co] Varopoulous, N. Th., Saloff–Coste, L., and Coulhon, T., *Analysis and Geometry on Groups*, Camb. Univ. Press, Cambridge, 1992.

[Weil] Weil, A., *Introduction à l'Étude des Variétés Kählériennes*, Hermann, 1958.

[White] White, P., Regular convergence, *Bull. Amer. Math. Soc.* **60** (1954), 431–443.

[Whit] Whitney, H., *Geometric Integration Theory*, Princeton Univ. Press, 1957.

[Why] Whyburn, H., *Analytical Topology*, 1946.

[Wolf]$_{\mathrm{Gro}}$ Wolf, J.A., Growth of finitely generated solvable groups and curvature of riemannian manifolds, *J. Diff. Geometry*, **2** (1968), 421–446.

[Wolf]$_{\mathrm{SCC}}$ Wolf, J.A., *Spaces of Constant Curvature*, Publish or Perish.

[Yau]$_{\text{RIG}}$ Yau, S.T., Remarks on the isometry group of a riemannian manifold, *Topology*, **16** (1977), 239–247.

[Yau]$_{\text{ICFE}}$ Yau, S.T., Isoperimetric constants and the first eigenvalue of a compact manifold, *Ann. Ec. Norm. Sup.*, **8** (1975), 487–507.

[Yau]$_{\text{FTPCRM}}$ Yau, S.T., Some function theoretic properties of complete riemannian manifolds and their applications to geometry, *Indiana U. Math. J.*, **25** (1976), 659–670.

[Zalg] Zalgaller, V.A., Isometric embeddings of polyhedra, *Dokl. Acad. Nauk. SSSR*, **123** (1958), 599–601.

[Zassen] Zassenhaus, H., Beweis eines Satzes über diskrete Gruppen, *Abhand. Math. Sem. Hansische Univ.*, **12** (1938), 282–312.

[Zor] Zoritch, V.A., M.A. Laurentiev's theorem on quasiconformal space maps, *Mat. Sb.* **74** (1967) 417 or *Math. USSR Sb.* **3** (1967).

[Zor–Kes] Zorich, V. A. and Keselman, V. M., On the conformal type of a Riemannian manifold, *Funktsional. Anal. i Prilozhen.* **30** (1996), 40–55.

[Zyg] Zygmund, A., *Trigonometric Series*, Volumes I and II, Cambridge University Press, 1979.

Glossary of Notation

$\mathrm{ObsInv}_Y(X, -\kappa)$, $3\frac{1}{2}{}_+$.F

$\mathcal{P}$, $3\frac{1}{2}{}_+$.51

$\mathrm{Rad}(X)$, 3.28

$\mathrm{Sep}(X; \kappa_0, \dots, \kappa_N)$, $3\frac{1}{2}{}_+$.30

$\mathrm{SpecDiam}(\mu)$, $3\frac{1}{2}{}_+$.37

$\mathrm{Supp}\mu$,$3\frac{1}{2}{}_+$.2

$\mathrm{syst}_k(V)$, 4.40

$\mathrm{Tk}\, X(\rho, \kappa)$, $3\frac{1}{2}{}_+$.8

$\mathrm{Tn}\, X(\rho, \kappa)$, $3\frac{1}{2}{}_+$.8

Tra_λ, $3\frac{1}{2}{}_+$.10

$\mathrm{vol}(f|_A)$, 2.7

$\mathrm{vol}(f)$, 2.7

$|\mathrm{vol}|(f)$, $2.8\frac{1}{2}{}_+$

$\mathrm{vol}^{(r)}$, $3.34\frac{1}{2}{}_+$

$\mathcal{X}$, $3\frac{1}{2}{}_+$.D

Index