美国数学会经典影印系列

AMS
AMERICAN MATHEMATICAL SOCIETY

Poincaré's Legacies, Part I

pages from year two of a mathematical blog

庞加莱的遗产，第Ⅰ部分

第二年的数学博客选文

Terence Tao

高等教育出版社

美国数学会经典影印系列

AMS
AMERICAN MATHEMATICAL SOCIETY

Poincaré's Legacies, Part I

pages from y
of a mathematical blog

庞加莱的遗产，第I部分

第二年的数学博客选文

Terence Tao

高等教育出版社·北京

图字：01-2016-2489 号

庞加莱的遗产，
第 I 部分
Pangjialai de Yichan, Di I Bufen

图书在版编目 (CIP) 数据

庞加莱的遗产 . 第 I 部分，第二年的数学博客选文 = Poincare's Legacies, Part I: pages from year two of a mathematical blog : 英文 / (澳) 陶哲轩 (Terence Tao) 著 . — 影印本 . — 北京 : 高等教育出版社，2017.4
ISBN 978-7-04-046995-0
Ⅰ. ①庞… Ⅱ. ①陶… Ⅲ. ①数学—文集—英文
Ⅳ. ①O1-53
中国版本图书馆 CIP 数据核字 (2017) 第 001230 号

策划编辑 李 鹏　　责任编辑 李 鹏
封面设计 张申申　　责任印制 毛斯璐

出版发行 高等教育出版社
社址 北京市西城区德外大街4号
邮政编码 100120
购书热线 010-58581118
咨询电话 400-810-0598
网址 http://www.hep.edu.cn
http://www.hep.com.cn
网上订购 http://www.hepmall.com.cn
http://www.hepmall.com
http://www.hepmall.cn
印刷:北京新华印刷有限公司

开本 787mm × 1092mm 1/16
印张 19.5
字数 490 千字
版次 2017 年 4 月第 1 版
印次 2017 年 4 月第 1 次印刷
定价 135.00元

[物 料 号 46995-00]

美国数学会经典影印系列

出版者的话

近年来，我国的科学技术取得了长足进步，特别是在数学等自然科学基础领域不断涌现出一流的研究成果。与此同时，国内的科研队伍与国外的交流合作也越来越密切，越来越多的科研工作者可以熟练地阅读英文文献，并在国际顶级期刊发表英文学术文章，在国外出版社出版英文学术著作。

然而，在国内阅读海外原版英文图书仍不是非常便捷。一方面，这些原版图书主要集中在科技、教育比较发达的大中城市的大型综合图书馆以及科研院所的资料室中，普通读者借阅不甚容易；另一方面，原版书价格昂贵，动辄上百美元，购买也很不方便。这极大地限制了科技工作者对于国外先进科学技术知识的获取，间接阻碍了我国科技的发展。

高等教育出版社本着植根教育、弘扬学术的宗旨服务我国广大科技和教育工作者，同美国数学会（American Mathematical Society）合作，在征求海内外众多专家学者意见的基础上，精选该学会近年出版的数十种专业著作，组织出版了“美国数学会经典影印系列”丛书。美国数学会创建于1888年，是国际上极具影响力的专业学术组织，目前拥有近30000会员和580余个机构成员，出版图书3500多种，冯·诺依曼、莱夫谢茨、陶哲轩等世界级数学大家都是其作者。本影印系列涵盖了代数、几何、分析、方程、拓扑、概率、动力系统等所有主要数学分支以及新近发展的数学主题。

我们希望这套书的出版，能够对国内的科研工作者、教育工作者以及青年学生起到重要的学术引领作用，也希望今后能有更多的海外优秀英文著作被介绍到中国。

高 等 教 育 出 版 社

2016年12月

To Garth Gaudry, who set me on the road;
To my family, for their constant support;
And to the readers of my blog, for their feedback and contributions.

Preface

In February of 2007, I converted my "What's new" web page of research updates into a blog at `terrytao.wordpress.com`. This blog has since grown and evolved to cover a wide variety of mathematical topics, ranging from my own research updates, to lectures and guest posts by other mathematicians, to open problems, to class lecture notes, to expository articles at both basic and advanced levels.

With the encouragement of my blog readers, and also of the AMS, I published many of the mathematical articles from the first year (2007) of the blog as [**Ta2008b**], which will henceforth be referred to as *Structure and Randomness* throughout this book. This gave me the opportunity to improve and update these articles to a publishable (and citeable) standard, and also to record some of the substantive feedback I had received on these articles from the readers of the blog. Given the success of the blog experiment so far, I am now doing the same for the second year (2008) of articles from the blog. This year, the amount of material is large enough that the blog will be published in two volumes.

As with *Structure and Randomness*, each part begins with a collection of expository articles, ranging in level from completely elementary logic puzzles to remarks on recent research, which are only loosely related to each other and to the rest of the book. However, in contrast to the previous book, the bulk of these volumes is dominated by the lecture notes for two graduate courses I gave during the year. The two courses stemmed from two very different but fundamental contributions to mathematics by Henri Poincaré, which explains the title of the book.

This is the first of the two volumes, and it focuses on ergodic theory, combinatorics, and number theory. In particular, Chapter 2 contains the lecture notes for my course on *topological dynamics and ergodic theory*, which originated in part from Poincaré's pioneering work in chaotic dynamical systems. Many situations in mathematics, physics, or other sciences can be modeled by a discrete or continuous *dynamical system*, which at its most abstract level is simply a space X, together with a shift $T : X \to X$ (or family of shifts) acting on that space, and possibly preserving either the topological or measure-theoretic structure of that space. At this level of generality, there are a countless variety of dynamical systems available for study, and it may seem hopeless to say much of interest without specialising to much more concrete systems. Nevertheless, there is a remarkable phenomenon that dynamical systems can largely be classified into "structured" (or "periodic") components, and "random" (or "mixing") components,[1] which then can be used to prove various *recurrence theorems* that apply to very large classes of dynamical systems, not the least of which is the *Furstenberg multiple recurrence theorem* (Theorem 2.10.3). By means of various *correspondence principles*, these recurrence theorems can then be used to prove some deep theorems in combinatorics and other areas of mathematics, in particular yielding one of the shortest known proofs of *Szemerédi's theorem* (Theorem 2.10.1) that all sets of integers of positive upper density contain arbitrarily long arithmetic progressions. The road to these recurrence theorems, and several related topics (e.g. ergodicity, and Ratner's theorem on the equidistribution of unipotent orbits in homogeneous spaces) will occupy the bulk of this course. I was able to cover all but the last two sections in a 10-week course at UCLA, using the exercises provided within the notes to assess the students (who were generally second or third-year graduate students, having already taken a course or two in graduate real analysis).

Finally, I close this volume with a third (and largely unrelated) topic (Chapter 3), namely a series of lectures on recent developments in additive prime number theory, both by myself and my coauthors, and by others. These lectures are derived from a lecture I gave at the annual meeting of the AMS at San Diego in January of 2007, as well as a lecture series I gave at Penn State University in November 2007.

A remark on notation

For reasons of space, we will not be able to define every single mathematical term that we use in this book. If a term is italicised for reasons other than emphasis or definition, then it denotes a standard mathematical object, result, or concept, which can be easily looked up in any number of references.

[1] One also has to consider *extensions* of systems of one type by another, e.g. mixing extensions of periodic systems; see Section 2.15 for a precise statement.

(In the blog version of the book, many of these terms were linked to their Wikipedia pages, or other on-line reference pages.)

I will however mention a few notational conventions that I will use throughout. The cardinality of a finite set E will be denoted $|E|$. We will use the asymptotic notation $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote the estimate $|X| \leq CY$ for some absolute constant $C > 0$. In some cases we will need this constant C to depend on a parameter (e.g. d), in which case we shall indicate this dependence by subscripts, e.g. $X = O_d(Y)$ or $X \ll_d Y$. We also sometimes use $X \sim Y$ as a synonym for $X \ll Y \ll X$.

In many situations there will be a large parameter n that goes off to infinity. When that occurs, we also use the notation $o_{n\to\infty}(X)$ or simply $o(X)$ to denote any quantity bounded in magnitude by $c(n)X$, where $c(n)$ is a function depending only on n that goes to zero as n goes to infinity. If we need $c(n)$ to depend on another parameter, e.g. d, we indicate this by further subscripts, e.g. $o_{n\to\infty;d}(X)$.

We will occasionally use the averaging notation

$$\mathbf{E}_{x\in X} f(x) := \frac{1}{|X|}\sum_{x\in X} f(x)$$

to denote the average value of a function $f : X \to \mathbf{C}$ on a non-empty finite set X.

Acknowledgments

The author is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman award.

Contents

Chapter 1

Expository Articles

1.1. The blue-eyed islanders puzzle

This is one of my favourite logic puzzles. It has a number of formulations, but I will use this one:

Problem 1.1.1. There is an island upon which a tribe resides. The tribe consists of 1000 people, with various eye colours. Yet, their religion forbids them to know their own eye color, or even to discuss the topic; thus, each resident can (and does) see the eye colors of all other residents, but has no way of discovering his or her own (there are no reflective surfaces). If a tribesperson does discover his or her own eye color, then their religion compels them to commit ritual suicide at noon the following day in the village square for all to witness. All the tribespeople are highly logical[1] and devout, and they all know that each other is also highly logical and devout (and they all know that they all know that each other is highly logical and devout, and so forth).

Of the 1000 islanders, it turns out that 100 of them have blue eyes and 900 of them have brown eyes, although the islanders are not initially aware of these statistics (each of them can of course only see 999 of the 1000 tribespeople).

One day, a blue-eyed foreigner visits the island and wins the complete trust of the tribe.

One evening, he addresses the entire tribe to thank them for their hospitality.

[1] For the purposes of this logic puzzle, "highly logical" means that any conclusion that can be logically deduced from the information and observations available to an islander, will automatically be known to that islander.

However, not knowing the customs, the foreigner makes the mistake of mentioning eye color in his address, remarking how unusual it is to see another blue-eyed person like myself in this region of the world.

What effect, if anything, does this *faux pas* have on the tribe?

The interesting thing about this puzzle is that there are two quite plausible arguments here, which give opposing conclusions:

Argument 1. The foreigner has no effect, because his comments do not tell the tribe anything that they do not already know (everyone in the tribe can already see that there are several blue-eyed people in their tribe). □

Argument 2. 100 days after the address, all the blue eyed people commit suicide. This is proven as a special case of Proposition 1.1.2 below. □

Proposition 1.1.2. *Suppose that the tribe had n blue-eyed people for some positive integer n. Then n days after the traveller's address, all n blue-eyed people commit suicide.*

Proof. We induct on n. When $n = 1$, the single blue-eyed person realizes that the traveler is referring to him or her, and thus commits suicide on the next day. Now suppose inductively that n is larger than 1. Each blue-eyed person will reason as follows: "If I am not blue-eyed, then there will only be $n-1$ blue-eyed people on this island, and so they will all commit suicide $n-1$ days after the traveler's address. But when $n-1$ days pass, none of the blue-eyed people do so (because at that stage they have no evidence that they themselves are blue-eyed). After nobody commits suicide on the $(n-1)^{\text{st}}$ day, each of the blue eyed people then realizes that they themselves must have blue eyes, and will then commit suicide on the n^{th} day." □

Which argument is logically valid? Or are the hypotheses of the puzzle logically impossible to satisfy?[2]

Notes. I will not spoil the solution to this puzzle in this article; but one can find much discussion on this problem at the comments to the web page for this puzzle, at `terrytao.wordpress.com/2008/02/05`. See also `xkcd.com/blue_eyes.html` for some further discussion.

1.2. Kleiner's proof of Gromov's theorem

In this article, I would like to present the recent simplified proof by Kleiner [**Kl2007**] of the celebrated theorem of Gromov [**Gr1981**] on groups of polynomial growth.

[2]Note that this is not the same as the hypotheses being extremely implausible, which of course they are.

Let G be an at most countable group generated by a finite set S of generators, which we can take to be symmetric (i.e., $s^{-1} \in S$ whenever $s \in S$). Then we can form the *Cayley graph* Γ, whose vertices are the elements of G, with g and gs connected by an edge for every $g \in G$ and $s \in S$. This is a connected regular graph, with a transitive left action of G. For any vertex x and $R > 0$, one can define the ball $B(x, R)$ in Γ to be the set of all vertices connected to x by a path of length at most R. We say that G has *polynomial growth* if we have the bound $|B(x, R)| = O(R^{O(1)})$ as $R \to \infty$; one can easily show that the left-hand side is independent of x, and that the polynomial growth property does not depend on the choice of generating set S.

Examples of finitely generated groups of polynomial growth include

(1) Finite groups;

(2) Abelian groups (e.g. $\mathbf{Z}^d$);

(3) Nilpotent groups (a generalisation of (2));

(4) *Virtually nilpotent* groups, i.e., those that have a nilpotent subgroup of finite index (a combination of (1) and (3)).

In [**Gr1981**], Gromov proved that these are the only examples:

Theorem 1.2.1 (Gromov's theorem). *Let G be a finitely generated group of polynomial growth. Then G is virtually nilpotent.*

Gromov's original argument used a number of deep tools, including the Montgomery-Zippin-Yamabe [**MoZi1955**] structure theory of locally compact groups (related to *Hilbert's fifth problem*), as well as various earlier partial results on groups of polynomial growth. Several proofs have subsequently been found. Recently, Kleiner [**Kl2007**] obtained a proof which was significantly more elementary, although it still relies on some non-trivial partial versions of Gromov's theorem. Specifically, it needs the following result proven by Wolf [**Wo1968**] and by Milnor [**Mi1968**]:

Theorem 1.2.2 (Gromov's theorem for virtually solvable groups). *Let G be a finitely generated group of polynomial growth which is* virtually solvable *(i.e., it has a solvable subgroup of finite index). Then it is virtually nilpotent.*

The argument also needs a related result:

Theorem 1.2.3. *Let G be a finitely generated amenable*[3] *group which is linear, thus $G \subset GL_n(\mathbf{C})$ for some n. Then G is virtually solvable.*

[3] In this context, one definition of amenability is that G contains a *Følner sequence* $F_1, F_2, \ldots$ of finite sets, thus $\bigcup_{n=1}^{\infty} F_n = G$ and $\lim_{n\to\infty} |gF_n \Delta F_n|/|F_n| = 0$ for all $g \in G$.

This theorem is an immediate consequence of the Tits alternative [**Ti1972**], but also has a short elementary proof, due to Shalom [**Sh1998**]. An easy application of the pigeonhole principle to the sequence $|B(x, R)|$ for $R = 1, 2, \ldots$ shows that every group of polynomial growth is amenable. Thus Theorems 1.2.2 and 1.2.3 already give Gromov's theorem for linear groups.

Other than the use of Theorems 1.2.2 and 1.2.3, Kleiner's proof of Theorem 1.2.1 is essentially self-contained. The argument also extends to groups of *weakly polynomial growth*, which means that $|B(x, R)| = O(R^{O(1)})$ for some sequence of radii R going to infinity. (This extension of Gromov's theorem was first established in [**vdDrWi1984**].) But for simplicity we only discuss the polynomial growth case here.

1.2.1. Reductions. The first few reductions follow the lines of Gromov's original argument. The first observation is that it suffices to exhibit an infinite abelianisation of G, or more specifically to prove:

Proposition 1.2.4 (Existence of infinite abelian representation). *Let G be an infinite finitely generated group of polynomial growth. Then there exists a subgroup G' of finite index whose abelianisation $G'/[G', G']$ is infinite.*

Indeed, if G' has infinite abelianisation, then one can find a non-trivial homomorphism $\alpha : G' \to \mathbf{Z}$. The kernel K of this homomorphism is a normal subgroup of G'. Using the polynomial growth hypothesis, one can show that K is also finitely generated; furthermore, it is of polynomial growth of one lower order (i.e., the exponent in the $O(R^{O(1)})$ bound for $|B(x, R)|$ is reduced by 1). An induction hypothesis then gives that K is virtually nilpotent, which easily implies that G' (and thus G) is virtually solvable. Gromov's theorem for infinite G then follows from Theorem 1.2.2. (The theorem is of course trivial for finite G.)

Remark 1.2.5. The above argument not only shows that G is virtually solvable, but moreover that G' is the semidirect product $K \rtimes_\phi \mathbf{Z}$ of a virtually nilpotent group K and the integers, which acts on K by some automorphism ϕ. Thus one does not actually need the full strength of Theorem 1.2.2 here, but only the special case of semidirect products of the above form. In fact, most proofs of Theorem 1.2.2 proceed by reducing to this sort of case anyway.

To prove Proposition 1.2.4, it suffices to prove

Proposition 1.2.6 (Existence of infinite linear representation). *Let G be an infinite finitely generated group of polynomial growth. Then there exists a finite-dimensional representation $\rho : G \to GL_n(\mathbf{C})$ whose image $\rho(G)$ is infinite.*

Indeed, the image $\rho(G) \subset GL_n(\mathbf{C})$ is also finitely generated with polynomial growth, and hence, by Theorems 1.2.3 and 1.2.2, is virtually nilpotent (actually, for this argument we do not need Theorem 1.2.2 and would be content with virtual solvability). If the abelianisation of $\rho(G)$ is finite, one can easily pass to a subgroup G' of finite index and reduce the (virtual) step of $\rho(G')$ by 1, so one can quickly reduce to the case when the abelianisation is infinite, at which point Proposition 1.2.4 follows. So all we need to do now is to prove Proposition 1.2.6.

1.2.2. Harmonic functions on Cayley graphs. Kleiner's approach to Proposition 1.2.6 relies on the notion of a (possibly vector-valued) *harmonic function* on the Cayley graph Γ. This is a function $f : G \to H$ taking values in a Hilbert space H such that $f(g) = \frac{1}{|S|} \sum_{s \in S} f(gs)$ for all $g \in G$. Formally, harmonic functions are local minimisers of the energy functional

$$E(f) := \frac{1}{2} \sum_{g \in G} |\nabla f(g)|^2,$$

where

$$|\nabla f(g)|^2 := \frac{1}{|S|} \sum_{s \in S} \|f(gs) - f(g)\|_H^2,$$

though of course with the caveat that $E(f)$ is often infinite. (This property is also equivalent to a certain graph Laplacian of f vanishing.)

Of course, every constant function is harmonic. But there are other harmonic functions too: for instance, on $\mathbf{Z}^d$, any linear function is harmonic (regardless of the actual choice of generators). Kleiner's proof of Proposition 1.2.6 follows by combining the following two results:

Proposition 1.2.7. *Let G be an infinite finitely generated group of polynomial growth. Then there exists an (affine-) isometric (left) action of G on a Hilbert space H with no fixed points, and a harmonic map $f : G \to H$ which is G-equivariant (thus $f(gh) = gf(h)$ for all $g, h \in G$). (Note that in view of equivariance and the absence of fixed points, this harmonic map is necessarily non-constant.)*

Proposition 1.2.8. *Let G be a finitely generated group of polynomial growth, and let $d \geq 0$. Then the linear space of harmonic functions $u : G \to \mathbf{R}$ with growth of order at most d (thus $u(g) = O(R^d)$ on $B(\mathrm{id}, R)$) is finite-dimensional.*

Indeed, if f is the vector-valued map given by Proposition 1.2.7, then from the G-equivariance it is easy to see that f is of polynomial growth (indeed it is Lipschitz). But the linear projections $\{f \cdot v : v \in H\}$ of f to scalar-valued harmonic maps lie in a finite-dimensional space, by Proposition 1.2.8. This implies that the range $f(G)$ of f lies in a finite-dimensional space

V. On the other hand, the obvious action of G on V has no fixed points (being a restriction of the action of G on H), and so the image of G in $GL_n(V)$ must be infinite, and Proposition 1.2.6 follows.

It remains to prove Proposition 1.2.7 and Proposition 1.2.8. Proposition 1.2.7 follows from some more general results of Korevaar-Schoen [**KoSc1997**] and Mok [**Mo1995**], though Kleiner provided an elementary proof, which we sketch below. Proposition 1.2.8 was initially proven by Colding and Minicozzi [**CoMi1997**] (for finitely presented groups, at least) using Gromov's theorem; Kleiner's key new observation was that Proposition 1.2.8 can be proven directly by an elementary argument based on a Poincaré inequality.

1.2.3. A non-constant equivariant harmonic function. We now sketch the proof of Proposition 1.2.6. The first step is to just get the action on a Hilbert space with no fixed points:

Lemma 1.2.9. *Let G be a countably infinite amenable group. Then there exists an action of G on a Hilbert space H with no fixed points.*

This is essentially the well-known assertion that countably infinite amenable groups do not obey *Property (T)*, but we can give an explicit proof as follows. Using amenability, one can construct a nested *Følner sequence* $F_1 \subset F_2 \subset \ldots \subset \bigcup_n F_n = G$ of finite sets with the property that $|(F_{n-1} \cdot F_n)\Delta F_n| \leq 2^{-n}|F_n|$ (say). (In the case of groups of polynomial growth, one can take $F_n = B(\mathrm{id}, R_n)$ for some rapidly growing, randomly chosen sequence of radii R_n.) We then look at $H := l^2(\mathbf{N}; l^2(G))$, the Hilbert space of sequences $f_1, f_2, \dots \in l^2(G)$ with $\sum_n \|f_n\|^2_{l^2(G)} < \infty$. This space has the obvious unitary action of G, defined as $g : (f_n(\cdot))_{n\in\mathbf{N}} \to (f_n(g\cdot))_{n\in\mathbf{N}}$. This action has a fixed point 0, but we can delete this fixed point by considering instead the affine-isometric action $f \mapsto gf + gh - h$, where h is the sequence $h = (\frac{1}{|F_n|^{1/2}} 1_{F_n})_{n\in\mathbf{N}}$. This sequence h does not directly lie in H, but observe that $gh - h$ lies in H for every g. One can then easily show that this action obeys the conclusions of Lemma 1.2.9.

Another way of asserting that an action of G on H has no fixed point is to say that the energy functional $E : H \to \mathbf{R}^+$ defined by $E(v) := \frac{1}{2}\sum_{s\in S} \|sv - v\|^2_H$ is always strictly positive. So Lemma 1.2.9 concludes that there exists an action of G on a Hilbert space on which E is strictly positive. It is possible to then conclude that there exists another action of G on another Hilbert space on which the energy E is not only strictly positive but actually attains its minimum at some vector v. This observation follows from more general results of Fisher and Margulis [**FiMa2005**], but one can also argue directly as follows. For every $0 < \lambda < 1$ and $A > 0$, there must exist a vector v which almost minimises E in the sense that $E(v') \geq \lambda E(v)$