© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1

Introduction

1.1 THE SUBJECT

In the present monograph we develop a structure theory for a class of finite structures whose description lies on the border between model theory and group theory. Model theoretically, we study large finite structures for a fixed finite language, with a bounded number of 4-types. In group theoretic terms, we study all sufficiently large finite permutation groups which have a bounded number of orbits on 4-tuples and which are k-closed for a fixed value of k. The primitive case is analyzed in [KLM; cf. Mp2]. The treatment of the general case involves the application of model theoretic ideas along lines pioneered by Lachlan.

We show that such structures fall into finitely many classes naturally parametrized by "dimensions" in the sense of Lachlan, which approximate finitely many infinite limit structures (a version of Lachlan's theory of shrinking and stretching), and we prove uniform finite axiomatizability modulo appropriate axioms of infinity (quasifinite axiomatizability). We also deal with issues of effectivity. At our level of generality, the proofs involve the extension of the methods of stability theory—geometries, orthogonality, modularity, definable groups—to this somewhat unstable context. Our treatment is relatively self-contained, although knowledge of the model theoretic background provides considerable motivation for the results and their proofs. The reader who is more interested in the statement of precise results than in the model theoretic background will find them in the next section.

On the model theoretic side, this work has two sources. Lachlan worked out the theory originally in the context of stable structures which are homogeneous for a finite relational language [La], emphasizing the parametrization by numerical invariants. Zilber, on the other hand, investigated totally categorical structures and developed a theory of finite approximations called "envelopes," in his work on the problems of finite axiomatizability. The class of $\aleph_0$-categorical, $\aleph_0$-stable structures provides a broad model theoretic context to which both aspects of the theory are relevant. The theory was worked out at this level in [CHL], including the appropriate theory of envelopes. These were used in particular to show that the corresponding theories are not finitely ax-

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

iomatizable, by Zilber's method. The basic tool used in [CHL], in accordance with Shelah's general approach to stability theory and geometrical refinements due to Zilber, was a "coordinatization" of an arbitrary structure in the class by a tree of standard coordinate geometries (affine or projective over finite fields, or degenerate. Other classical geometries involving quadratic forms were conspicuous only by their absence at this point.

The more delicate issue of finite axiomatizability modulo appropriate "axioms of infinity," which is closely connected with other finiteness problems as well as problems of effectivity, took some time to resolve. In [AZ1] Ahlbrandt and Ziegler isolated the relevant combinatorial property of the coordinatizing geometries, which we refer to here as "geometrical finiteness," and used it to prove quasifinite axiomatizability in the case of a single coordinatizing geometry. The case of $\aleph_0$-stable, $\aleph_0$-categorical structures in general was treated in [HrTC].

The class of *smoothly approximable* structures was introduced by Lachlan as a natural generalization of the class of $\aleph_0$-categorical $\aleph_0$-stable structures, in essence taking the theory of envelopes as a definition. Smoothly approximable structures are $\aleph_0$-categorical structures which can be well approximated by finite structures in a sense to be given precisely in §2.1. One of the achievements of the structure theory for $\aleph_0$-categorical $\aleph_0$-stable theories was the proof that they are smoothly approximable in Lachlan's sense. While this was useful model theoretically, Lachlan's point was that in dealing with the model theory of large finite structures, one should also look at the reverse direction, from smooth approximability to the structure theory. We show here, confirming this not very explicitly formulated conjecture of Lachlan, that the bulk of the structure theory applies to smoothly approximable structures, or even, as stated at the outset, to sufficiently large finite structures with a fixed finite language, having a bounded number of 4-types.

Lachlan's project was launched by Kantor, Liebeck, and Macpherson in [KLM] with the classification of the primitive smoothly approximable structures in terms of various more or less classical geometries (the least classical being the "quadratic" geometry in characteristic 2, described in §2.1.2). These turn up in projective, linear, and affine flavors, and in the affine case there are some additional nonprimitive structures that play no role in [KLM] but will be needed here ("affine duality," §2.3). Bearing in mind that any $\aleph_0$-categorical structure can be analyzed to some degree in terms of its primitive sections, the results of [KLM] furnish a rough coordinatization theorem for smoothly approximable structures. This must be massaged a bit to give the sort of coordinatization that has been exploited previously in an ω-stable context. We will refer to a structure as "Lie coordinatizable" if it is bi-interpretable with a structure which has a nice coordinatization of the type introduced below. Lie coordinatizability will prove to be equivalent to smooth approximability, in one direction largely because of [KLM], and in the other by the analog of Zilber's

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

theory of envelopes in this context. One tends to work with Lie coordinatizability as the basic technical notion in the subject. The analysis in [KLM] was in fact carried out for primitive structures with a bound on the number of orbits on 5-tuples, and in [Mp2] it was indicated how the proof may be modified so as to work with a bound on 4-tuples. (Using only [KLM], we would also be forced to state everything done here with 5 in place of 4.)

In model theory, techniques for going from a good description of primitive pieces to meaningful statements about imprimitive structures generally fall under the heading of "geometrical stability theory," whose roots lie in early work of Zilber on $\aleph_1$-categorical theories, much developed subsequently. Though the present theory lies slightly outside stability theory (it can find a home in the more recent developments relating to simple theories), geometrical stability theory provided a very useful template [Bu, PiGS].

Before entering into greater detail regarding the present work, we make some comments on the Galois correspondence between structures and permutation groups implicit in the above, and on its limitations.

Let X be a finite set. There is then a Galois correspondence between subgroups of the symmetric group *Sym*(X) on X, and model theoretic structures with universe X, associating to a permutation group the invariant relations, and to a structure its automorphism group. This correspondence extends to $\aleph_0$-categorical structures ([AZ1, Introduction], [CaO]).

When we consider infinite families of finite structures in general, or a passage to an infinite limit, this correspondence is not well behaved. For instance, the automorphism group of a large finite random graph of order n (with constant and nontrivial edge probability) is trivial with probability approaching 1 as n goes to infinity, while the natural model theoretic limit is the random countable graph, which has many automorphisms.

It was shown in [CHL], building on work of Zilber for totally categorical structures, that structures which are both $\aleph_0$-categorical and $\aleph_0$-stable can be approximated by finite structures simultaneously in both categories. Lachlan emphasized the importance of this property, which will be defined precisely in §2.1, and proposed that the class of structures with this property, the *smoothly approximable structures*, should be amenable to a strong structure theory, appropriately generalizing [CHL]. Moreover, Lachlan suggested that the direction of the analysis can be reversed, from the finite to the infinite: one could classify the large finite structures that appear to be "smooth approximations" to an infinite limit, or in other words, classify the families of finite structures which appear to be Cauchy sequences both as structures and as permutation groups. This line of thought was suggested by Lachlan's work on stable finitely homogeneous structures [La], much of which predates the work in [CHL], and provided an additional ideological framework for that paper.

In the context of stable finitely homogeneous structures this analysis in terms of families parametrized by dimensions was carried out in [KL] (cf. [CL, La]),

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

but was not known to go through even in the totally categorical case. Harrington pointed out that this reversal would follow immediately from compactness if one were able to work systematically within an elementary framework [Ha]. This idea is implemented here: we will replace the original class of "smoothly approximable structures" by an elementary class, a priori larger. Part of our effort then goes into developing the structure theory for the ostensibly broader class.

From the point of view of permutation group theory, it is natural to begin the analysis with the case of finite primitive structures. This was carried out using group theoretic methods in [KLM], and we rely on that analysis. However, there are model theoretic issues which are not immediately resolved by such a classification, even for primitive structures. For instance, if some finite graphs G_n are assumed to be primitive, and to have a uniformly bounded number of 4-types, our theory shows that an ultraproduct G^* of the G_n is bi-interpretable with a Grassmannian structure, which does not appear to follow from [KLM] by direct considerations. The point here is that if G_n is "the same as" a Grassmannian structure in the category of permutation groups, then it is bi-interpretable with such a structure on the model theoretic side. To deal with families, one must deal (at least implicitly) with the uniformity of such interpretations; see §8.3, and the sections on reducts. It is noteworthy that our proof in this case actually passes through the theory for imprimitive structures: any nonuniform interpretation of a Grassmannian structure on G_n gives rise to a certain structure on G^*, a reduct of the structure which would be obtained from a uniform interpretation, and one argues that finite approximations (on the model theoretic side) to G^* would have too many automorphisms. In other words, we can obtain results on uniformity (and hence effectivity) by ensuring that the class for which we have a structure theory is closed under reducts. This turns out to be a very delicate point, and perhaps the connection with effectivity explains why it should be delicate.

1.2 RESULTS

A rapid but thorough summary of this theory was sketched in [HrBa], with occasional inaccuracies. For ease of reference we now repeat the main results of the theory as presented there, making use of a considerable amount of specialized terminology which will be reintroduced in the present work. The various finiteness conditions referred to are all given in Definition 2.1.1.

Theorem 1 (Structure Theory)

Let $\mathcal{M}$ be a Lie coordinatizable structure. Then $\mathcal{M}$ can be presented in a finite language. Assuming $\mathcal{M}$ is so presented, there are finitely many definable dimension invariants for $\mathcal{M}$ which are infinite, up to equivalence of

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

such invariants. If C is a set of representatives for such definable dimension invariants, then there is a sentence $\varphi = \varphi_{\mathcal{M}}$ with the following properties:

1. *Every model of φ in which the definable dimension invariants of C are well-defined is determined up to isomorphism by these invariants.*
2. *Any sufficiently large reasonable sequence of dimension invariants is realized by some model of φ.*
3. *The models of φ for which the definable dimension invariants of C are well-defined embed homogeneously into $\mathcal{M}$ and these embeddings are unique up to an automorphism of $\mathcal{M}$.*

There are a considerable number of terms occurring here which will be defined later. Readers familiar with "shrinking" and "stretching" in the sense of Lachlan should recognize the situation. Definable dimension invariants are simply the dimensions of coordinatizing geometries which occur in families of geometries of constant dimension; when the appropriate dimensions are not constant within each family, the corresponding invariants are no longer well-defined. A dimension invariant is reasonable if its parity is compatible with the type of the geometry under consideration; in particular, infinite values are always reasonable.

The statements of the next two theorems are slight deformations of the versions given in [HrBa]. We include more clauses here, and we use definitions which vary slightly from those used in [HrBa].

Theorem 2 (Characterizations)

The following conditions on a model $\mathcal{M}$ are equivalent:

1. *$\mathcal{M}$ is smoothly approximable.*
2. *$\mathcal{M}$ is weakly approximable.*
3. *$\mathcal{M}$ is strongly quasifinite.*
4. *$\mathcal{M}$ is strongly 4-quasifinite.*
5. *$\mathcal{M}$ is Lie coordinatizable.*
6. *The theory of $\mathcal{M}$ has a model $\mathcal{M}^*$ in a nonstandard universe whose size is an infinite nonstandard integer, and for which the number of internal n-types $s_n^*(\mathcal{M}^*)$ satisfies*

$$s_n^*(\mathcal{M}^*) \leq c^{n^2}$$

for some finite c, and in which internal n-types and n-types coincide. (Here n varies over standard natural numbers.)

The class characterized above is not closed under reducts. For the closure under reducts we have:

Theorem 3 (Reducts)

The following conditions on a model $\mathcal{M}$ are equivalent:

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1. *$\mathcal{M}$ has a smoothly approximable expansion.*
2. *$\mathcal{M}$ has a weakly approximable expansion.*
3. *$\mathcal{M}$ is quasifinite.*
4. *$\mathcal{M}$ is 4-quasifinite.*
5. *$\mathcal{M}$ is weakly Lie coordinatizable*
6. *The theory of $\mathcal{M}$ has a model $\mathcal{M}^*$ in a nonstandard universe whose size is an infinite nonstandard integer, and for which the number of internal n-types $s_n^*(\mathcal{M}^*)$ satisfies:*

$$s_n^*(\mathcal{M}^*) \leq c^{n^2}$$

for some finite c. (Here n varies over standard natural numbers.)

On the other hand, once the class is closed under reducts it is closed under interpretation, hence:

Theorem 4 (Interpretations)
The closure of the class of Lie coordinatizable structures under interpretation is the class of weakly Lie coordinatizable structures.

An earlier claim that the class of Lie coordinatizable structures is closed under interpretations was refuted by an example of David Evans which will be given below.

Theorem 5 (Decidability)
For any k and any finite language, the theory of finite structures with at most k 4-types is decidable, uniformly in k. The same applies in an extended language with dimension comparison quantifiers and Witt defect quantifiers. Thus one can decide effectively whether a sentence in such a language has a finite model with a given number of 4-types.

This is a distant relation of a family of theorems in permutation group theory giving explicit classifications of primitive permutation groups with very few 2-types. Dimension comparison quantifiers do not allow us to quantify over the dimensions of spaces, but they allow us to compare the dimensions of any two geometries. Witt defect quantifiers are more technical (§2.1, Definition 2.1.1).

Theorem 6 (Finite structures)
Let L be a finite language and k a natural number. Then the class of finite L-structures having at most k 4-types can be divided into families $\mathcal{F}_1, \ldots, \mathcal{F}_n$ for some effectively computable n such that

1. *Each family $\mathcal{F}_i$ is finitely axiomatizable in a language with dimension comparison and Witt defect quantifiers.*
2. *Each family $\mathcal{F}_i$ is associated with a single countable Lie coordinatizable structure $\mathcal{M}_i$. The family $\mathcal{F}_i$ is the class of "envelopes" of $\mathcal{M}_i$, which are the structures described in Theorem 1, parametrized by freely vary-*

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

ing definable dimension invariants (above a certain minimal bound, with appropriate parity constraints).

3. *For $\mathcal{M}$, $\mathcal{N}$ in $\mathcal{F}_i$, if the dimension invariants satisfy $d(\mathcal{M}) \leq d(\mathcal{N})$ then there is a homogeneous embedding of $\mathcal{M}$ in $\mathcal{N}$, unique up to an automorphism of $\mathcal{N}$.*
4. *Membership in each of the families $\mathcal{F}_i$ (and in particular, in their union) can be determined in polynomial time, and the dimension invariants can be computed in polynomial time. Thus the isomorphism problem in the class of finite structures with a bounded number of types can be solved in polynomial time.*
5. *The cardinality of an envelope of dimension d is an exponential polynomial in d; specifically, a polynomial in exponentials of the entries of d (with bases roughly the sizes of the base fields involved). The structure $N_i(d)$ which is the member of $\mathcal{F}_i$ of specified dimensions d can be constructed in time which is polynomial in its cardinality.*

Theorem 7 (Model Theoretic Analysis)

The weakly Lie coordinatizable structures $\mathcal{M}$ are characterized by the following nine model theoretic properties:

LC1. $\aleph_0$*-categoricity.*

LC2. *Pseudofiniteness.*

LC3. *Finite rank.*

LC4. *Independent type amalgamation.*

LC5. *Modularity in* $\mathcal{M}^{\mathrm{eq}}$.

LC6. *The finite basis property in groups.*

LC7. *General position of large* 0*-definable sets.*

LC8. $\mathcal{M}$ *does not interpret the generic bipartite graph.*

LC9. *For every vector space V interpreted in $\mathcal{M}$, the definable dual V^* (the set of all definable linear maps on V) is interpreted in $\mathcal{M}$.*

Some of these notions were first introduced in [HrBa], sometimes using different terminology. In particular, the rank function is not a standard rank function, the finite basis property in groups (or "linearity") reduces to local modularity in the stable case, and the general position (or "rank/measure") property is an additional group theoretic property that arises in the unstable case, when groups tend to have many definable subgroups of finite index. The eighth condition is peculiarly different from the ninth. This is a corrected version of Theorem 6 of [HrBa].

David Evans made several contributions to the theory given here, notably the observation that the orientation of quadratic geometries is essential, and bears on the problem of reducts. The detection of all such points is critical. Evans also gave a treatment of weak elimination of imaginaries in linear geometries, in [EvSI].

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

We will say a few words about the development of this material, using technical notions explained fully in the text. The first author on reading [KLM] understood that one could extract stably embedded geometries from the analysis of primitive smoothly approximable structures given there, and that the group theory gives a decent orthogonality theory (but the orthogonality theory given here will be based more on geometry than on group theory). These ingredients seemed at first to be enough to reproduce the Ahlbrandt–Ziegler analysis, after the routine verification that the necessary geometrical finiteness principle follows from Higman's lemma; all of this follows the lead of [AZ1], along the lines developed in [HrTC]. An attempt to implement this strategy failed, in part because at this stage there was no hint of "affine duality."

The second author then produced affine duality and gave a complete proof of quasifinite axiomatizability, introducing some further modifications of the basic strategy, notably canonical projectives and a closer analysis of the affine case. The theme in all of this is that one should worry even more about the interactions of affine geometries than one does in the stable case. This can perhaps be explained by the following heuristic. Only the projective geometries are actually coordinatizing geometries; the linear and affine geometries are introduced to analyze definable group structures, in keeping with the general philosophy that structures are built from basic 1-dimensional pieces, algebraic closure, and definable groups. Here higher dimensional groups are not needed largely because of the analog of 1-basedness, referred to below as the finite basis property. The developments that go beyond what is needed for quasifinite axiomatizability are all due to the second author. The extension of a considerable body of geometric stability theory to this context is essential to further developements. The high points of these developments, as far as applications are concerned, are the analysis of *reducts* and its applications to issues of *effectivity*. It may be noted also that the remarkable quadratic geometries have been known for some time, and play an essential role in [KLM], in particular. In our view they add considerably to the appeal of the theory.

The treatment of reducts requires a considerably more elaborate transference of techniques of stability theory to this unstable setting than would be required for the quasifinite axiomatizability alone. This would not be indispensable for the treatment of structures already equipped with a Lie coordinatization; but to apply these results to classes which are closed under interpretation requires the ability to recognize an appropriate coordinatization, starting from global properties of the structure; thus one must find the model theoretic content of the property of coordinatizability by the geometries on hand.

Our subject has also been illuminated by recent developments in connection with Shelah's "simple theories," and is likely to be further illuminated by that theory.

Various versions of this material, less fully worked out, have been in circulation for a considerable period of time (beginning with notes written in Spring

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1990) and have motivated some of the work in simple theories. In particular, versions of sections 5.1 [KiP], 5.4, and 6.1 [PiGr] have been obtained in that very general context; all of this rests on the theoretical foundation provided by the original paper of Shelah [ShS] and subsequent work by Kim [Ki].

Some comments on the relationship of this theory to Shelah's "simple theories" are in order. Evidently a central preoccupation of the present work is the extension of methods of stability theory to an unstable context. Stability theory is a multilayered edifice. The first layer consists of a theory of rank and the related combinatorial behavior of definable sets. The next layer includes the theory of orthogonality, regular types, and modularity, and was initially believed to be entirely dependent on the foundational layer in its precise form. One of the key conclusions of the present work is that is possible to recover the second "geometric model theory" layer over an unstable base. Because we have $\aleph_0$-categoricity and finiteness of the rank, our basic rank theory becomes as simple as possible; nonetheless, almost all of the "second-level" phenomena connected with simplicity appear in our context with their full complexity—the main exception being the Lascar group. It was perhaps this combination of circumstances that facilitated a very successful generalization of the "geometric theory" to the simple context, once the first layer was brought into an adequate state by Kim's thesis [KiTh].

As far as the present work is concerned, the development of a sufficiently general theory was often due to necessity rather than insight. For example, if we—or the creator of the finite simple groups—had been able to exclude from consideration the orthogonal geometries in characteristic 2, we would have had a considerably simpler theory of generics in groups, with *Stab* $=$ *Stab*$_\circ$ (cf. §6.1, Definition 6.1.9, and the Example following). Such a simplified theory would have been much less readily generalizable to the simple context; in addition, under the same hypothesis, this simplified theory would have largely obviated the need for the theory of the semi-dual cover.

A number of features of the theory exposed here have been generalized with gratifying success to the context of simple theories, but some have not. On the positive side, one has first of all the theorem which we originally called *the independence theorem*. This name has become standard in the literature, although in the present manuscript it was eventually renamed "the type amalgamation property." In any case this is still a misnomer, as this amalgamation involves a triple over a base rather than a pair. Compare the following "homological" description. Let $I(n)$ be the space of n-types, over some fixed base, of *independent* n-tuples (whose elements are themselves finite sequences of elements). We have "projection" maps $\pi_i : I(n) \to I(n-1)$ obtained by deletion of one coordinate. The uniqueness of forking in stability theory is the statement that the induced map $I(2) \to I(1)^2$ is *injective*. We replace this by an *exactness* property, characterizing the image of $I(3)$ in $I(2)^3$ by minimal coherence conditions.

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

The first proof found for this theorem consisted of inspection in the 1-dimensional case, followed by an induction on rank. In the course of related work, an abstract proof was found, assuming finite simplicity rank and definability of the rank. This proof was later generalized by Kim and Pillay, and together with their realization of the relevance of the Lascar group, it became the central pillar of simplicity theory. In §5.1 we retain the original clumsy inductive proof. This may be of use in situations where simplicity is not known in advance.

The main point in any case is not the proof of this theorem but the realization that the uniqueness of nonforking extensions, which seemed characteristic of stability theory and essential to its fabric, can be replaced “densely often” with an appropriate *existential* statement.

The definition of *modularity* could largely be taken over from the stable case. A new idea was required (cf. §5.4) to produce enough geometric imaginaries for proof of the local–global principle; this idea survives in the contemporary treatment of canonical bases in simple theories. The consequences of modularity for groups are not as decisive in general as in the stable case, even generically, so we had to consider stronger variants. The recognition theorems in rank one which use these properties serve to situate the basic geometries model theoretically to a degree. One would like to see these theorems generalized, as Zilber’s characterizations of modular groups were extended from the totally categorical to the strongly minimal case.

The strong presence of duality is also a new feature as far as the model theory is concerned. Initially it arose as a particular instance of instability, which we sought to circumscribe and neutralize as much as possible. At the outset duals must be recognized in order to render the basic geometries stably embedded; the dual space of a finite vector space is also a prime example of a nonuniform interpretation. Eventually duality also emerged as a positive tool, useful for certain purposes even in contexts where stability is initially assumed: see §6.5, on the semi-dual cover, and also the treatment of second-order quantifiers in Chapter 8, dealing with effectivity. It seems possible that linear duality, like modularity, has some significance in general model theoretic frameworks, but at this time our situation remains isolated, awaiting further illumination.

The proof of Theorem 2 will be largely complete by the end of §3.5 (see the discussion in §3.5 for more on this). The final section (§8.4) contains some retrospective remarks on the structure of our development.

Various versions of this paper have benefited from remarks by a variety of model theorists. We thank particularly Ambar Chowdhury, David Evans, Bradd Hart, Dugald Macpherson, Anand Pillay, and Frank Wagner for their remarks. We thank Virginia Dunn, Amélie Cherlin, and Jakob Kellner for various forms of editorial and technical assistance. The first author also thanks Amaal for diverting correspondence during the preparation of the final version.

For general queries, contact webmaster@press.princeton.edu