Jahresber Dtsch Math-Ver (2014) 116:1–2
DOI 10.1365/s13291-014-0084-8

PREFACE

Preface Issue 1-2014

Hans-Christoph Grunau

Last year, the Zentralblatt für Mathematik and the Jahresbericht der DMV began a collaboration to review again classical books and papers and the effect that they have had or could have had on the development of specific mathematical branches from today's point of view. It has turned out that these renewed reviews build a category of articles of its own and that they cannot in general be considered as book reviews or historical articles. For this reason we introduce the new category "Classics Revisited" and shall try to have one article of this kind in each issue. We begin with a new review by Jean-Pierre Otal of William Thurston's 1982 paper on "Three-dimensional manifolds, Kleinian groups and hyperbolic geometry". This research announcement did not only contain his Geometrization Conjecture and his Hyperbolization Theorem for so called Haken manifolds but also a list of 24 of at that time open problems, of which 22 had been solved by 2012. Only few papers have been of such a strong influence and inspiration.

Bosons are particles in quantum mechanics which do not obey the Pauli principle. If a large fraction of particles in a Bose fluid occupies the same quantum state, the so called Bose-Einstein condensation occurs and leads to physically remarkable phenomena like superfluidity. This means that the viscosity of a fluid vanishes completely at low temperatures. Robert Seiringer's survey article on "The excitation spectrum for Bose fluids with weak interaction" reviews recent progress in understanding underlying mathematical models.

Mass aggregation phenomena like e.g. polymerisation may be modelled by means of Smoluchowski's coagulation equation. Barbara Niethammer's survey article addresses the key question in understanding the large time behaviour of solutions

H.-Ch. Grunau (✉)
Institut für Analysis und Numerik, Fakultät für Mathematik, Otto-von-Guericke-Universität, Postfach 4120, 39016 Magdeburg, Germany
e-mail: hans-christoph.grunau@ovgu.de

whether this can be described with the help of particular self-similar solutions. Existence and uniqueness results are reviewed and some recent and more elegant proofs are indicated.

Last but not least, recently released books concerning “Clifford algebras and Lie theory”, “Mathematics in image processing”, and “Oblique derivative problems for elliptic equations” are reviewed.

Jahresber Dtsch Math-Ver (2014) 116:3–20
DOI 10.1365/s13291-014-0079-5

CLASSICS REVISITED

William P. Thurston: "Three-dimensional manifolds, Kleinian groups and hyperbolic geometry"

Bull. Am. Math. Soc., New Ser. 6, 357–379 (1982)

Jean-Pierre Otal

Published online: 25 February 2014

This paper is the written version of a talk that Thurston gave at the AMS Symposium on the Mathematical Heritage of Henri Poincaré held at Bloomington in April 1980. It is a Research announcement.[1]

In his "Analysis Situs" paper, Poincaré introduces simplicial homology and the fundamental group, with the aim of finding invariants which could distinguish 3-manifolds. He concludes the 5th complement [84] discussing the question: Is any simply connected closed 3-manifold homeomorphic to the 3-sphere? This question, often understood as a conjecture, motivated much of the development of 3-dimensional topology. At the beginning of his paper Thurston proposes, "fairly confidently", his Geometrization Conjecture: *The interior of any compact 3-manifold has a canonical decomposition into pieces that carry a geometric structure.* By Kneser and Milnor any orientable compact 3-manifold is homeomorphic to the connected sum of manifolds which are either irreducible or homeomorphic to a sphere bundle over S^1. By Jaco-Shalen and Johannson any orientable irreducible 3-manifold can be decomposed along a disjoint union of embedded tori into pieces that are atoroidal or Seifert-fibered (see in [14] the corresponding statement for non-orientable manifolds). Saying that a

[1] Before 1978 the Bulletin of the AMS did not accept Research announcements longer than 100 lines. For this reason, Thurston could not submit there his preprint about surface diffeomorphisms [96] which dates back to 1976 and which is written in the same informal style as this one. After a couple of refusals, he stopped trying to publish it and waited until 1988 to do so (see the introduction to [96]).

This is a slightly revised version of a review which appeared first in Zentralblatt für Mathematik (Zbl. 0496.57005).

J.-P. Otal (✉)
Institut Mathématique de Toulouse, Université Paul Sabatier, 118, Route de Narbonne, 31062 Toulouse, France
e-mail: otal@math.univ-toulouse.fr

manifold carries a *geometric structure* means that it has a complete Riemannian metric which is locally modelled on one of the eight geometries (which are introduced in Sect. 4 of the paper): the constant curvature ones $\mathbb{H}^3$, $\mathbb{S}^3$, $\mathbb{R}^3$, the Seifert-fibered ones $\widetilde{\mathrm{PSL}(2,\mathbb{R})}$, $\mathbb{H}^2 \times \mathbb{R}$, $\mathbb{S}^2 \times \mathbb{R}$, Nil and Sol. The pieces which are referred to in the Geometrization Conjecture are essentially the pieces of the Jaco-Shalen-Johannson decomposition; more precisely, all the pieces except that one needs to keep undecomposed any component of the Kneser-Milnor decomposition which is a torus bundle over S^1 with Anosov monodromy, since those bundles carry a Sol-geometry. This conjecture contains the Poincaré Conjecture as a very special case. But the main difference with the Poincaré Conjecture is that the Geometrization Conjecture proposes a global vision of *all* compact 3-manifolds. Furthermore, the geometric structure on a given 3-manifold when it exists is often unique, up to isometry; consequently, all the Riemannian invariants of the geometric structure on a 3-manifold like the volume, the diameter, the length of the shortest closed geodesic, the Chern-Simons invariant, etc. are then topological invariants.

The Geometrization Conjecture was proven in 2003 by Grigori Perelman, who followed an analytical approach that had been initiated by Richard Hamilton in [40]. This approach was entirely different from that of Thurston, but there is no doubt that the results announced in the present paper contributed to give a very strong evidence for the truth of the Geometrization Conjecture, and therefore for the truth of the Poincaré Conjecture (in the introduction of [69], John Morgan points out this apparent paradox that a conjecture might look less plausible than another one which is much stronger).

Among the results announced in this paper which gave strong evidence for the validity of the Geometrization Conjecture, the most important is certainly the Hyperbolization Theorem for Haken manifolds (Theorem 2.5): *Any Haken 3-manifold which is homotopically atoroidal is hyperbolic, except the particular case of the twisted I-bundle over the Klein bottle*. One corollary is that the Geometrization Conjecture is true for all Haken manifolds, and in particular for any irreducible 3-manifold with non-empty boundary (once accepted the Hyperbolization Theorem, the proof of this corollary reduces to checking that Seifert fibered spaces and T^2-bundles over the circle are geometric too).

Haken manifolds form a large class of 3-manifolds for which the Geometrization Conjecture is true. However, as Thurston writes after the statement of the Hyperbolization Theorem, even if the incompressible surface is an essential tool in the proof of the Hyperbolization Theorem for Haken manifolds, this surface seems to have little to do with the existence of the hyperbolic metric. And then he discusses the manifolds obtained by Dehn surgery on a 3-manifold with boundary M when the interior of M carries a finite volume hyperbolic metric. Theorem 2.6, *the Hyperbolic Dehn Surgery Theorem*, states that for all surgery data, except those involving a finite set of slopes on each boundary component, the resulting 3-manifold is hyperbolic. Furthermore, few of the resulting manifolds are Haken: Thurston had shown this when M is the figure-eight knot complement [92], and Hatcher-Thurston had shown this when M is a 2-bridge knot complement [41] (a paper which circulated as a preprint as early as 1979). So these manifolds form a new family of hyperbolic manifolds which are not covered by the Hyperbolization Theorem.

Thurston explains that he found using a computer program that most Dehn surgeries on certain punctured tori bundles over the circle were hyperbolic and that he constructed then by hand geometric structures for all the other Dehn surgeries. He insists on the beauty of the geometric structures "when you learn to see them" that is often revealed by computer pictures.

The following section "Applications" describes some properties of Haken manifolds and of hyperbolic manifolds.

- When M is Haken and atoroidal, the group of isotopy classes of homeomorphisms of M is finite and it lifts to a group of homeomorphisms of M.
- The fundamental group of a Haken manifold is *residually finite*.
- Theorem 3.4: the Gromov-Thurston theorem on the behavior of volume under a non-zero degree map between hyperbolic manifolds $M \to N$: $\operatorname{vol} M \geq \text{degree} \cdot \operatorname{vol} N$ with equality only when the covering map is homotopic to a regular cover.
- Theorem 3.5: the Jørgensen-Thurston theorem that says that the set of volumes of hyperbolic manifolds is well-ordered. In particular, there is a finite volume manifold with smallest volume.
- Thurston suggests that the volume and the eta invariant of a hyperbolic 3-manifold should be considered as the real and imaginary parts of a complex number.

Then Thurston describes concrete approaches "to make the calculation of hyperbolic structures routine" and he explains how Robert Riley used computers for constructing hyperbolic structures [86]. Riley was maybe the first person who tried to "hyperbolize" a given topological 3-manifold, namely the complement in $\mathbb{S}^3$ of certain knots (see [87]).

The last section addresses two theorems which are directly related to the proof of the hyperbolization of Haken manifolds. Theorem 5.7, *the Double Limit Theorem*, is the main step of the hyperbolization for manifolds which fiber over the circle; it gives a condition for a sequence (ρ_i) of quasi-Fuchsian groups to contain a subsequence which converges up to conjugacy in the space of representations of $\pi_1(S)$ in $\mathrm{PSL}(2, \mathbb{C})$. The condition, formulated in terms of the Ahlfors-Bers coordinates of ρ_i is that the two coordinates diverge to laminations whose reunion fills up the surface. This theorem led to beautiful objects, in particular to *the Cannon-Thurston map* which is a sphere-filling curve $S^1 \to \partial\mathbb{H}^3$ which conjugates the action of the surface group on S^1 as a Fuchsian group to its action on $\partial\mathbb{H}^3$ as the limiting Kleinian group; furthermore this map can be described in geometric terms. Theorem 5.8 is *the Compactness Theorem of the space of discrete and faithful representations into* $\mathrm{PSL}(2, \mathbb{C})$ *of the fundamental group of acylindrical 3-manifolds*, one of the key ingredients of the proof in the non-fibered case.

The paper contains neither proofs nor hints of proofs. But the Princeton lecture notes [92] were circulating widely since the end of the 70's. These notes were not aimed to give a proof of the Hyperbolization Theorem, but they did contain important material which is relevant to the proof (in particular in the notoriously difficult Chaps. 8 and 9 which discuss algebraic versus geometric convergence of hyperbolic 3-manifolds). Also Thurston had already lectured about the Hyperbolization Theorem at several conferences (see [93] and [94]). The first versions of [95] (proving the Compactness Theorem 5.8 above) and [97] (the Double Limit Theorem 5.7) were circulating since 1980. The preprint [98] which contains a refinement of the main result of [95] was distributed later, in 1986. In 1982, the most detailed presentation of the proof of the Hyperbolization Theorem in the non-fibered case, written by John Morgan, was published in the Proceedings of the Smith Conjecture conference [68], a conference held in 1979; this presentation follows the proof described in [94]. Dennis Sullivan had reported about the proof in the fibered case at the Bourbaki seminar in 1980 [91].

The Hyperbolization Theorem for Haken manifolds was sometimes named *the Monster Theorem*, because of the length of the proof and the unusual amount of different techniques that were required. It became a challenge to find other approaches. John Hubbard observed a deep relation between the main step of the proof for the non-fibered case *the fixed point theorem for the skinning map*—and a conjecture of Irwin Kra in Teichmüller theory saying that the classical Theta operator acting on holomorphic quadratic differentials is a contraction for the L^1-norm. This conjecture was solved by Curt McMullen who showed how to apply it to the fixed point problem [58, 59] (see also [80]). A proof of the fibered case is contained in [79]; it is based on a proof of the Double Limit Theorem, different from Thurston's and uses $\mathbb{R}$-trees instead of the delicate "Uniform injectivity of doubly incompressible pleated surfaces" Theorem (see also [48]).

Thurston concludes his paper with a list of problems about 3-manifolds and about Kleinian groups which contains many of the most difficult problems of these fields that had already been posed. Any mathematician can appreciate the importance and the impact of this paper through the fact that among these 24 problems, 22 were solved by 2012 (and indeed the two remaining ones, Problem 19—on properties of arithmetic hyperbolic manifolds—is more a research theme than a problem and Problem 23—on the rational independence of volumes of hyperbolic manifolds—leads to difficult conjectures in number theory). Any topologist knows how much the new perspective presented in this paper influenced and inspired research in 3-dimensional topology, in hyperbolic geometry and even much beyond: it contributed in a lot of ways to the developments of geometric group theory, of complex dynamics, to the study of spaces of representations, to the study of the Weil-Petersson geometry of Teichmüller space etc. Reporting about these influences is not the purpose of this review and I will end by trying to describe each problem and, for all but two, by giving references to the papers which brought the solution. As the copious references to be found in Zentralblatt or in MathSciNet testify, this work has inspired a vast literature. For reasons of space, I have been unable to cite many important papers which solved significant intermediate steps.

1 The List of Problems

I have classified the problems according to the following themes: Geometrization of 3-manifolds and of 3-orbifolds, Topology of 3-manifolds, Kleinian groups, Subgroups of the fundamental groups of 3-manifolds, Computer programs and tabulations, Arithmetic properties of Kleinian groups.

It appeared that there were connections between certain of the problems. For the exposition, I followed the chronological order of the solutions within a given theme, rather than the order in which Thurston presented the problems.

2 Geometrization of 3-Manifolds and of 3-Orbifolds: Problems 1, 3

1. *The Geometrization Conjecture for* 3*-manifolds.*

This conjecture is proven by Perelman in [81–83]. The content of these preprints is explained with details by Bruce Kleiner and John Lott in [52]. John Morgan and Gang Tian give complete proofs of the Poincaré Conjecture and of the geometrization of 3-manifolds with finite fundamental groups in [70]. A complete proof of the Geometrization Conjecture for arbitrary 3-manifolds, including a simplification of the original argument (but assuming the Hyperbolization Theorem for Haken manifolds) is contained in [9].

3. *The Geometrization Conjecture for* 3*-orbifolds.*

Thurston makes the comment that this problem contains Problem 2 on the classification of finite group actions on 3-manifolds. Indeed the language of orbifold gives a way to "encode" non-free actions of finite groups (but it does cover much more examples). If G is a finite group of diffeomorphisms acting on a 3-manifold M, the quotient space M/G is naturally the underlying space of an orbifold whose singular locus is the image of the set of points in M which are fixed by some non-trivial element in G and labeled by the isotropy groups. Showing that the action of G on M is geometric is equivalent to showing that this orbifold is geometric. The Geometrization Conjecture for 3-orbifolds can be formulated in a way close to the Geometrization Conjecture for 3-manifolds [14]. In a footnote added in proof, Thurston announces that he has proven this conjecture when the singular locus of the orbifold has dimension at least 1: this is the *Orbifold Theorem*. Thurston did not write anything about the Orbifold Theorem comparable to what he wrote about the Hyperbolization Theorem, although he planned to do so in 1986 (cf. the Introduction to [95]). In lectures at Durham in 1984, he outlined his proof. A complete proof of the Orbifold Theorem (for orientable orbifolds) is given by Michel Boileau, Bernhard Leeb and Joan Porti in [11]; previously Boileau and Porti had solved the case when the singular locus is a 1-dimensional submanifold [12]. The proofs in both papers differ from Thurston's at the delicate point of recognizing Seifert fibered pieces: those pieces are identified after a simplicial volume computation, using Gromov's vanishing theorem. Another reference, closer to Thurston's original proof, is [30].

3 Topology of 3-Manifolds: Problems 2, 4, 24

2. *Is any finite group action on a (geometric)* 3*-manifold equivalent to an isometric action?*

This problem is a broad extension of *the Smith conjecture*: *Any orientation preserving periodic diffeomorphism of* $\mathbb{S}^3$ *is conjugated to an orthogonal rotation* solved in [67] (see in particular the history at the end of [89]). When the geometry on M is $\widetilde{PSL(2,\mathbb{R})}$, $\mathbb{H}^2 \times \mathbb{R}$, Nil, $\mathbb{R}^3$ and Sol, Problem 2 is solved by Williams Meeks and Peter Scott in [62] using minimal surfaces and topological techniques. The case of the $\mathbb{S}^2 \times \mathbb{R}$ geometry is solved by Meeks and Yau in [63], except for actions of the alternating group A_5. The case when the geometry is $\mathbb{S}^3$ or $\mathbb{H}^3$ is solved by Jonathan Dinkelbach and Bernhard Leeb in [33] by making equivariant Perelman's proof; it is important to observe that this more recent result applies to an *arbitrary* finite group action on $\mathbb{S}^3$ or $\mathbb{H}^3$ whereas those (non-orientable) actions which have only isolated fixed points cannot be covered by the Orbifold theorem.

4. *Develop a global theory of hyperbolic Dehn surgery.*

Let M be a finite volume hyperbolic 3-manifold with k cusps; choose disjoint horoball neighborhoods of these cusps which are bounded by tori denoted T_i. If for $i = 1, \ldots, k$, s_i is *a slope*, i.e., a non trivial isotopy class of simple closed curves on T_i, $M(s_1, \ldots, s_k)$ denotes the manifold obtained from M by Dehn filling along $s = (s_i)$. The *exceptional set* is the set of the s's such that $M(s)$ is not hyperbolic. Ian Agol and Mark Lackenby find independently conditions on the slopes s_i that imply that $M(s)$ is irreducible with a *Gromov hyperbolic* fundamental group ([1, 53]). When ∂M is connected, Lackenby and Meyerhoff prove that the exceptional set contains at most 10 slopes [55], as it was conjectured by Cameron Gordon; the figure-eight knot has 10 exceptional slopes ([92], §4) and is still conjectured to be the only knot with this property. A different approach to a quantitative version of the Dehn Surgery Theorem is due to Craig Hodgson and Steve Kerckhoff using deformations of hyperbolic cone-manifolds ([43, 44]). When ∂M is connected, other important universal properties of the exceptional set are also the *Cyclic Dehn Surgery Theorem* of Mark Culler, Cameron Gordon, John Luecke, and Peter Shalen [32] which says that there are at most three slopes such that $M(s)$ has cyclic fundamental group, and the *Finite Dehn Surgery Theorem* of Steve Boyer and Xingru Zhang which says that there are at most six slopes such that $M(s)$ has a finite fundamental group [20].

24. *Show that most* 3*-manifolds with Heegaard diagrams of a given genus have hyperbolic structures.*

By the Geometrization Theorem, this problem amounts to show that most 3-manifolds with Heegaard diagrams of a given genus are irreducible and atoroidal. These properties can be checked on the curve complex of the Heegaard surface in terms of the distance between the two subcomplexes which are respectively generated by the meridian systems of each splitting [42]. But Thurston had maybe in mind a more concrete description of the hyperbolic metric, closer to the description that comes with the Hyperbolic Dehn Surgery Theorem. In [71] Hossein Namazi and Juan Souto consider 3-manifolds obtained by glueing 2 copies of a handlebody by a power ϕ^n of a pseudo-Anosov diffeomorphism which satisfies some genericity assumption. For all sufficiently high powers n, they construct a negatively curved Riemannian metric on the resulting manifold with curvature arbitrarily close to -1. Also in this case the Geometrization Theorem implies that those manifolds are hyperbolic. However, the advantage of the approach in [71] is to show the quasi-isometry type of the hyperbolic metric in terms of ϕ and n.

4 Kleinian Groups: Problems 5–14

In the 60's, Lipman Bers constructed an embedding of Teichmüller space as a bounded domain in $\mathbb{C}^n$. He deduced using a topological argument that most points in the frontier of this embedding did correspond to Kleinian groups which are *degenerate*, in the sense that they are not *geometrically finite* (see also [38]). Explicit examples of degenerate groups had been given by Troels Jørgensen in [45] (see also [46], a preprint from around 1975 which describes the fundamental domains for doubly degenerate once punctured torus groups). But before Thurston, no general geometric properties of degenerate Kleinian groups had been established. For proving the Hyperbolization Theorem, Thurston needed to consider those Kleinian groups and he began their study. Problems 5 to 14 address mostly those groups.

5. *Are all Kleinian groups geometrically tame?*

Let G be a finitely generated Kleinian group with infinite covolume that we will also suppose without parabolic elements for this exposition. By a Theorem of Peter Scott, the quotient manifold $N = \mathbb{H}^3/G$ has *a compact core*, i.e., a codimension 0 submanifold M such that the inclusion $M \hookrightarrow N$ is a homotopy equivalence. Each component of ∂M separates the interior of M from an *end* of N. The groups studied by Thurston in §8 and §9 of [92] have the following two properties: (1) M is boundary-incompressible (i.e., G does not split as a free product or as an HNN extension over the trivial group) and (2) G is a limit of *geometrically finite groups* (i.e., the inclusion homomorphism $G \to \mathrm{PSL}(2,\mathbb{C})$ is a limit of group embeddings $\rho_i : G \to \mathrm{PSL}(2,\mathbb{C})$ such that for all i, the manifold $\mathbb{H}^3/\rho_i(G)$ has a compact core with convex boundary). A consequence of (1) is that for any boundary component S of ∂M bounding an end E, the inclusion $S \hookrightarrow E$ is a homotopy equivalence. Thurston introduced the notion of *pleated surface* (also called *uncrumpled surface* in [92], §8). He showed that when G satisfies (2), then each end of N is *geometrically tame*, i.e., either E is geometrically finite, or there exists a sequence of pleated surfaces $f_i : S \to E$ homotopic to the inclusion map $S \hookrightarrow E$ and such that the maps f_i tend to ∞. The existence of this sequence has two important consequences. First N is *topologically tame*: it is diffeomorphic to the interior of a compact 3-manifold. Second, N is *analytically tame*, a property that implies in particular that G satisfies *the Ahlfors conjecture* which says that the Lebesgue measure of the limit set of any finitely generated Kleinian group is 0 or 1; even more, analytical tameness implies that the action of G on its limit set is ergodic when this limit set has full measure. The notion of "geometric tameness" which appears in Problem 5 is this one. As indicated before, Thurston had solved this problem for the Kleinian groups which satisfy (1) and (2) in [92].

The first breakthrough on Problem 5 is due to Francis Bonahon. He shows in [13] that any finitely generated Kleinian group is geometrically tame when it satisfies property (1). In fact, his theorem applies to groups G which satisfy a property weaker than (1), which requires that no parabolic element of G is conjugated to an element which is contained in a factor of a decomposition of G as a free product or as an HNN extension over the trivial group. One of the tools that Bonahon introduced in his proof are *the geodesic currents*; these are the transverse measures to the geodesic foliation on the unit tangent bundle of S. Those objects had been much studied before, in particular in the context of Anosov flows, but they were considered by Bonahon as a natural generalization of the notion of measured geodesic laminations.

The next important progress on Problem 5 after Bonahon is due to Richard Canary: he shows in his thesis, using a beautiful branched covering trick, that, when $N = \mathbb{H}^3/G$ is the interior of a compact 3-manifold, then it is geometrically tame [25]. In the 70's, Al Marden had conjectured that for any finitely generated Kleinian group G, $\mathbb{H}^3/G$ is *topologically tame*, meaning that it is the interior of a compact manifold [61]; this *Marden tameness conjecture* is Problem 9 from Thurston's list. Therefore, if the Marden conjecture is true, it will follow from Canary's theorem that any Kleinian group is geometrically tame.

9. *Are all Kleinian groups topologically tame?*

The Marden conjecture was proven in 2004 independently by Ian Agol [2] and by Danny Calegari and David Gabai [23] following distinct approaches. Teruhiko Soma gave a simplification of the argument of Calegari-Gabai [90] making use of the notion of "disk-busting" that Agol used. Brian Bowditch exposes a self-contained proof in [21] which simplifies the original approach and avoids in particular the use of the "end-reductions" as in the previous references. Canary surveys the history and applications of the tameness theorem in [28].

10. *The Ahlfors measure* 0 *problem.*

Lars Ahlfors showed in [5] that if G is a geometrically finite Kleinian group, then its limit set either has measure 0, or is equal to the whole sphere and he conjectured that the same holds for any finitely generated Kleinian group. Thurston had shown in [92] that when a Kleinian group is geometrically tame and indecomposable, then it satisfies the Ahlfors conjecture (and furthermore the action of G on the boundary of $\mathbb{H}^3$ is ergodic when its limit set has full measure). In [25] Canary shows how to adapt Thurston's argument to the decomposable case. Therefore the Ahlfors conjecture follows from the positive answer to the geometric tameness question (Problem 5) which follows from the truth of the Marden conjecture (Problem 9).

11. *Classify geometrically tame representations of a given group*.

This problem is the *Ending Lamination Conjecture*. It comes with

12. *Describe the quasi-isometry type of a given group*.

Let $N = \mathbb{H}^3/G$ be an hyperbolic 3-manifold with indecomposable fundamental group and let $M \hookrightarrow N$ be a compact core. When N is geometrically tame, Thurston defined *end invariants* which retain geometric information about the ends of N. By Bonahon's theorem, his definition applies to any hyperbolic manifold with indecomposable fundamental group. For simplicity, suppose again that G has no parabolic elements: then each component of ∂M faces exactly one end of N. If the end E facing the component S is geometrically finite, one classical invariant of E is the *conformal structure at infinity*, which is an element of the Teichmüller space of S. When the (geometrically tame) end facing S is not geometrically finite, Thurston defined an *ending lamination*: it is a measured geodesic lamination on S, well-defined up to the transversal measure. Problem 11 asks wether this set of invariants, the elements in Teichmüller space and the ending laminations are sufficient to reconstruct N. When N has indecomposable fundamental group, the problem can be reduced to the case when N has the homotopy type of a closed surface. Minsky solves the case when N has *bounded geometry* (the length of the closed geodesics is bounded from below by a positive constant) in [64]. His proof provides also a *geometric model* for N, i.e., a metric space constructed directly from the end invariants which is quasi-isometric to N (this is the "formula" sought for in Problem 12): in the *doubly degenerate case*, this model is the metric on $S \times \mathbb{R}$ such the metric on the slice $S \times \{t\}$ describes a Teichmüller geodesic between the two ending laminations. Understanding the case in the presence of short geodesics required a lot of efforts and the development of new techniques; one can say that much of the study from the geometric view point of the *curve complex of a surface* that was initiated by Howard Masur and Yair Minsky was motivated by this problem. In [65], Minsky shows that the property that N has bounded geometry in one end can be read on the ending lamination. In [66], he introduces a new model, which depends on the end invariants. This model is a metric on $S \times \mathbb{R}$ which predicts (in terms of the geodesic in the curve complex joining the end invariants) which should be the short geodesics of N and where they should be located; Minsky constructs also a map from this model to N which is Lipschitz on the thick part. In [18] Jeff Brock, Canary and Minsky showthat this model is biLipschitz equivalent to N; they announce also the same result (the existence of a model) for any finitely generated Kleinian group. Another approach to the same result is also explained in [22]. Therefore, if two hyperbolic manifolds have the same end invariants, they are biLipschitz equivalent and then by Sullivan's No Invariant Line Fields Theorem, they are isometric.

6. *Is every Kleinian group a limit of geometrically finite groups?*

Let G be a finitely generated Kleinian group. The space of representations of G into $\mathrm{PSL}(2,\mathbb{C})$ that send each parabolic element to a parabolic element is an affine algebraic set; the faithful representations with discrete image form a closed subset $\mathcal{DF}(G)$ of this space. The geometrically finite representations without accidental parabolics form an open subset $\mathcal{GF}(G) \subset \mathcal{DF}(G)$. The question asks if this subset is dense in $\mathcal{DF}(G)$. When G is a surface group and for representations contained in a *Bers slice*, this had been conjectured by Bers in [7] and it was solved by Kenneth Bromberg in [17] using cone manifolds technics. A positive answer to Thurston's question in full generality is given independently by Hossein Namazi and Juan Souto in [72] and by Ken'ichi Ohshika in [76]; previously Brock and Bromberg solve the case of indecomposable groups without parabolic elements in [19] (see also [18]). The proofs in [72] and [76] depend on the Ending Lamination Theorem (Problem 11). By this theorem, any hyperbolic 3-manifold is determined up to isometry by its topology and by its end invariants. Therefore it suffices to show that for any $\rho \in \mathcal{DF}(G)$, there is a representation ρ_∞ in the closure of $\mathcal{GF}(G)$ which has the same end invariants. The existence of such a ρ_∞ comes from a generalization of the Double Limit Theorem. A different solution to Problem 6, which does not use the full strength of the Ending Lamination Theorem, has been announced by Bromberg and Souto.

8. *Analyse limits of quasi-Fuchsian groups with accidental parabolics.*

One important and difficult step of Thurston's original proof of the Hyperbolization of Haken manifolds is the theorem which says that when a sequence of geometrically tame representations into $\mathrm{PSL}(2,\mathbb{C})$ of the fundamental group of a 3-manifold with incompressible boundary which preserves the type of the elements and converges algebraically, then it converges also geometrically (up to possibly extracting a subsequence) if the limit representation has *no accidental parabolics*. This last hypothesis is really necessary since Kerckhoff and Thurston give an example of a sequence of quasi-Fuchsian representations of a punctured torus group which converges algebraically but such the geometric limit contains a $\mathbb{Z}+\mathbb{Z}$ parabolic subgroup [49]. The geometric limit of an algebraically converging sequence of quasi-Fuchsian groups can even be non finitely generated and for many reasons ([15, 16]). Ohshika and Soma have announced a complete description of the topological type of the limit hyperbolic manifolds, and also a classification up to isometry of those limits in the spirit of the Ending Lamination Theorem [77].

7. *Develop a theory of Schottky groups and their limits.*

In [92], Thurston studies Kleinian groups which satisfy properties (1) and (2) (cf. the discussion of Problem 5), in particular surface groups which are limits of quasi-Fuchsian groups. For those groups he defines the end invariants in §8 and he shows in §9 the theorem mentioned in the last paragraph: "algebraic convergence without accidental parabolics implies geometric convergence". Problem 7 is to develop a similar study for the Kleinian groups which are limits of *Schottky groups*, i.e., limits of geometrically finite free groups.

Brock, Canary and Minsky announce in [18] that using Canary's work on Problem 5, the solution of Marden's conjecture (Problem 9) and extending their work on the Ending lamination Theorem in the boundary incompressible case, they have constructed biLipschitz models for the geometry of $\mathbb{H}^3/G$ when G is an arbitrary finitely generated Kleinian group; they show indeed that each end of $\mathbb{H}^3/G$ is quasi-isometric to an end of a degenerate surface group (with quasi-isometric constants depending on G).

Jørgensen conjectured that for any finitely generated Kleinian group G, a sequence of representations $\rho_i : G \to \mathrm{PSL}(2,\mathbb{C})$ which converges algebraically to a representation ρ without accidental parabolics converges also geometrically to ρ up to extracting a subsequence. This is proven by Jim Anderson and Canary in [6] when the domain of discontinuity of ρ is non empty. When the domain of discontinuity is empty, then any end of $\mathbb{H}^3/\rho(G)$ is degenerate; it follows from the tameness of the algebraic limit and from the Thurston-Canary covering theorem [27] that ρ has finite index in the limit ρ_∞ of any subsequence which converges geometrically, and therefore $\rho = \rho_\infty$.

One could also interpet this problem as Thurston asking for a generalization of the Double Limit Theorem to the context of Schottky groups. He had conjectured a boundedness criterion for a sequence of geometrically finite representations of a free group that he formulated in terms close to those of the Double Limit Theorem but involving laminations in the *Masur domain*. First attempts to prove this conjecture are [78] and [49] (see also [26]). In its most general form, when the free group is replaced by an arbitrary finitely generated Kleinian group, the conjecture is solved by Namazi-Souto in [72] and by Ohshika in [76] (cf. also the announcement by Ingkang Kim, Cyril Lecuire and Ohshika of a slightly more general result in [50]).

13. *If the Hausdorff dimension of the limit set of a Kleinian group is* < 2*, is it geometrically finite?*

Bishop and Jones give a positive answer to this question using analytical tools in [10]. Another solution can be deduced from the geometric tameness of finitely generated Kleinian groups (which was proven after [10]). Canary observes in [24] that if N is a geometrically tame hyperbolic 3-manifold which is not geometrically finite, its Cheeger constant is 0 and therefore $\lambda_0(N)$, the lowest eigenvalue of the Laplacian on N is 0. By a result of Sullivan, for any hyperbolic 3-manifold $N = \mathbb{H}^3/G$, $\lambda_0(N) = \delta_G(2-\delta_G)$, when *the critical exponent* δ_G of G is ≥ 1 or $\lambda_0(N) = 1$. Therefore $\delta_G = 2$ when G is tame and not geometrically finite. By another result from [10] (true for any non-elementary discrete group of isometries of $\mathbb{H}^n$, for any dimension) δ_G equals the Hausdorff dimension of *the radial limit set* of G. Therefore the limit set of G has Hausdorff dimension 2.

14. *Existence of Cannon-Thurston maps.*

Let H be a finitely generated Kleinian group and let M be a compact core of $N = \mathbb{H}^3/H$. It follows from the Hyperbolisation Theorem that there exists a convex cocompact Kleinian group G such that M is homeomorphic to the quotient by G of the reunion $\mathbb{H}^3 \cup \Omega(G)$ of $\mathbb{H}^3$ and of the domain of discontinuity of G. The group G is not unique but its limit set $L(G)$ and the action of G on it are. They can be defined indeed in purely combinatorial terms, due to the fact that G is convex cocompact: $L(G)$ is equivariantly homeomorphic to *the Floyd boundary* of G. Denote by ρ the isomorphism between G and H induced by the homotopy equivalence $M \hookrightarrow N$. The problem asks if there is a continuous G-equivariant map from $L(G)$ to $L(\rho(G))$; such a map is called *a Cannon-Thurston map* since it had been shown to exist when N is the cyclic cover of an hyperbolic manifold which fibers over the circle by Cannon and Thurston in [29] (previously Bill Floyd had constructed the map when N is geometrically finite in [35]). Progress on this problem followed progress on the Ending lamination conjecture. Minsky solves the case when N has the homotopy type of a surface and has bounded geometry [64]. Mahan Mj solves the case of pared manifolds with incompressible boundary and bounded geometry in [56] (see also [51]). More recently, using the Minsky model for surface groups, Mahan Mj solves the general case when N has the homotopy type of a closed surface [57]. He also announces the solution for an arbitrary hyperbolic 3-manifold.

5 Subgroups of the Fundamental Group of 3-Manifolds: Problems 15–18

15. *Are finitely generated subgroups of a Kleinian group separable?*

A group G is LERF (*locally extended residually finite*) if each finitely generated subgroup $H \subset G$ is *separable*, i.e., H equals the intersection of the finite index subgroups of G which contain H. This algebraic property for a group has important geometric consequences: Peter Scott had shown in [88] that closed surface groups are LERF and deduced that any finitely generated subgroup of a surface group $\pi_1(S)$ is the fundamental group of a subsurface in a finite cover of S. Thurston asks the question because of its potential implications to the Geometrization Conjecture. Even if the Geometrization was obtained by quite different routes, the research on the LERF property has led to a vast literature which shows strong and profound connections between geometric group theory and hyperbolic manifolds. In particular, Daniel Wise developed a vast research program to attack this problem which is centered around the notion of *cube complexes* [100]. This program is an elaboration of Scott's approach (for proving that surface groups are LERF, Scott exploited the property that the fundamental group of the non-orientable closed surface of Euler characteristic -1—which is a quotient of any closed surface—acts on $\mathbb{H}$ with a right-angled pentagon as fundamental domain).

A *cube complex* is a cell complex whose cells are isomorphic to cubes $[-1, 1]^n$ and such that the attaching maps between cells are isometries. One says that a group G is *cubulated* when it acts freely and cocompactly on a simply connected CAT(0) cube complex. In 2007, Frédéric Haglund and Daniel Wise introduced the notion of *special cube complex* which imposes some restrictions on the behavior of the *hyperplanes* [39]; they prove that if a compact cube complex X is special and has a Gromov hyperbolic fundamental group, then all the *quasi-convex subgroups* of $\pi_1(X)$ are separable. Wise conjectures in [100] that any compact cube complex which has a Gromov hyperbolic fundamental group has a finite cover that is *special*. This is precisely the conjecture that Agol solves in [4]. The solution of Problems 15 to 18 follows from this.

The separability of quasi-convex subgroups (i.e., convex cocompact subgroups) of a cocompact Kleinian group is a direct application of this result since Nicolas Bergeron and Daniel Wise proved in 2009 that any cocompact Kleinian group is cubulated [8] (subgroups which are not convex cocompact are virtual fibers as a consequence of the tameness and the Thurston-Canary covering theorem). Bergeron and Wise used a method introduced by Michah Sageev who constructed, from any finite collection of quasi-convex subgroups of a group G (which satisfy certain conditions) an action of G on a CAT(0) cube complex. This theorem of Bergeron and Wise could be proven in such a generality thanks to the recent result of Jeremy Kahn and Vlad Markovic which says that any cocompact Kleinian group contains (many) quasi-Fuchsian subgroups [47].

16. *Is every irreducible* 3*-manifold with infinite fundamental group virtually Haken?*

This problem contains the question whether the fundamental group of an irreducible 3-manifold contains a surface group when it is infinite. For a cocompact Kleinian group, this is solved by Kahn and Markovic in [47]; they use in particular fine properties of the geodesic flow on the unit tangent bundle of hyperbolic manifolds such as the *exponential mixing with respect to the Liouville measure*. By the LERF property of Kleinian groups, any surface subgroup of a cocompact Kleinian group G provided by this theorem is the fundamental group of a closed surface which is embedded in an appropriate finite cover of $\mathbb{H}^3/G$. This solves Problem 16 for the case of hyperbolic manifolds which was the remaining case (by the Geometrization Theorem and since all the other geometries can be handled directly).

17. *Does every aspherical* 3*-manifold have a finite cover with positive first Betti number?*

18. *Does every finite volume hyperbolic* 3*-manifold have a finite cover which fibers over the circle?*

In [3], Agol introduces a new class of groups, the *residually finite rational solvable* (RFRS) groups and shows that this class contains all *right-angled Artin groups*. He observes that if the fundamental group of a 3-manifold M is not abelian but RFRS, then M has a first Betti number virtually infinite; he shows also that when M is irreducible with $\chi(M) = 0$ and $\pi_1(M)$ RFRS, then M virtually fibers over the circle. In [39], among the properties of special cube complexes they study, Haglund and Wise show that if a group is cubulated by a special cube complex, then some finite index subgroup of it embeds into a right-angled Artin group. Therefore, any cocompact Kleinian group has a finite index subgroup which is RFRS and therefore Problems 17 and 18 have positive answers.

6 Computer Programs and Tabulations: Problems 20–22

In his PhD thesis, Jeff Weeks wrote the program `SnapPea` which computes the hyperbolic structure of a link complement in $\mathbb{S}^3$ when this structure exists (http://www.geometrygames.org/SnapPea/). This program provides a lot of useful information: it shows the Ford domain, it gives the shape of the cusps and shows the induced tessellation of those, it computes many invariants of the hyperbolic metric like the volume, the length of its shortest closed geodesic etc. [99]. It allows to tabulate many families of 3-manifolds like the Callahan-Hildebrand-Weeks census of non compact finite volume hyperbolic 3-manifolds, or the Hodgson-Weeks census of small volume closed hyperbolic 3-manifolds. It shows evidence for several problems in the list; in particular in [34], Nathan Dunfield and Thurston check the virtual Haken conjecture for all 3-manifolds from the Hodgson-Weeks census (all have indeed positive virtual Betti number). The *Weeks manifold* was observed to have the smallest volume among the known hyperbolic manifolds and was conjectured to be the closed hyperbolic manifold with smallest volume: this conjecture is now solved by David Gabai, Robert Meyerhoff and Peter Milley (see [36], [37]). Furthermore, `SnapPea` permits to "feel" the well ordering of hyperbolic manifolds by the volume as Thurston asks in Problem 22: it gives evidence that this ordering is compatible with the topological complexity given by the *Mom number* [36]. `SnapPea` is a powerful tool, which in addition to helping the intuition with many problems, suggests also new ones. For instance, its computation of the hyperbolic structure on a link complement exhibits in all cases a triangulation of this complement by *geodesic ideal tetrahedra*; however there are no theoretical proofs yet of the existence of such a triangulation.

Other software doing computations with 3-manifolds or with surface diffeomorphisms can be found at the CompuTop.org Software Archive website http://www.math.illinois.edu/~nmd/computop/index.html.

7 Arithmetic Properties of Kleinian Groups: Problems 19, 23

19. *Find topological and geometric properties of quotient spaces of arithmetic subgroups of* $\mathrm{PSL}(2, \mathbb{C})$.

This problem is more a theme of research than a problem like the ones above. Certain topological problems from the list were first established for arithmetic hyperbolic 3-manifolds: Lackenby proves the Surface subgroup conjecture for arithmetic manifolds in [54] and Agol applies his criterion for virtual fibering to Bianchi groups in [3]. A geometric property of arithmetic hyperbolic manifolds that is often used is that each commensurability class of arithmetic 3-manifolds contains many representatives with a non-trivial isometry group. An example of this is the following property which characterizes the arithmetic manifolds among all hyperbolic finite volume 3-manifolds [31]: *any closed geodesic γ in an arithmetic hyperbolic 3-manifold N lifts to a geodesic $\tilde{\gamma}$ in some finite cover N_γ of N which admits an isometric involution whose fixed point set contains $\tilde{\gamma}$*. One central question, closely related to questions in number theory like the Lehmer conjecture, is the Short Geodesic Conjecture: it asks wether the injectivity radius of arithmetic hyperbolic 3-orbifolds is bounded from below by a global positive number (cf. [60], §12). For more information regarding this problem, see the survey by Alan Reid [85].

23. *Show that the volumes of hyperbolic* 3*-manifolds are not all rationally related*.

Today, one knows very little about arithmetic properties of volumes of hyperbolic 3-manifolds and this problem is far from being solved; one does not even know of one single hyperbolic 3-manifold for which one could decide whether its volume is rational or irrational. However, the algebraic framework for studying arithmetic properties of volumes is now well established (see [75]). Given a field $k \subset \mathbb{C}$, its *Bloch group* $\mathcal{B}(k)$ is defined as a certain subspace of a certain quotient of the free $\mathbb{Z}$-module generated by the elements of $k \setminus \{0, 1\}$; there is also a *Bloch regulator map* $\rho : B(k) \to \mathbb{C}/\mathbb{Q}$. In [75], Walter Neumann and Jun Yang assign to any finite volume hyperbolic 3-manifold $N = \mathbb{H}^3/G$ an element $\beta(N) \in \mathcal{B}(k(N)) \subset \mathcal{B}(\mathbb{C})$, where $k(N)$ is the *invariant trace field of* N (i.e., the subfield of $\mathbb{C}$ generated by the squares of the traces of the elements of G). They show that, up to a constant multiple, the volume of N and its Chern-Simons invariant are respectively, the imaginary part and the real part of $\rho(\beta(N))$ (this is one realization of Thurston's hint that volume and Chern-Simons invariant should be considered simultaneously as the real and imaginary parts of the same complex number (see also [101])). It is conjectured that when $k = \overline{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$, the imaginary part of the Bloch regulator map is injective. If this was true, this would imply that two hyperbolic 3-manifolds with the same volume, have Dirichlet domains which are scissors congruent. See [73] and [74] for a detailed discussion and for applications to the study of Chern-Simons invariant.

See also [74] for a discussion of the conjecture that any number field $k \subset \mathbb{C}$ can appear as the invariant trace field $k(N)$ of some hyperbolic 3-manifold N, a conjecture which is directly relevant to Problems 19 and 23.

Acknowledgements I thank Brian Bowditch, Jeff Brock, Dick Canary, Alan Reid and Jeff Weeks for their comments on a first version of this text.

References

1. Agol, I.: Bounds on exceptional Dehn filling. Geom. Topol. **4**, 431–449 (2000)
2. Agol, I.: Tameness of hyperbolic 3-manifolds. math.GT/0405568
3. Agol, I.: Criteria for virtual fibering. Topology **1**, 269–284 (2008)
4. Agol, I., Groves, D., Manning, J.: The virtual Haken conjecture. arXiv:1204.2810
5. Ahlfors, L.: Fundamental polyhedrons and limit point sets of Kleinian groups. Proc. Natl. Acad. Sci. USA **55**, 251–254 (1966)
6. Anderson, J., Canary, R.: Cores of hyperbolic 3-manifolds and limits of Kleinian groups II. J. Lond. Math. Soc. (2) **61**(2), 489–505 (2000)
7. Bers, L.: On boundaries of Teichmüller spaces and on Kleinian groups. Ann. Math. **91**, 570–600 (1970)
8. Bergeron, N., Wise, D.: A boundary criterion for cubulation. Am. J. Math. **134**(3), 843–859 (2012)
9. Bessières, L., Besson, G., Boileau, M., Maillot, S., Porti, J.: Geometrisation of 3-manifolds. EMS Tracts in Mathematics, vol. 13. European Mathematical Society, Zurich (2010)
10. Bishop, C., Jones, P.: Hausdorff dimension and Kleinian groups. Acta Math. **179**, 1–39 (1997)
11. Boileau, M., Leeb, B., Porti, J.: Geometrization of 3-dimensional orbifolds. Ann. Math. **162**, 195–290 (2005)
12. Boileau, M., Porti, J.: Geometrization of 3-Orbifolds of Cyclic Type. Astérisque, vol. 272. SMF, Paris (2001)
13. Bonahon, F.: Bouts des variétés hyperboliques de dimension 3. Ann. Math. **124**, 71–158 (1986)
14. Bonahon, F.: Geometric structures on 3-manifolds. In: Daverman, R.J., Sher, R.B. (eds.) Handbook of Geometric Topology, pp. 93–164. Elsevier, Amsterdam (2001)
15. Bonahon, F., Otal, J.-P.: Variétés hyperboliques avec géodésiques arbitrairement courtes. Bull. Lond. Math. Soc. **20**, 255–261 (1988)
16. Brock, J.: Iteration of mapping classes and limits of hyperbolic 3-manifolds. Invent. Math. **143**, 523–570 (2001)
17. Bromberg, K.: Projective structures with degenerate holonomy and the Bers density conjecture. Ann. Math. (2) **166**, 77–93 (2007)
18. Brock, J., Canary, R., Minsky, Y.: The classification of Kleinian surface groups, II: the ending lamination conjecture. Ann. Math. (2) **176**(1), 1–149 (2012)
19. Brock, J., Bromberg, K.: On the density of geometrically finite Kleinian groups. Acta Math. **192**, 33–93 (2004)
20. Boyer, S., Zhang, X.: Finite Dehn surgery on knots. J. Am. Math. Soc. **9**(4), 1005–1050 (1996)
21. Bowditch, B.: Notes on tameness. Enseign. Math. (2) **56**(3–4), 229–285 (2010)
22. Bowditch, B.: The ending lamination theorem. Preprint Warwick, September 2011
23. Calegari, D., Gabai, D.: Shrinkwrapping and the taming of hyperbolic 3-manifolds. J. Am. Math. Soc. **19**, 385–446 (2006)
24. Canary, R.: On the Laplacian and the geometry of hyperbolic 3-manifolds. J. Differ. Geom. **36**(2), 349–367 (1992)
25. Canary, R.: Ends of hyperbolic 3-manifolds. J. Am. Math. Soc. **6**(1), 1–35 (1993)
26. Canary, R.: Algebraic convergence of Schottky groups. Trans. Am. Math. Soc. **337**, 235–258 (1993)
27. Canary, R.: A covering theorem for hyperbolic 3-manifolds and its applications. Topology **35**(3), 751–778 (1996)
28. Canary, R.: Marden's tameness conjecture: history and applications. In: Geometry, Analysis and Topology of Discrete Groups. Adv. Lect. Math., vol. 6, pp. 137–162. International Press, Somerville (2008)
29. Cannon, J., Thurston, W.: Group invariant Peano curves. Geom. Topol. **11**, 1315–1356 (2007)
30. Cooper, D., Hodgson, C., Kerckhoff, S.: Three-dimensional orbifolds and cone-manifolds. MSJ Memoirs, vol. 5. Math. Soc. of Japan, Tokyo (2000)
31. Cooper, D., Long, D., Reid, A.: On the virtual Betti numbers of arithmetic hyperbolic 3-manifolds. Geom. Topol. **11**, 2265–2276 (2007)
32. Culler, M., Gordon, C., Luecke, J., Shalen, P.: Dehn surgery on knots. Ann. Math. **125**, 237–300 (1987)
33. Dinkelbach, J., Leeb, B.: Equivariant Ricci flow with surgery and applications to finite group actions on geometric 3-manifolds. Geom. Topol. **13**(2), 1129–1173 (2009)
34. Dunfield, N., Thurston, W.: The virtual Haken conjecture: experiments and examples. Geom. Topol. **7**, 399–441 (2003)

35. Floyd, W.: Group completions and limit sets of Kleinian groups. Invent. Math. **57**, 205–218 (1980)
36. Gabai, D., Meyerhoff, R., Milley, P.: Mom technology and hyperbolic 3-manifolds. In: Bonk, M., et al. (eds.) the Tradition of Ahlfors-Bers, V. Contemp. Math., vol. 510, pp. 81–108 (2010)
37. Gabai, D., Meyerhoff, R., Milley, P.: Mom technology and volumes of hyperbolic 3-manifolds. Comment. Math. Helv. **86**, 145–188 (2011)
38. Greenberg, L.: Fundamental polyhedra for kleinian groups. Ann. Math. **84**, 433–441 (1966)
39. Haglund, F., Wise, D.: Special cube complexes. Geom. Funct. Anal. **17**, 1551–1620 (2008)
40. Hamilton, R.: Three-manifolds with positive Ricci curvature. J. Differ. Geom. **17**, 255–306 (1982)
41. Hatcher, A., Thurston, W.: Incompressible surfaces in 2-bridge knot complements. Invent. Math. **79**(2), 225–246 (1985)
42. Hempel, J.: 3-Manifolds as viewed from the curve complex. Topology **40**(3), 631–657 (2001)
43. Hodgson, C., Kerckhoff, S.: Universal bounds for hyperbolic Dehn surgery. Ann. Math. (2) **162**(1), 367–421 (2005)
44. Hodgson, C., Kerckhoff, S.: The shape of hyperbolic Dehn surgery space. Geom. Topol. **12**(2), 1033–1090 (2008)
45. Jørgensen, T.: Compact 3-manifolds of constant negative curvature fibering over the circle. Ann. Math. (2) **106**(1), 61–72 (1977)
46. Jørgensen, T.: On pairs of once-punctured tori. In: Kleinian Groups and Hyperbolic 3-Manifolds, Warwick, 2001. London Math. Soc., Lecture Notes Ser., vol. 299, pp. 183–207. Cambridge University Press, Cambridge (2003)
47. Kahn, J., Markovic, V.: Immersing almost geodesic surfaces in a closed hyperbolic three-manifold. Ann. Math. **175**(3), 1127–1190 (2012)
48. Kapovich, M.: Hyperbolic Manifolds and Discrete Groups. Birkhäuser Classics, Boston (2001)
49. Kerckhoff, S., Thurston, W.: Non-continuity of the action of the modular group at Bers boundary of Teichmüller space. Invent. Math. **100**, 25–47 (1990)
50. Kim, I., Lecuire, C., Ohshika, K.: Convergence of freely decomposable Kleinian groups. arXiv:0708.3266
51. Klarreich, E.: Semi-conjugacies between Kleinian group actions on the Riemann sphere. Am. J. Math. **121**(5), 1031–1078 (1999)
52. Kleiner, B., Lott, J.: Notes on Perelman's papers. Geom. Topol. **12**, 2587–2858 (2008)
53. Lackenby, M.: Word hyperbolic Dehn surgery. Invent. Math. **140**, 243–282 (2000)
54. Lackenby, M.: Surface subgroups of Kleinian groups with torsion. Invent. Math. **179**, 175–190 (2010)
55. Lackenby, M., Meyerhoff, R.: The maximal number of exceptional Dehn surgeries. Invent. Math. **191**, 341–382 (2013)
56. Mahan, M.: Cannon-Thurston maps for pared manifolds of bounded geometry. Geom. Topol. **13**, 189–245 (2009). arXiv:math/0503581
57. Mahan, M.: Cannon-Thurston maps for surface groups. arXiv:math/0607509
58. McMullen, C.: Amenability, Poincaré series and quasiconformal maps. Invent. Math. **97**, 95–128 (1989)
59. McMullen, C.: Iteration on Teichmüller space. Invent. Math. **99**, 425–454 (1990)
60. Maclachlan, C., Reid, A.: The Arithmetic of Hyperbolic 3-Manifolds. Graduate Texts in Mathematics, vol. 219. Springer, New York (2003)
61. Marden, A.: The geometry of finitely generated Kleinian groups. Ann. Math. **99**, 383–462 (1974)
62. Meeks, W. III, Scott, P.: Finite group actions on 3-manifolds. Invent. Math. **86**, 287–346 (1986)
63. Meeks, W. III, Yau, S.-T.: Group actions on $\mathbb{R}^3$. In: Morgan, J., Bass, H. (eds.) The Smith Conjecture, New York, 1979. Pure Appl. Math., vol. 112, pp. 167–179. Academic Press, New York (1984)
64. Minsky, Y.: On rigidity, limit sets and end invariants of hyperbolic 3-manifolds. J. Am. Math. Soc. **7**, 539–588 (1994)
65. Minsky, Y.: Bounded geometry for Kleinian groups. Invent. Math. **146**(1), 143–192 (2001)
66. Minsky, Y.: The classification of Kleinian surface groups. I. Models and bounds. Ann. Math. **171**, 1–107 (2010)
67. Morgan, J., Bass, H. (eds.): The Smith Conjecture, New York, 1979. Pure Appl. Math., vol. 112. Academic Press, New York (1984)
68. Morgan, J.: On Thurston's uniformization theorem for three-dimensional manifolds. In: Morgan, J., Bass, H. (eds.) The Smith Conjecture, New York, 1979. Pure Appl. Math., vol. 112, pp. 37–125. Academic Press, New York (1984)
69. Morgan, J.: Recent progress on the Poincaré conjecture and the classification of 3-manifolds. Bull. Am. Math. Soc. **42**, 57–78 (2005)

70. Morgan, J., Tian, G.: Ricci Flow and the Poincaré Conjecture. Clay Mathematics Monographs, vol. 3 (2007)
71. Namazi, H., Souto, J.: Heegaard splittings and pseudo-Anosov maps. Geom. Funct. Anal. **19**, 1195–1228 (2009)
72. Namazi, H., Souto, J.: Non-realizability and ending laminations: proof of the density conjecture. Acta Math. **209**, 323–395 (2012)
73. Neumann, W.: Hilbert's 3rd problem and invariants of 3-manifolds. In: The Epstein Birthday Schrift. Geometry & Topology Monographs, vol. 1, pp. 383–411 (1998)
74. Neumann, W.: Realizing arithmetic invariants of hyperbolic 3-manifolds. In: Interactions Between Hyperbolic Geometry, Quantum Topology and Number Theory. Contemp. Math., vol. 541, pp. 233–246. AMS, Reading (2011)
75. Neumann, W., Yang, J.: Bloch invariants of hyperbolic 3-manifolds. Duke Math. J. **96**(1), 29–59 (1999)
76. Ohshika, K.: Realising end invariants by limits of minimally parabolic, geometrically finite groups. Geom. Topol. **15**, 827–890 (2011)
77. Ohshika, K., Soma, T.: Geometry and topology of geometric limits, I. arXiv:1002.4266v1
78. Otal, J.-P.: Sur la dégénérescence des groupes de Schottky. Duke Math. J. **74**, 777–792 (1994)
79. Otal, J.-P.: Le théorème d'hyperbolisation pour les variétés fibrées de dimension 3. Astérisque, vol. 235. Soc. Math. France, Paris (1996)
80. Otal, J.-P.: Thurston's hyperbolization of Haken manifolds. In: Hsiung, C.-C., Yau, S.-T. (eds.) Surveys in Differential Geometry II, pp. 77–194. International Press, Boston (1998)
81. Perelman, G.: The entropy formula for the Ricci flow and its geometric applications. arXiv:math.DG/0211159
82. Perelman, G.: Ricci flow with surgery on three-manifolds. arXiv:math.DG/0303109
83. Perelman, G.: Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. arXiv:math.DG/0307245
84. Poincaré, H.: Cinquième complément à l'analysis situs. Rend. Circ. Mat. Palermo **18**, 45–110 (1904). See also: Oeuvres, Tome VI, 1953
85. Reid, A.: The geometry and topology of arithmetic hyperbolic 3-manifolds. In: Proc. Symposium Topology, Complex Analysis and Arithmetic of Hyperbolic Spaces, Kyoto, 2006. RIMS Kokyuroku Series, vol. 1571, pp. 31–58 (2007)
86. Riley, R.: Discrete parabolic representations of link groups. Mathematika **22**, 141–150 (1975)
87. Riley, R.: A personal account of the discovery of hyperbolic structures on some knot complements. Expo. Math. **31**, 104–115 (2013)
88. Scott, P.: Subgroups of surface groups are almost geometric. J. Lond. Math. Soc. **17**(3), 555–565 (1978)
89. Shalen, P.: The proof in the case of no incompressible surface. In: Morgan, J., Bass, H. (eds.) The Smith Conjecture, New York, 1979. Pure Appl. Math., vol. 112, pp. 21–36. Academic Press, New York (1984)
90. Soma, T.: Existence of ruled wrappings in hyperbolic 3-manifolds. Geom. Topol. **10**, 1173–1184 (2006)
91. Sullivan, D.: Travaux de Thurston sur les groupes quasi-fuchsiens et sur les variétés hyperboliques de dimension 3 fibrées sur $\mathbb{S}^1$. In: Séminaire Bourbaki 554. Lecture Notes in Math., vol. 842, pp. 196–214. Springer, Berlin (1981)
92. Thurston, W.: The geometry and topology of 3-manifolds. Lecture notes, Princeton Math. Dept, 1979
93. Thurston, W.: Hyperbolic geometry and 3-manifolds. In: Low-dimensional topology, Bangor, 1979. London Math. Soc., Lecture Notes Ser., vol. 48, pp. 9–25. Cambridge University Press, Cambridge (1982)
94. Thurston, W.: Hyperbolic structures on 3-manifolds: the overall logic. Notes of a workshop at Bowdoin (1980)
95. Thurston, W.: Hyperbolic structures on 3-manifolds, I, deformations of acylindrical manifolds. Ann. Math. **124**, 203–246 (1986). arXiv:math/9801019
96. Thurston, W.: On the geometry and dynamics of surface diffeomorphisms. Bull. Am. Math. Soc. **19**, 417–431 (1988)
97. Thurston, W.: Hyperbolic structures on 3-manifolds, II: surface groups and 3-manifolds which fiber over the circle. arXiv:math/9801058

98. Thurston, W.: Hyperbolic Structures on 3-manifolds, III: deformations of 3-manifolds with incompressible boundary. arXiv:math/9801045
99. Weeks, J.: Computation of hyperbolic structures in knot theory. In: Menasco, B., Thistlethwaite, M. (eds.) Handbook of Knot Theory. Elsevier, Amsterdam (2005)
100. Wise, D.: The structure of groups with a quasi-convex hierarchy. Elec. Res. Announ. Math. Sci. **16**, 44–55 (2009)
101. Yoshida, T.: The η-invariant of hyperbolic 3-manifolds. Invent. Math. **81**, 473–514 (1985)

Jahresber Dtsch Math-Ver (2014) 116:21–41
DOI 10.1365/s13291-014-0083-9

SURVEY ARTICLE

The Excitation Spectrum for Bose Fluids with Weak Interactions

Robert Seiringer

Published online: 20 February 2014

Abstract We review recent progress towards a rigorous understanding of the excitation spectrum of bosonic quantum many-body systems. In particular, we explain how one can rigorously establish the predictions resulting from the Bogoliubov approximation in the mean field limit. The latter predicts that the spectrum is made up of elementary excitations, whose energy behaves linearly in the momentum for small momentum. This property is crucial for the superfluid behavior of the system. We also discuss a list of open problems in this field.

Keywords Schrödinger equation · Quantum statistical mechanics · Bose–Einstein condensation · Dilute Bose gas · Superfluidity · Excitation spectrum

Mathematics Subject Classification 82B10 · 82-02 · 46N50

1 Introduction

Many interesting effects in quantum mechanics result from the interactions among the fundamental particles that constitute the system. A famous example of such an effect is superconductivity, the vanishing of electrical resistance in certain materials at low enough temperature. The relevant fundamental particles in this case are the electrons, which are *fermions* and obey the Pauli exclusion principle, which demands that each quantum state can be occupied by at most one particle. *Bosons*, on the other hand, are particles in quantum mechanics that do not obey the Pauli principle. Examples of bosons include photons (the quanta of the electromagnetic field) or also composite

R. Seiringer (✉)
IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria
e-mail: robert.seiringer@ist.ac.at

particles like atoms, which themselves consists of fermions (electrons and nucleons) but behave as bosons if the number of fermionic constituents is even. There is no bound on the number of bosons occupying the same quantum state. This leads to the phenomenon of *Bose–Einstein condensation* (BEC), which occurs if a macroscopic fraction of all the particles occupy the same quantum state. The resulting state of matter displays various interesting phenomena, like superfluidity, for instance, where the viscosity of a fluid vanishes completely at low temperature.

BEC in cold atomic gases was first achieved experimentally in 1995 [1, 9]. After initial failed attempts with spin-polarized atomic hydrogen, the first successful demonstrations of this phenomenon used gases of rubidium and sodium atoms, respectively. In these experiments, a large number of (bosonic) atoms is confined to a trap and cooled to very low temperatures; below a critical temperature condensation of a large fraction of particles into the same one-particle state occurs. Since then there has been a surge of activity in this field, with ingenious experiments putting forth more and more astonishing results about the behavior of matter at very cold temperatures. BEC has now been achieved by more than a dozen different research groups working with gases of different types of atoms. Literally thousands of scientific articles, concerning both theory and experiment, have been published in recent years. Various interesting quantum phenomena have been explored, like the appearance of quantized vortices in rotating systems and the property of superfluidity. The latter is related to the low-energy excitation spectrum of the system. We refer to [3, 6, 8, 13] for reviews of the recent developments in this field of physics.

The theoretical investigation of BEC goes back much further, and even pre-dates the modern formulation of quantum mechanics. It was investigated in two papers by Einstein [12] in 1924 and 1925, respectively, following up on a work by Bose [5] on the derivation of Planck's law for black-body radiation. Einstein's result, in its modern formulation, can be found in any textbook on quantum statistical mechanics, and was concerned with ideal, i.e., non-interacting gases. The understanding of BEC in the presence of interparticle interactions poses a formidable challenge to mathematical physics. One of the key contributions to the theory of weakly interacting Bose gases is Bogoliubov's 1947 paper [4], where he introduces an approximate model (now referred to as the Bogoliubov approximation) to explain its superfluid behavior. In this paper, we will summarize recent progress made towards a rigorous justification of this approximation.

2 The Bose Gas: A Quantum Many-Body Problem

The quantum-mechanical description of a system of N bosons is given in terms of the Hamiltonian H_N, acting as a linear operator on a suitable Hilbert space $\mathscr{H}_N$. Typically, H_N is an unbounded operator, defined only on a dense subspace of $\mathscr{H}_N$, but it should be bounded from below in order to describe a stable physical system. For bosons interacting via a pair-interaction potential denoted by $v(x)$, the Hamiltonian is given, in appropriate units, by

$$H_N = -\sum_{i=1}^{N} \Delta_i + \sum_{1\leq i<j\leq N} v(x_i - x_j). \tag{1}$$

The kinetic energy is described by Δ, the Laplacian on a suitable domain in $\mathbb{R}^3$, which we will typically take to be a cube of side length L, i.e., $[0, L]^3$. Suitable boundary conditions have to be imposed in order for Δ to define a self-adjoint operator, with periodic boundary conditions being a typical example. The subscript i indicates, as usual, that the second derivative is with respect to $x_i \in \mathbb{R}^3$.

As appropriate for bosons, the Hamiltonian H_N acts on the Hilbert space of *permutation-symmetric* wave functions $\Psi(x_1, \ldots, x_N)$ in $\bigotimes^N L^2([0, L]^3)$, which we shall denote by $\mathscr{H}_N$:

$$\mathscr{H}_N = \bigotimes_{\text{sym}}^N L^2\big([0, L]^3\big). \tag{2}$$

The interaction v is a real-valued function $v : \mathbb{R}^3 \to \mathbb{R}$, which we assume to be bounded and symmetric, i.e., $v(x) = v(-x)$. It acts as a multiplication operator on $\mathscr{H}_N$.

Of fundamental importance is the *spectrum* of H_N, i.e., the complement of the subset of $\mathbb{C}$ where $z - H_N$ has a bounded inverse. For the Hamiltonian H_N acting on the Hilbert space $\mathscr{H}_N$, it is not difficult to see that the spectrum is discrete, i.e., it consists of eigenvalues of H_N of finite multiplicity, which are bounded from below and accumulate at $+\infty$. The corresponding eigenfunctions describe the stationary states of the system.

The following quantities, derived from the Hamiltonian H_N, will interest us here.

- *Ground state energy*, defined as the lowest value of the spectrum of the Hamiltonian,

$$E_0(N, L) = \inf \operatorname{spec} H_N. \tag{3}$$

- The ground state wave function Ψ_0 is the eigenfunction of H_N corresponding to eigenvalue $E_0(N, L)$, i.e.,

$$H_N \Psi_0 = E_0(N, L)\Psi_0. \tag{4}$$

  For large particle number N, it is typically much too complicated to compute. Instead one considers the corresponding reduced density matrices of Ψ_0, the simplest of which is the *one-particle density matrix*, given by the integral kernel

$$\gamma_0\big(x, x'\big) = N \int_{\mathbb{R}^{3(N-1)}} \Psi_0(x, x_2, \ldots, x_N)\overline{\Psi_0\big(x', x_2, \ldots, x_N\big)}dx_2 \cdots dx_N. \tag{5}$$

  It satisfies $0 \le \gamma_0 \le N$ as an operator, and its trace equals $\operatorname{Tr} \gamma_0 = N$. With the aid of creation and annihilation operators (to be reviewed in Sect. 4) one can also write

$$\gamma_0\big(x, x'\big) = \big\langle a^\dagger\big(x'\big)a(x)\big\rangle, \tag{6}$$

  and this definition generalizes to arbitrary mixed states as well.
- The diagonal of the one-particle density matrix is the *particle density*

$$\varrho_0(x) = \gamma_0(x, x) = N \int_{\mathbb{R}^{3(N-1)}} |\Psi_0(x, x_2, \ldots, x_N)|^2 dx_2 \cdots dx_N, \tag{7}$$

with $\int \varrho_0(x)dx = N$. For translation invariant systems, ϱ_0 is a constant and does not depend on x, but for inhomogeneous systems the spatial variation of ϱ_0 represents a non-trivial question.

- By definition, *Bose–Einstein condensation* in a state Ψ_0 means that the one-particle density matrix γ_0 has an eigenvalue of order N, i.e., that $\|\gamma_0\| \geq cN$ for some $c > 0$ and all (large) N, with $\|\cdot\|$ denoting the operator norm. The corresponding eigenfunction is called the *condensate wave function*.
- Of particular interest to us will be the structure of the *excitation spectrum*, i.e., the spectrum of H_N above the ground state energy $E_0(N, L)$, and the relation of the corresponding eigenstates to the ground state. For translation invariant systems, H_N commutes with the total momentum operator

$$P = -i \sum_{j=1}^{N} \nabla_j, \tag{8}$$

and hence one can look at their joint spectrum. Of particular relevance is the infimum

$$E_q(N, L) = \inf \operatorname{spec} H_N \restriction_{P=q} . \tag{9}$$

In contrast to the non-interacting case, for interacting particles one expects a linear behavior of $E_q(N, L)$ in q for not too large values of $|q|$. For a review of various questions related to the excitation spectrum of Bose gases we refer to [7].

The particle number N is typically very large. This large number of variables involved in the problem is the main reason why the quantities above are very hard to compute. We will be interested in their behavior as $N \to \infty$.

3 The Ideal Bose Gas

For *non-interacting bosons*, i.e., in the case $v \equiv 0$, the ground state energy is simply N times the lowest eigenvalue of the Laplacian. In the case of periodic boundary conditions, i.e., the Laplacian on the flat torus $[0, L]^3$, this is simply zero:

$$E_0(N, L) = 0 \quad \text{for all } N \text{ and } L. \tag{10}$$

The corresponding ground state wave function Ψ_0 is the constant function in $\mathscr{H}_N$.

Also the excitation spectrum can easily be computed explicitly for the ideal gas. The spectrum of $-\Delta$ on the flat torus $[0, L]^3$ equals

$$\left\{ |p|^2 : p \in \left(\frac{2\pi}{L} \mathbb{Z} \right)^3 \right\}, \tag{11}$$

with corresponding eigenfunctions $L^{-3/2} e^{ip\cdot x}$. The spectrum of N bosons is then simply

$$\sum_p |p|^2 n_p, \tag{12}$$

where the sum is over $p \in (\frac{2\pi}{L}\mathbb{Z})^3$ and $n_p \in \{0, 1, 2, \ldots\}$ for each p, with $\sum_p n_p = N$. The latter are called the *occupation numbers* of the corresponding momentum states. The eigenstate of H_N corresponding to an eigenvalue of the form (12) is given by

$$\mathcal{S} \prod_j \varphi_j(x_j) \tag{13}$$

where $\mathcal{S}$ denotes symmetrization with respect to permutations, and all the φ_j are eigenfunctions of the Laplacian, the one corresponding to momentum p appearing n_p times in the product. Note that for each set of occupations numbers $\{n_p\}$ there is exactly one such eigenstate in $\mathscr{H}_N$.

4 Second Quantization on Fock Space

In the following, it will be convenient to regard $\mathscr{H}_N = \bigotimes_{\mathrm{sym}}^N L^2([0, L]^3)$ as a subspace of the bosonic *Fock space*

$$\mathcal{F} = \bigoplus_{n=0}^{\infty} \mathscr{H}_n, \qquad \mathscr{H}_0 \equiv \mathbb{C}. \tag{14}$$

On this space, the particle number N is now an operator, which acts simply as multiplication by n on the subspace $\mathscr{H}_n$ of $\mathcal{F}$.

A basis of $L^2([0, L]^3)$ is given by the plane waves $L^{-3/2}e^{ip\cdot x}$ for $p \in (\frac{2\pi}{L}\mathbb{Z})^3$, and we introduce the corresponding *creation and annihilation operators*, which satisfy the canonical commutation relations (CCR)

$$[a_p, a_q] = \left[a_p^\dagger, a_q^\dagger\right] = 0, \qquad \left[a_p, a_q^\dagger\right] = \delta_{p,q}. \tag{15}$$

Here, the creation operator $a_p^\dagger$ is the adjoint of the annihilation operator a_p. The latter maps $\mathscr{H}_n$ to $\mathscr{H}_{n-1}$ for $n \geq 1$ and acts as

$$(a_p \psi)(x_1, \ldots, x_{n-1}) = \sqrt{\frac{n}{L^3}} \int_{[0,L]^3} e^{-ip\cdot x_n} \psi(x_1, \ldots, x_n) dx_n. \tag{16}$$

We consider again the Hamiltonian H_N, with the particles moving on the flat torus $[0, L]^3$. It is then natural to assume that the interaction v in (1) is a periodic function on $\mathbb{R}^3$, with period L in all three coordinate directions. In other words,

$$v(x) = L^{-3} \sum_{p \in (\frac{2\pi}{L}\mathbb{Z})^3} \widehat{v}(p) e^{ip\cdot x} \tag{17}$$

where the

$$\widehat{v}(p) = \int_{[0,L]^3} v(x) e^{-ip\cdot x} dx \tag{18}$$

are the Fourier coefficients of v.

A simple calculation shows that the Hamiltonian H_N in (1) is equal to the restriction of

$$\mathbb{H} = \sum_p |p|^2 a_p^\dagger a_p + \frac{1}{2L^3} \sum_p \widehat{v}(p) \sum_{q,k} a_{q+p}^\dagger a_{k-p}^\dagger a_k a_q \tag{19}$$

to the subspace $\mathscr{H}_N \subset \mathcal{F}$. Here, all sums are over $(\frac{2\pi}{L}\mathbb{Z})^3$. The expression (19) for the Hamiltonian on Fock space serves as a basis for the approximation introduced by Bogoliubov in [4], which we shall discuss next.

5 The Bogoliubov Approximation

At low energy, and for sufficiently weak interactions, one expects the occurrence of Bose–Einstein condensation. That is, the zero momentum mode is expected to be macroscopically occupied, meaning that $a_0^\dagger a_0 \sim N$. In particular, the $p = 0$ mode plays a special role.

The *Bogoliubov approximation* consists of

- dropping all terms in $\mathbb{H}$ higher than quadratic in $a_p^\dagger$ and a_p for $p \neq 0$;
- replacing $a_0^\dagger$ and a_0 in $\mathbb{H}$ by the number $\sqrt{N}$.

The resulting Hamiltonian is quadratic in the $a_p^\dagger$ and a_p, and equals[1]

$$\begin{aligned}\mathbb{H}^{\text{Bog}} &= \frac{N(N-1)}{2L^3}\widehat{v}(0) \\ &\quad + \sum_{p\neq 0}\Big(\big(|p|^2 + \varrho\widehat{v}(p)\big)a_p^\dagger a_p + \frac{1}{2}\varrho\widehat{v}(p)\big(a_p^\dagger a_{-p}^\dagger + a_p a_{-p}\big)\Big),\end{aligned} \tag{20}$$

with $\varrho = N/L^3$ the particle density. It can be explicitly diagonalized via a *Bogoliubov transformation*:

Let $b_p = \cosh(\alpha_p)a_p + \sinh(\alpha_p)a_{-p}^\dagger$, with

$$\tanh(\alpha_p) = \frac{|p|^2 + \varrho\widehat{v}(p) - \sqrt{|p|^4 + 2|p|^2\varrho\widehat{v}(p)}}{\varrho\widehat{v}(p)}. \tag{21}$$

Here, the right side is interpreted as 0 if $\widehat{v}(p) = 0$. Moreover, we have to *assume* that $|p|^2 + 2\varrho\widehat{v}(p) \geq 0$ for all p in order for the square root to be well-defined. The b_p and $b_p^\dagger$ again satisfy CCR (for any choice of real numbers α_p, in fact). A simple calculation shows that

$$\mathbb{H}^{\text{Bog}} = E_0^{\text{Bog}} + \sum_{p\neq 0} e_p b_p^\dagger b_p, \tag{22}$$

[1] Note that the contribution of $p = 0$ to the second sum in (19) is exactly equal to $N(N-1)\widehat{v}(0)/(2L^3)$, hence the substitution of $a_0^\dagger$ and a_0 by $\sqrt{N}$ was not applied to this term.

where

$$E_0^{\rm Bog} = \frac{N(N-1)}{2L^3}\widehat{v}(0) - \frac{1}{2}\sum_{p\neq 0}\left(|p|^2 + \varrho\widehat{v}(p) - \sqrt{|p|^4 + 2|p|^2\varrho\widehat{v}(p)}\right) \tag{23}$$

and

$$e_p = \sqrt{|p|^4 + 2|p|^2\varrho\widehat{v}(p)}. \tag{24}$$

Note that in contrast to the non-interacting case, where $e_p = p^2$, the function e_p in (24) behaves linearly in p for small p (assuming that $\widehat{v}(p)$ does not vanish near zero).

The Bogoliubov approximation thus predicts that the ground state energy equals the value $E_0^{\rm Bog}$ displayed in (23). Moreover, it also allows to compute the complete excitation spectrum. In fact, from (22) we see that the spectrum of $\mathbb{H}^{\rm Bog} - E_0^{\rm Bog}$ is given by

$$\sum_p e_p n_p \quad \text{with } n_p \in \{0, 1, 2, \ldots\}, \tag{25}$$

with e_p defined in (24). It has the exact same structure as for non-interacting particles (12), except for the replacement of $|p|^2$ by e_p. Moreover, the corresponding eigenstates can be constructed out of the ground state $\Psi_0 \in \mathcal{F}$ of $\mathbb{H}^{\rm Bog}$ by *elementary excitations* of the form

$$b_{p_n}^\dagger \cdots b_{p_1}^\dagger \Psi_0, \tag{26}$$

with $b_p^\dagger = \cosh(\alpha_p)a_p^\dagger + \sinh(\alpha_p)a_{-p}$, as before.

One can also calculate the ground state energy $E_q^{\rm Bog}$ in a sector of total momentum q, and arrives at

$$\begin{aligned} E_q^{\rm Bog} - E_0^{\rm Bog} &= \text{subadditive hull of } e_p \\ &= \inf_{\sum_p p n_p = q} \sum_p e_p n_p. \end{aligned} \tag{27}$$

In particular, also $E_q^{\rm Bog} - E_0^{\rm Bog}$ behaves linearly in q for not too large $|q|$.

For a detailed discussion of variants of the Bogoliubov approximation we refer the interested reader to [39].

6 The Mean-Field (Hartree) Limit

It is a major open problem to understand the regime of validity of the Bogoliubov approximation for many-body quantum systems of interacting particles. The progress made in recent years was mainly limited to giving bounds on the ground state energy of the systems, and we refer to [14, 15, 25, 26, 30, 36–38] for various interesting results in this direction. Virtually nothing is known concerning the excitation spectrum of such systems in general, however.

A simple case where the analysis of the validity of the Bogoliubov approximation can be extended beyond the ground state energy is the Hartree limit. This is an extreme form of a *mean-field limit*, where the interaction potential extends over the whole size of the system, but the interaction is sufficiently weak (of order $1/N$) in order for the interaction energy to be of the same order as the kinetic energy.

We consider again a system of N bosons in a cubic box, with periodic boundary conditions. For simplicity, let us choose units such that the length of the box L equals 1. The Hamiltonian of the systems is thus given by

$$H_N = -\sum_{i=1}^{N} \Delta_i + \frac{1}{N-1} \sum_{1\leq i<j\leq N} v(x_i - x_j), \tag{28}$$

and it acts on the Hilbert space

$$\mathscr{H}_N = \bigotimes_{\text{sym}}^{N} L^2\big([0,1]^3\big). \tag{29}$$

Here we wrote the interaction potential as $(N-1)^{-1}v(x)$, reflecting the weakness of the potential as mentioned above. The case of fixed, N-independent v corresponds to the mean-field or *Hartree limit*.

It is not difficult to see that the ground state energy is determined, to leading order in N for large N, by minimizing the energy $\langle \Psi | H_N | \Psi \rangle$ over product states of the form

$$\Psi(x_1,\ldots,x_N) = \phi(x_1)\cdots\phi(x_N). \tag{30}$$

This has been shown, in a much more general setting than what is discussed here, in [19]. For a constant ϕ, corresponding to a homogeneous system, the resulting Hartree energy is then simply equal to $\frac{1}{2}N\int v$.

It is also known that starting from a product state of the form (30), a solution to the time-dependent Schrödinger equation $i\partial_t \Psi = H_N \Psi$ stays roughly a product at later times, with the factors in the limit $N\to\infty$ determined by the time-dependent Hartree equation

$$i\partial_t \phi = -\Delta\phi + 2\big(|\phi|^2 * v\big)\phi, \tag{31}$$

where $*$ denotes convolution. For a history of this problem and a review of recent results, we refer to [31].

Going beyond the leading order, where the Hartree equation applies, we can ask the following questions.

- Given that the ground state energy $E_0(N) = \inf \operatorname{spec} H_N$ satisfies $E_0(N) = \frac{1}{2}N\widehat{v}(0) + o(N)$ for fixed (i.e., N-independent) v, what is the next order correction? It turns out that it is actually $O(1)$, and the $O(1)$-term can be explicitly computed and agrees with the prediction from the Bogoliubov approximation.
- What is the spectrum of $H_N - E_0(N)$, i.e., the excitation spectrum of the system? Does it converge as $N\to\infty$? Is the Bogoliubov approximation valid? The latter predicts a dispersion law for elementary excitations that is *linear* for small momentum, as discussed in Sect. 5.

- What fraction of particles are in a Bose–Einstein condensate? Recall that Bose–Einstein condensation concerns the largest eigenvalue of the one-particle density matrix γ of a many-body wave function Ψ, defined via the matrix elements

$$\langle f|\gamma|g\rangle = N\int \overline{f(x)}\Psi(x,x_2,\ldots,x_N)g(y)\overline{\Psi(y,x_2,\ldots,x_N)}dxdydx_2\cdots dx_N. \tag{32}$$

  For fixed v, the Bogoliubov approximation predicts that $\|\gamma\| \geq N - O(1)$ in the ground state, and this can actually be proved to be correct.

6.1 Main Results

For our analysis of the excitation spectrum, we assume that $v(x)$ is bounded and of positive type, i.e.,

$$v(x) = \sum_{p\in(2\pi\mathbb{Z})^3} \widehat{v}(p)e^{ip\cdot x} \tag{33}$$

with

$$\widehat{v}(p) \geq 0 \quad \forall p \in (2\pi\mathbb{Z})^3, \qquad \sum_{p\in(2\pi\mathbb{Z})^3} \widehat{v}(p) < \infty. \tag{34}$$

Under these assumptions, the following theorem holds.

Theorem 1 *The ground state energy $E_0(N)$ of H_N equals*

$$E_0(N) = \frac{N}{2}\widehat{v}(0) + E_0^{\mathrm{Bog}} + O\big(N^{-1/2}\big) \quad \textit{as } N \to \infty, \tag{35}$$

with

$$E_0^{\mathrm{Bog}} = -\frac{1}{2}\sum_{p\neq 0}\Big(|p|^2 + \widehat{v}(p) - \sqrt{|p|^4 + 2|p|^2\widehat{v}(p)}\Big). \tag{36}$$

Moreover, the excitation spectrum of $H_N - E_0(N)$ below an energy ξ is equal to finite sums of the form

$$\sum_{p\in(2\pi\mathbb{Z})^3\setminus\{0\}} e_p n_p + O\big(\xi^{3/2}N^{-1/2}\big), \tag{37}$$

where

$$e_p = \sqrt{|p|^4 + 2|p|^2\widehat{v}(p)} \tag{38}$$

and $n_p \in \{0,1,2,\ldots\}$ for all $p \neq 0$.

Theorem 1 is proved in [34]. The proof consists of constructing a unitary operator U that makes $UH_NU^\dagger$ close to the operator

$$\frac{N}{2}\widehat{v}(0) + E_0^{\mathrm{Bog}} + \sum_{p\in(2\pi\mathbb{Z})^3\setminus\{0\}} e_p a_p^\dagger a_p. \tag{39}$$

In particular, the proof implies that the excited eigenfunctions can be (approximately) obtained by acting with products of $U^\dagger a_p^\dagger a_0 U$ on the ground state.

Let us comment on the error terms in (35) and (37). Both the ground state energy and all excited energy levels a distance $O(1)$ from the ground state agree with the prediction obtained via Bogoliubov's approximation up to errors of order $N^{-1/2}$ for large N. Moreover, an excitation energy a distance ξ from the ground state energy is necessarily of the form $\sum_p e_p n_p(1+o(1))$ as long as $\xi^{3/2}N^{-1/2} \ll \xi$, i.e., for $\xi \ll N$. That is, the Bogoliubov approximation gives the correct excitation energies to leading order in a very large window above the ground state energy, whose size has to be small compared with N. This restriction is presumably optimal. The existence of Bose–Einstein condensation is only guaranteed for excitation energies small compared to N, and the existence of BEC is one of the key assumptions entering the Bogoliubov approximation.

Theorem 1 implies the following corollary concerning the momentum dependence of the spectrum of H_N.

Corollary 1 *Let $E_P(N)$ denote the ground state energy of H_N in the sector of total momentum P. We have*

$$E_P(N) - E_0(N) = \min_{\{n_p\},\sum_p p n_p = P} \sum_{p\neq 0} e_p n_p + O\big(|P|^{3/2}N^{-1/2}\big). \tag{40}$$

In particular,

$$E_P(N) - E_0(N) \geq |P| \min_p \sqrt{2\widehat{v}(p) + |p|^2} + O\big(|P|^{3/2}N^{-1/2}\big). \tag{41}$$

The bound (41) implies that $E_P(N) - E_0(N)$ behaves linearly in P for not too large P (assuming that $\widehat{v}(p)$ does not vanish for small $|p|$). Note that this fact is caused by the interactions among the particles, non-interacting systems do not show this behavior. The linear behavior is very important physically and is responsible for the superfluid behavior of the system. According to Landau, the coefficient multiplying $|P|$ in (41) is, in fact, the critical velocity for frictionless flow. We refer to [7] for further details on this correspondence.

Note that under the unitary transformation

$$\widetilde{U} = \exp\left(-iq\cdot\sum_{j=1}^N x_j\right), \quad q \in (2\pi\mathbb{Z})^3, \tag{42}$$

the Hamiltonian H_N transforms as

$$\widetilde{U}^\dagger H_N \widetilde{U} = H_N + N|q|^2 - 2q\cdot P, \tag{43}$$

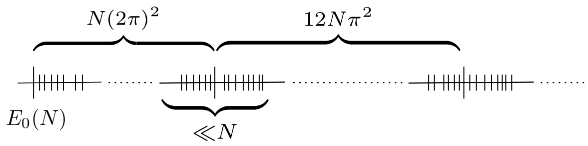


Fig. 1 Sketch of the parts of the spectrum that are correctly determined by the Bogoliubov approximation in the Hartree limit

where $P = -i\sum_{j=1}^{N} \nabla_j$ denotes again the total momentum operator. Hence our results apply equally also to the parts of the spectrum of H_N with excitation energies close to $N|q|^2$, corresponding to *collective excitations* where the particles move uniformly with momentum q; cf. Fig. 1.

6.2 Ideas in the Proof

In the language of second quantization, the Hamiltonian H_N is the restriction of the operator

$$\mathbb{H} = \sum_{p\in(2\pi\mathbb{Z})^3} |p|^2 a_p^\dagger a_p + \frac{1}{2(N-1)} \sum_p \widehat{v}(p) \sum_{q,k} a_{q+p}^\dagger a_{k-p}^\dagger a_k a_q \tag{44}$$

to the N-particle subspace of the Fock space $\mathcal{F}$. Note that N has two different roles here. It determines the particle number, but also appears as a parameter in the Hamiltonian $\mathbb{H}$.

As discussed in Sect. 5, the Bogoliubov approximation consists of

- dropping all terms higher than quadratic in $a_p^\dagger$ and a_p, $p \neq 0$;
- replacing $a_0^\dagger$ and a_0 by $\sqrt{N}$.

The resulting quadratic Hamiltonian is $\frac{N}{2}\widehat{v}(0) + \mathbb{H}^{\mathrm{Bog}}$, where

$$\mathbb{H}^{\mathrm{Bog}} = \sum_{p\neq 0} \left(\left(|p|^2 + \widehat{v}(p)\right) a_p^\dagger a_p + \frac{1}{2}\widehat{v}(p) \left(a_p^\dagger a_{-p}^\dagger + a_p a_{-p} \right) \right). \tag{45}$$

It is diagonalized via a Bogoliubov transformation of the form

$$b_p = \cosh(\alpha_p) a_p + \sinh(\alpha_p) a_{-p}^\dagger, \tag{46}$$

leading to

$$\mathbb{H}^{\mathrm{Bog}} = E_0^{\mathrm{Bog}} + \sum_{p\neq 0} e_p b_p^\dagger b_p \tag{47}$$

for an appropriate choice of α_p, with E_0^{Bog} and e_p defined in (36) and (38), respectively.

The proof of Theorem 1 consists of *two main steps*:

1. As a first step, one shows that H_N is well approximated by an operator similar to the Bogoliubov Hamiltonian $\mathbb{H}^{\mathrm{Bog}}$ in (45), but with a_p and $a_p^\dagger$ replaced by

$$a_p^\dagger \to c_p^\dagger := \frac{a_p^\dagger a_0}{\sqrt{N}}, \qquad a_p \to c_p := \frac{a_p a_0^\dagger}{\sqrt{N}}. \tag{48}$$

Note that the operators c_p and $c_p^\dagger$ conserve the particle number. The resulting Hamiltonian is quadratic in $c_p^\dagger$ and c_p and is, in particular, also particle number conserving. Hence it has a chance of being close to H_N on the subspace of particle number N. The original Bogoliubov Hamiltonian (45) does not leave this subspace invariant, and hence can not be directly compared with H_N.

2. Mimicking the Bogoliubov transformation (46), we introduce the operators $d_p = \cosh(\alpha_p)c_p + \sinh(\alpha_p)c_{-p}^\dagger$. It turns out that the modified Hamiltonian from Step 1 is close to

$$E_0^{\rm Bog} + \sum_{p\neq 0} e_p d_p^\dagger d_p, \tag{49}$$

whose spectrum now has to be analyzed. This analysis is complicated by the fact that the operators d_p and $d_p^\dagger$ do *not* satisfy CCR. It turns out that they do, however, approximately on the subspace where $a_0^\dagger a_0$ is close to N, which is sufficient for our purpose.

In the following, we shall explain these two steps in greater detail. For further details, we refer to [34].

6.2.1 Step 1: Approximation by a Quadratic Hamiltonian

Under our assumptions on the interaction potential v, it is not difficult to see that

$$N - a_0^\dagger a_0 \leq \text{const.}\big[1 + H_N - E_0(N)\big]. \tag{50}$$

This proves that the excitation energy dominates the condensate depletion. In particular, if the excitation energy is small compared with N, most particles occupy the zero momentum mode, i.e., Bose–Einstein condensation occurs.

To show that cubic and quartic terms in $a_p^\dagger$ and a_p, $p \neq 0$, in the Hamiltonian are negligible, one needs to prove a stronger bound of the form

$$\big(N - a_0^\dagger a_0\big)^2 \leq \text{const.}\big[1 + \big(H_N - E_0(N)\big)^2\big], \tag{51}$$

however. It implies that also the fluctuations in the number of particles outside the condensate are suitably small.

The first statement (50) follows easily from positivity of $\widehat{v}(p)$. Positivity implies that

$$\sum_{p\in(2\pi\mathbb{Z})^3\setminus\{0\}} \widehat{v}(p) \left| \sum_{j=1}^{N} e^{ip\cdot x_j} \right|^2 \geq 0, \tag{52}$$

which can be rewritten as

$$\sum_{1\leq i<j\leq N} v(x_i - x_j) \geq \frac{N^2}{2}\widehat{v}(0) - \frac{N}{2}v(0). \tag{53}$$

Thus H_N is bounded from below as

$$H_N \geq \frac{N}{2}\widehat{v}(0) + T - \frac{N}{2(N-1)}\big(v(0) - \widehat{v}(0)\big), \tag{54}$$

where T denotes the kinetic energy

$$T = -\sum_{i=1}^{N} \Delta_i. \tag{55}$$

The statement (50) follows from (54) since $T \geq (2\pi)^2(N - a_0^\dagger a_0)$.

For the second statement (51) one has to work a bit more. It turns out to be useful to actually prove a slightly stronger bound, namely the inequality

$$\big(N - a_0^\dagger a_0\big)T \leq \text{const.}\big[1 + \big(H_N - E_0(N)\big)^2\big]. \tag{56}$$

Since $T \geq (2\pi)^2(N - a_0^\dagger a_0)$ (and the two operators commute), this indeed implies the bound (51).

For the proof of (56), let us introduce the notation

$$N^> = N - a_0^\dagger a_0 = \sum_{i=1}^{N} Q_i \tag{57}$$

for the number of particles outside the condensate, where Q denotes the projection onto the subspace of $L^2([0,1]^3)$ of co-dimension one orthogonal to the constant function. For any bosonic (i.e., permutation-symmetric) wave function Ψ, we can write

$$\begin{aligned}\langle\Psi|N^> T|\Psi\rangle &= N\langle\Psi|Q_1 T|\Psi\rangle \\ &= N\langle\Psi|Q_1 S|\Psi\rangle + \langle\Psi|N^>\big(H_N - E_0(N)\big)|\Psi\rangle,\end{aligned} \tag{58}$$

where

$$\begin{aligned} S &= T - H_N + E_0(N) \\ &= E_0(N) - (N-1)^{-1}\sum_{i<j} v(x_i - x_j). \end{aligned} \tag{59}$$

With the aid of the Cauchy–Schwarz inequality, the last term in (58) can be bounded as

$$\langle\Psi|N^>\big(H_N - E_0(N)\big)|\Psi\rangle \leq \langle\Psi|\big(N^>\big)^2|\Psi\rangle^{1/2}\langle\Psi|\big(H_N - E_0(N)\big)^2|\Psi\rangle^{1/2}. \tag{60}$$

We split S into two parts, $S = S_a + S_b$, with

$$S_a = E_0(N) - \frac{1}{N-1}\sum_{2\leq i<j\leq N} v(x_i - x_j) \tag{61}$$

and

$$S_b = -\frac{1}{N-1}\sum_{j=2}^{N} v(x_1 - x_j). \tag{62}$$

Note that S_a does not depend on x_1. By using positivity of $\widehat{v}(p)$ as in (52), but with the sum over j running from 2 to N only, as well as the simple upper bound $E_0(N) \leq \frac{N}{2}\widehat{v}(0)$ on the ground state energy, we see that

$$S_a \leq \frac{1}{2}\big(\widehat{v}(0) + v(0)\big). \tag{63}$$

In particular, this implies that

$$N\langle\Psi|Q_1 S_a|\Psi\rangle \leq \frac{1}{2}\big(\widehat{v}(0) + v(0)\big)\langle\Psi|N^>|\Psi\rangle. \tag{64}$$

To bound the contribution of S_b, we use

$$\begin{aligned} -\langle\Psi|Q_1 S_b|\Psi\rangle &= \langle\Psi|Q_1 v(x_1 - x_2)|\Psi\rangle \\ &= \langle\Psi|Q_1 Q_2 v(x_1 - x_2)|\Psi\rangle \\ &\quad + \langle\Psi|Q_1 P_2 v(x_1 - x_2) P_2|\Psi\rangle \\ &\quad + \langle\Psi|Q_1 P_2 v(x_1 - x_2) Q_2|\Psi\rangle, \end{aligned} \tag{65}$$

where $P = 1 - Q$ denotes the rank-one projection onto the constant function in $L^2([0,1]^3)$. The second term on the right side of (65) is positive. For the first and the third, we use Schwarz's inequality and $\|v\|_\infty = v(0)$ to conclude that

$$\langle\Psi|Q_1 S_b|\Psi\rangle \leq v(0)\langle\Psi|Q_1 Q_2|\Psi\rangle^{1/2} + v(0)\langle\Psi|Q_1|\Psi\rangle. \tag{66}$$

Since

$$\langle\Psi|Q_1 Q_2|\Psi\rangle = \frac{\langle\Psi|N^>(N^> - 1)|\Psi\rangle}{N(N-1)} \leq \frac{\langle\Psi|(N^>)^2|\Psi\rangle}{N^2}, \tag{67}$$

we have thus shown that

$$\begin{aligned} \langle\Psi|N^> T|\Psi\rangle &\leq \frac{1}{2}\big(\widehat{v}(0) + 3v(0)\big)\langle\Psi|N^>|\Psi\rangle \\ &\quad + \big(v(0) + \langle\Psi|\big(H_N - E_0(N)\big)^2|\Psi\rangle^{1/2}\big)\langle\Psi|\big(N^>\big)^2|\Psi\rangle^{1/2}. \end{aligned} \tag{68}$$

Using that $N^> \leq (2\pi)^{-2}T$ in the last factor, this further implies that

$$\begin{aligned} \langle\Psi|N^> T|\Psi\rangle &\leq \left(\frac{v(0) + \langle\Psi|(H_N - E_0(N))^2|\Psi\rangle^{1/2}}{2\pi}\right)^2 \\ &\quad + \big(3v(0) + \widehat{v}(0)\big)\langle\Psi|N^>|\Psi\rangle. \end{aligned} \tag{69}$$

The desired result (56) then follows from (50).

6.2.2 An Algebraic Identity

The inequalities (50) and (56) allow us to conclude that $\mathbb{H}$ is, at low energy, well approximated by

$$\frac{N}{2}\widehat{v}(0) + \frac{1}{2}\sum_{p\neq 0}\left[A_p\left(c_p^\dagger c_p + c_{-p}^\dagger c_{-p}\right) + B_p\left(c_p^\dagger c_{-p}^\dagger + c_p c_{-p}\right)\right], \tag{70}$$

where $A_p = |p|^2 + \widehat{v}(p)$ and $B_p = \widehat{v}(p)$, and the operators c_p are defined in (48). A simple identity, which does *not* use the CCR, is

$$\begin{aligned} &A_p\left(c_p^\dagger c_p + c_{-p}^\dagger c_{-p}\right) + B_p\left(c_p^\dagger c_{-p}^\dagger + c_p c_{-p}\right) \\ &\quad= \sqrt{A_p^2 - B_p^2}\left(\frac{(c_p^\dagger + \beta_p c_{-p})(c_p + \beta_p c_{-p}^\dagger)}{1-\beta_p^2} + \frac{(c_{-p}^\dagger + \beta_p c_p)(c_{-p} + \beta_p c_p^\dagger)}{1-\beta_p^2}\right) \\ &\qquad - \frac{1}{2}\left(A_p - \sqrt{A_p^2 - B_p^2}\right)\left(\left[c_p, c_p^\dagger\right] + \left[c_{-p}, c_{-p}^\dagger\right]\right), \end{aligned} \tag{71}$$

where

$$\beta_p = \begin{cases} \frac{1}{B_p}\left(A_p - \sqrt{A_p^2 - B_p^2}\right) & \text{if } B_p > 0 \\ 0 & \text{if } B_p = 0. \end{cases} \tag{72}$$

Note that if the operators c_p and $c_p^\dagger$ satisfied CCR, the term in the last line of (71) would be a constant. Its deviation from a constant can be controlled in terms of the condensate depletion, and the inequality (56) can be used to control the error made by simple replacing it by the value it would take in the case of CCR.

Introducing the operators

$$d_p = \frac{c_p + \beta_p c_{-p}^\dagger}{\sqrt{1-\beta_p^2}} \tag{73}$$

and their adjoints, we conclude that $\mathbb{H}$ is, in fact, close to the operator

$$\frac{N}{2}\widehat{v}(0) + E_0^{\mathrm{Bog}} + \sum_{p\neq 0} e_p d_p^\dagger d_p, \tag{74}$$

where we used that

$$E_0^{\mathrm{Bog}} = -\frac{1}{2}\sum_{p\neq 0}\left(A_p - \sqrt{A_p^2 - B_p^2}\right) \tag{75}$$

and

$$e_p = \sqrt{A_p^2 - B_p^2}. \tag{76}$$

6.2.3 Step 2: The Spectrum of $d_p^\dagger d_p$

If the operators d_p and $d_p^\dagger$ satisfied CCR, we could immediately read off the spectrum of the operator in (74), and we would be done. However, without CCR we do not know the spectrum of $d_p^\dagger d_p$. Moreover, the various summands in (74) do not actually commute in our case.

The usual Bogoliubov transformation (46) is of the form

$$b_p = \cosh(\alpha_p) a_p + \sinh(\alpha_p) a_{-p}^\dagger = e^{-X} a_p e^X, \tag{77}$$

where X is the anti-hermitian operator

$$X = \frac{1}{2} \sum_{p\neq 0} \alpha_p \left(a_p^\dagger a_{-p}^\dagger - a_p a_{-p} \right). \tag{78}$$

This identity can easily be verified using the CCR $[a_p, a_q^\dagger] = \delta_{p,q}$. Our operators $c_p = a_p a_0^\dagger / \sqrt{N}$, on the other hand, satisfy

$$\left[c_p, c_q^\dagger \right] = \delta_{p,q} \frac{a_0 a_0^\dagger}{N} - \frac{a_p a_q^\dagger}{N}. \tag{79}$$

We now define, in analogy to (78), the particle-number conserving anti-hermitian operator

$$\widetilde{X} = \frac{1}{2} \sum_{p\neq 0} \alpha_p \left(c_p^\dagger c_{-p}^\dagger - c_p c_{-p} \right). \tag{80}$$

In order to compute the spectrum of $d_p^\dagger d_p$, we apply the unitary $e^{\widetilde{X}}$, and argue that the resulting operator is close $a_p^\dagger a_p$, at least in the subspace of low energy. More precisely, we show that

$$e^{-\widetilde{X}} a_p e^{\widetilde{X}} = \overbrace{\cosh(\alpha_p) c_p + \sinh(\alpha_p) c_{-p}^\dagger}^{d_p} + \mathrm{Error}_p \tag{81}$$

for suitable small error term. Here it is important that actually the sum over all error terms (depending on p) is still (relatively) small as long as $(N - a_0^\dagger a_0)^2 \ll N^2$. The proof of (81) is somewhat lengthy and will be skipped here. It proceeds by studying $e^{-t\widetilde{X}} a_p e^{t\widetilde{X}}$ as a function of $t \in [0, 1]$, using a Grönwall type estimate. The details are presented in [34].

6.3 Conclusions and Generalizations

The mean-field or Hartree limit may be somewhat unphysical when it comes to the description of cold atomic gases. It can be used as a toy model, however, which is analytically much easier to handle than the Gross-Pitaevskii limit of dilute gases [24, 27–29, 35], for instance. The results reviewed in this paper are the first rigorous results concerning the excitation spectrum of an interacting Bose gas, in a suitable limit of weak, long-range interactions. With the notable exception of exactly solvable models in one dimension (like the Lieb–Liniger model [22, 23], for instance), this is the only model where rigorous results on the excitation spectrum are available. The results verify Bogoliubov's prediction that the spectrum consists of sums of elementary excitations. In the translation invariant case, the excitation energy turns out to be linear in the momentum for small momentum. In particular, Landau's criterion for superfluidity is verified.

The methods presented in this paper can be generalized to inhomogeneous systems without translation invariance. This was shown in [16], where the excitation spectrum of the Hamiltonian

$$H_N = \sum_{i=1}^{N}\big(-\Delta_i + V(x_i)\big) + \frac{1}{N-1}\sum_{1\le i<j\le N} v(x_i - x_j) \tag{82}$$

on the Hilbert space $\bigotimes_{\text{sym}}^N L^2(\mathbb{R}^3)$ was studied, with a trap potential V that is locally bounded and tends to infinity at infinity, in order to ensure that all the particles are confined and cannot escape to infinity. Moreover, v is assumed to be non-negative, bounded, and of positive type. To leading order in N, the ground state energy of (82) is determined by minimizing the *Hartree functional*

$$\begin{aligned}\mathcal{E}^{\mathrm{H}}(\phi) = &\int_{\mathbb{R}^3}\big(|\nabla\phi(x)|^2 + V(x)|\phi(x)|^2\big)dx \\ &+ \frac{1}{2}\int_{\mathbb{R}^6} |\phi(x)|^2 v(x-y)|\phi(y)|^2 dx dy,\end{aligned} \tag{83}$$

with minimal energy $E^{\mathrm{H}} = \inf\{\mathcal{E}^{\mathrm{H}}(\phi) : \int|\phi|^2 = 1\}$. Under the stated conditions on v and V, it is not difficult to see that there exists a unique minimizer ϕ_0 (up to a constant phase, of course, which we can choose such that ϕ_0 is positive) with $E^{\mathrm{H}} = \mathcal{E}^{\mathrm{H}}(\phi_0)$. The corresponding Euler-Lagrange equation for the minimizer ϕ_0 can be written as $K^{\mathrm{H}}\phi_0 = 0$, where K^{H} is the Hartree operator

$$K^{\mathrm{H}} = -\Delta + V(x) + v * |\phi_0|^2(x) - \varepsilon_0, \tag{84}$$

with $\varepsilon_0 = E^{\mathrm{H}} + \frac{1}{2}\int_{\mathbb{R}^6} |\phi_0(x)|^2 v(x-y)|\phi_0(y)|^2 dx dy$ and $*$ denoting convolution.

The excitation spectrum of (82) turns out to have a similar structure as in (37), i.e., it consists of sums of elementary excitations. These are described by an effective one-body operator given by

$$E = \left(\sqrt{K^{\mathrm{H}}}\big(K^{\mathrm{H}} + 2W\big)\sqrt{K^{\mathrm{H}}}\right)^{1/2}, \tag{85}$$

where W denotes the operator with integral kernel $\phi_0(x)v(x-y)\phi_0(y)$. More precisely, to leading order in N the spectrum of $H_N - E_0(N)$ is of the form $\sum_i e_i n_i$, with $n_i \in \{0, 1, 2, \dots\}$ and e_i the (non-zero) eigenvalues of E. We refer to [16] for details.

By using different techniques, this result was further generalized in [21], where the validity of the Bogoliubov approximation in the Hartree limit was shown for a much larger class of Hamiltonians and interaction potentials. The method of [21] does not require that v has positive Fourier transform, for instance, one merely needs to assume that the corresponding Hartree functional has a unique minimizer and that its Hessian is strictly positive at the minimum. While the result of [21] applies to a much larger class of models, it does not yield so precise estimates on the error terms as the ones obtained in Theorem 1, and is restricted to studying the excitation spectrum in a smaller window above the ground state energy.

It remains to be seen to what extent the methods in [16, 34] or the method in [21] can be generalized to the study of less restrictive parameter regimes, away from the Hartree limit. A first step in this direction was recently taken in [10], where bounds were given on the maximally allowed rate at which the system size is allowed to grow with N in order for the Bogoliubov approximation to remain valid. Equivalently, one can let the interaction potential v depend on N and ask at what rate it is allowed to tend to a δ-function as $N \to \infty$. Since all error terms in Theorem 1 are explicit, an estimate of this kind is actually contained in Theorem 1, but the dependence of the error terms on v was greatly improved in [10].

Finally, we mention that the validity of the Bogoliubov approximation in the Hartree limit can also be investigated concerning the dynamics generated by the Hamiltonian H_N. We refer to [20] and the references there for recent results in this direction.

7 Open Problems

In this final section, we collect a list of open problems related to the Bogoliubov approximation for many-boson systems. Some of these problems have already been mentioned in the preceding sections.

- One of the key assumptions motivating the Bogoliubov approximation is the existence of *Bose–Einstein condensation*. While this property is easy to demonstrate in the Hartree limit discussed in the previous section, it is not known how to prove it in more general cases. In particular, the existence of BEC in the usual thermodynamic limit ($N \to \infty$, $L \to \infty$ with N/L^3 fixed) remains an open problem. The only model where the occurrence of BEC has been proved in the thermodynamic limit is the hard-core lattice gas at exactly half-filling, which is equivalent to the quantum XY spin model [11, 18]. BEC is also known to occur in the Gross-Pitaevskii limit of dilute trapped gases [24, 28, 29, 35].
- The results in Sect. 6 on the excitation spectrum concern the mean-field or Hartree limit, where the interaction among the particles is very weak and of long range. In fact, the range is of the same order as the system size. In view of applications to cold atomic gases, a physically more relevant limit would be the *Gross-Pitaevskii limit* [24, 28, 29, 35], where the interaction potential takes the form

$$v(x) = N^2 w(Nx) \tag{86}$$

  for some fixed, N-independent function w. As discussed in more detail in [16], one expects that in this limit the excitation spectrum is still of the form (37), but with $\widehat{v}(p)$ replaced by $8\pi a$, where a denotes the scattering length of w.

- An even more challenging problem concerns the low energy excitation spectrum in the *thermodynamic limit*, and to study its relation to the property of *superfluidity*. There are no rigorous results available up to now, not even rough bounds are known. In fact, not even the absence of a *spectral gap* in the thermodynamic limit of an interacting Bose gas is rigorously known. We refer to [7] for further discussion of this topic.
- Also for the Hartree limit discussed in Sect. 6 there are interesting open problems. One of them concerns the existence of *collective excitations* which should be described by solutions of the Hartree equation

$$-\Delta\phi(x) + V(x)|\phi(x)|^2 + v * |\phi|^2(x)\phi(x) = \mu\phi(x) \tag{87}$$

for some $\mu \in \mathbb{R}$, that are different from ϕ_0 and hence correspond to (non-linear) excited states of the Hartree functional. In the translation invariant case, collective excitations are related to the ground state via a Galileo transformation, as explained in Sect. 6.1. In the absence of translation invariance, there is no such symmetry, and the existence of such states is therefore an open problem in general.

Moreover, the results in [10, 16, 21, 34] are all limited to the case where the Hartree functional (83) has a unique minimizer (up to a constant phase). However, at least in the case of attractive interactions, uniqueness will not hold, in general (see, e.g. [2, 17]). Even with repulsive interactions, uniqueness can fail in the presence of magnetic fields or, equivalently, the case of rotating Bose gases [32, 33]. In this case, there can even be uncountably many minimizers. This happens, for instance, in rotating systems if the system is rotation invariant with respect to the axis of rotation, and the rotation speed is large enough for quantized vortices to form. If there is more than one such vortex, the rotation symmetry is necessarily broken in the minimizer, and hence there are infinitely many minimizers, which are all related via rotation. It would be nice to extend the results about the excitation spectrum in the Hartree limit to the case of multiple Hartree minimizers.

References

1. Anderson, M.H., Ensher, J.R., Matthews, M.R., Wieman, C.E., Cornell, E.A.: Observation of Bose–Einstein condensation in a dilute atomic vapor. Science **269**, 198–201 (1995)
2. Aschbacher, W.H., Fröhlich, J., Graf, G.M., Schnee, K., Troyer, M.: Symmetry breaking regime in the nonlinear Hartree equation. J. Math. Phys. **43**, 3879–3891 (2002)
3. Bloch, I., Dalibard, J., Zwerger, W.: Many-body physics with ultracold gases. Rev. Mod. Phys. **80**, 885–964 (2008)
4. Bogolubov, N.: On the theory of superfluidity. USSR J. Phys. **11**, 23–32 (1947)
5. Bose, S.N.: Plancks Gesetz und Lichtquantenhypothese. Z. Phys. **26**, 178–181 (1924)
6. Cooper, N.R.: Rapidly rotating atomic gases. Adv. Phys. **57**, 539–616 (2008)
7. Cornean, H.D., Dereziński, J., Ziń, P.: On the infimum of the energy-momentum spectrum of a homogeneous Bose gas. J. Math. Phys. **50**, 062103 (2009)
8. Dalfovo, F., Giorgini, S., Pitaevskii, L.P., Stringari, S.: Theory of Bose–Einstein condensation in trapped gases. Rev. Mod. Phys. **71**, 463–512 (1999)
9. Davis, K.B., Mewes, M.O., Andrews, M.R., van Druten, N.J., Durfee, D.S., Kurn, D.M., Ketterle, W.: Bose–Einstein condensation in a gas of sodium atoms. Phys. Rev. Lett. **75**, 3969–3973 (1995)

10. Dereziński, J., Napiórkowski, M.: Excitation spectrum of interacting bosons in the mean-field infinite-volume limit. Preprint. arXiv:1305.3641
11. Dyson, F.J., Lieb, E.H., Simon, B.: Phase transitions in quantum spin systems with isotropic and nonisotropic interactions. J. Stat. Phys. **18**, 335–383 (1978)
12. Einstein, A.: Quantentheorie des einatomigen idealen Gases. Sitzber. Preuss. Akad. Wiss., phys.-math. Kl., 261–267 (1924); 3–14 (1925)
13. Fetter, A.L.: Rotating trapped Bose–Einstein condensates. Rev. Mod. Phys. **81**, 647–691 (2009)
14. Ginibre, J.: On the asymptotic exactness of the Bogoliubov approximation for many boson systems. Commun. Math. Phys. **8**, 26–51 (1968)
15. Giuliani, A., Seiringer, R.: The ground state energy of the weakly interacting Bose gas at high density. J. Stat. Phys. **135**, 915–934 (2009)
16. Grech, P., Seiringer, R.: The excitation spectrum for weakly interacting bosons in a trap. Commun. Math. Phys. **322**, 559–591 (2013)
17. Guo, Y., Seiringer, R.: On the mass concentration for Bose–Einstein condensates with attractive interactions. Lett. Math. Phys. **104**, 141–156 (2014)
18. Kennedy, T., Lieb, E.H., Shastry, B.S.: The XY model has long-range order for all spins and all dimensions greater than one. Phys. Rev. Lett. **61**, 2582–2584 (1988)
19. Lewin, M., Nam, P.T., Rougerie, N.: Derivation of Hartree's theory for generic mean-field Bose systems. Preprint. arXiv:1303.0981
20. Lewin, M., Nam, P.T., Schlein, B.: Fluctuations around Hartree states in the mean-field regime. Preprint. arXiv:1307.0665
21. Lewin, M., Nam, P.T., Serfaty, S., Solovej, J.P.: Bogoliubov spectrum of interacting Bose gases. Preprint. arXiv:1211.2778
22. Lieb, E.H.: Exact analysis of an interacting Bose gas. II. The excitation spectrum. Phys. Rev. **130**, 1616–1624 (1963)
23. Lieb, E.H., Liniger, W.: Exact analysis of an interacting Bose gas. I. The general solution and the ground state. Phys. Rev. **130**, 1605–1616 (1963)
24. Lieb, E.H., Seiringer, R.: Proof of Bose–Einstein condensation for dilute trapped gases. Phys. Rev. Lett. **88**, 170409 (2002)
25. Lieb, E.H., Solovej, J.P.: Ground state energy of the one-component charged Bose gas. Commun. Math. Phys. **217**, 127–163 (2001). See also: Errata **225**, 219–221 (2002)
26. Lieb, E.H., Solovej, J.P.: Ground state energy of the two-component charged Bose gas. Commun. Math. Phys. **252**, 485–534 (2004)
27. Lieb, E.H., Yngvason, J.: Ground state energy of the low density Bose gas. Phys. Rev. Lett. **80**, 2504–2507 (1998)
28. Lieb, E.H., Seiringer, R., Yngvason, J.: Bosons in a trap: a rigorous derivation of the Gross-Pitaevskii energy functional. Phys. Rev. A **61**, 043602 (2000)
29. Lieb, E.H., Seiringer, R., Solovej, J.P., Yngvason, J.: The Mathematics of the Bose Gas and Its Condensation, Oberwolfach Seminars Bd. 34. Birkhäuser, Basel (2005). arXiv:cond-mat/0610117
30. Lieb, E.H., Seiringer, R., Yngvason, J.: Justification of c-number substitutions in bosonic Hamiltonians. Phys. Rev. Lett. **94**, 080401 (2005)
31. Schlein, B.: Effective equations for quantum dynamics. Preprint. arXiv:1208.0185
32. Seiringer, R.: Gross-Pitaevskii theory of the rotating Bose gas. Commun. Math. Phys. **229**, 491–509 (2002)
33. Seiringer, R.: Ground state asymptotics of a dilute, rotating gas. J. Phys. A, Math. Gen. **36**, 9755–9778 (2003)
34. Seiringer, R.: The excitation spectrum for weakly interacting bosons. Commun. Math. Phys. **306**, 565–578 (2011)
35. Seiringer, R.: Hot topics in cold gases—a mathematical physics perspective. Jpn. J. Math. **8**, 185–232 (2013)
36. Solovej, J.P.: Upper bounds to the ground state energies of the one- and two-component charged Bose gases. Commun. Math. Phys. **266**, 797–818 (2006)
37. Sütő, A.: Equivalence of Bose–Einstein condensation and symmetry breaking. Phys. Rev. Lett. **94**, 080402 (2005)
38. Yau, H.-T., Yin, J.: The second order upper bound for the ground energy of a Bose gas. J. Stat. Phys. **136**, 453–503 (2009)
39. Zagrebnov, V.A., Bru, J.-B.: The Bogoliubov model of weakly imperfect Bose gas. Phys. Rep. **350**, 291–434 (2001)

Robert Seiringer obtained his Ph.D. in 2000 at the University of Vienna under the supervision of Jakob Yngvason. Then he went to Princeton, first as a Post-Doc (2001–2002) and then as Assistant Professor (2003–2010). In 2010 he became Associate Professor at McGill University in Montréal. Since last spring he is Professor at the Institute for Science and Technology Austria. Robert Seiringer's research focuses on quantum many-body systems. He was awarded several prizes and awards including the Henri Poincaré Prize of the International Association of Mathematical Physics, the NSERC E.W.R. Steacie Memorial Fellowship, and the William Dawson Scholarship.

Jahresber Dtsch Math-Ver (2014) 116:43–65
DOI 10.1365/s13291-014-0085-7

SURVEY ARTICLE

Self-similarity in Smoluchowski's Coagulation Equation

Barbara Niethammer

Published online: 6 March 2014

Abstract Smoluchowski's coagulation equation is one of the fundamental deterministic models that describe mass aggregation phenomena. A key question in the analysis of this equation is whether the large-time behaviour of solutions is universal and described by special self-similar solutions. This issue is however only well-understood for the small class of solvable kernels while the analysis for non-solvable kernels still poses many challenging problems.

Our main focus in this article will be to describe recent progress in the analysis of self-similar solutions of Smoluchowski's equation for non-solvable kernels. Existence results for self-similar solutions with finite mass have been available for some time for a large class of kernels. In contrast, the uniqueness of such solutions has been an open problem for some time. A first uniqueness result has recently been obtained for kernels that are in a certain sense close to constant. We present here a shorter proof under an additional assumption on the kernels that makes the analysis significantly simpler. We also give an overview of recent results on the existence of fat tail solutions for non-solvable kernels.

Keywords Smoluchowski's coagulation equations · Self-similar solutions · Uniqueness

1 Introduction

In 1916 Smoluchowski [33] derived a mean-field model for mass aggregation in order to develop a mathematical theory for coagulation processes. His interest in this topic arose from a correspondence with R. Zsigmondys who investigated the coagulation

B. Niethammer (✉)
Institute of Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany
e-mail: niethammer@iam.uni-bonn.de

of particles in a homogeneous colloidal gold solution and Smoluchowski also used the results of these experiments to validate his theory.

Since Smoluchowski's groundbreaking work his model has been used in a surprisingly diverse range of applications. One of the largest application area are aerosols, where one is interested in describing the statistical behaviour of a large number of liquid or solid particles suspended in a gas. The goal is to predict the time evolution of measurable macroscopic characteristics of the aerosol cloud, such as mean sizes or weights or higher order moments of the distribution function. Similar questions have been investigate in the kinetics of polymerization processes, or on very large scales, such as in the description of clustering of planets and stars [32].

In Smoluchowski's model particles are solely characterized by their size $i \in \mathbb{N}$, the number of so-called atoms or monomers in the cluster. The functions $n_i = n_i(t)$ denote the number densities of clusters of size i at time t. The crucial assumptions in the model are that one only needs to take binary coagulation into account, while coagulation of three or more particles can be neglected, and that the rate of coagulation of clusters of size i and j is given by a rate kernel $K(i, j)$ which is nonnegative and symmetric. Then the rate of change of n_i is given by the system of rate equations

$$\frac{d}{dt}n_i(t) = \frac{1}{2}\sum_{j=1}^{i-1} K(i-j, j)n_{i-j}(t)n_j(t) - n_i(t)\sum_{j=1}^{\infty} K(i, j)n_j(t), \quad i \in \mathbb{N}. \tag{1}$$

The properties of solutions to this equation depend of course on the properties of the kernel K in which all microscopic details of the coagulation process are subsumed. For example, Smoluchowski derived in [33] the kernel

$$K(i, j) = D\big(i^{1/3} + j^{1/3}\big)\big(i^{-1/3} + j^{-1/3}\big) \tag{2}$$

under the assumption that the clusters move independently by Brownian motion and coagulate quickly when they become close. Here $i^{1/3} + j^{1/3}$ is proportional to the sum of the diameters of two clusters, which are assumed to be approximately balls, while $i^{-1/3} + j^{-1/3}$ is, according to Einstein's formula, proportional to the diffusion constant of the clusters.

While Smoluchowski was not able to say something about solutions to (1) with K as in (2), he already noticed that if one chooses K to be constant, one can solve (1) explicitly. Indeed, summing (1) over i, one obtains for $N(t) := \sum_{i=1}^{\infty} n_i(t)$ the differential equation $\dot{N}(t) = -\frac{1}{2}N^2(t)$ which gives $N(t) = \frac{1}{t+\frac{1}{N(0)}}$. This formula can be plugged into the right hand side of (1) and allows to solve the equations successively for $i = 2, 3, \ldots$. Similar procedures are possible for the kernels $K(i, j) = i + j$ and $K(i, j) = ij$ and for this reason these three kernels are called the solvable kernels.

However, most kernels that are derived for specific applications, do not have the form of one of the solvable kernels. The most prominent one is Smoluchowski's kernel (2), but one can find in addition a huge variety of kernels in different applications, that are non-solvable. For example, in [24] the kernel

$$K(i, j) = a\big(i^{1/3} + j^{1/3}\big)^2\left(\frac{1}{i} + \frac{1}{j}\right)^{1/2}$$

and generalization of it are used to describe soot agglomeration or [32] use

$$K(i,j) = \left(i^{1/3} + j^{1/3}\right)^2$$

to describe coalescence of stellar fragments. We refer to [1, 6, 16] for further examples of coagulation kernels.

Smoluchowski did not take the possible fragmentation of clusters into smaller ones into account. There are obvious extensions of (1) that include this effect, however we will focus in this article solely on the issues that appear in the pure coagulation process. We also do not discuss other related models that take additional effects, such as space dependence or higher order coagulation into account.

In a pure coagulation process the average cluster size is always increasing. In order to describe the process on the relevant scales it is therefore appropriate to consider the continuous variant of (1). Then $\xi \in (0,\infty)$ describes the size or mass of cluster and if $n(\xi,t)$ is the number density of clusters of size ξ at time t then (1) turns into

$$\partial_t n(\xi,t) = \frac{1}{2}\int_0^\xi K(\xi,\eta)n(\xi-\eta,t)n(\eta,t)d\eta - n(\xi,t)\int_0^\infty K(\xi,\eta)n(\eta,t)d\eta. \quad (3)$$

Another way of viewing this is the fact that (3) turns into (1) if $n(\xi,t) = \sum_{i=1}^\infty n_i(t)\delta_{(\xi=i)}$. In the following our main focus will be on the long-time behaviour of solutions to Smoluchowski's equation and from now on we will always consider the continuous version of Eq. (3).

2 Well-Posedeness and Mass Conservation

A fundamental feature of (3) is mass conservation. Since mass is conserved on the microscopic level one would expect that it is also conserved on the level of the macroscopic equation. In fact, multiplying (3) by a test function ϕ one obtains formally that

$$\begin{aligned}
&\frac{d}{dt}\int_0^\infty \phi(\xi)n(\xi,t)d\xi \\
&= \int_0^\infty \phi(\xi)\frac{1}{2}\int_0^\xi K(\xi-\eta,\eta)n(\xi-\eta,t)n(\eta,t)d\eta d\xi \\
&\quad - \int_0^\infty\int_0^\infty \phi(\xi)K(\xi,\eta)n(\xi,t)n(\eta,t)d\eta d\xi \\
&= \frac{1}{2}\int_0^\infty\int_\eta^\infty \phi(\xi)K(\xi-\eta,\eta)n(\xi-\eta,t)n(\eta,t)d\xi d\eta \\
&\quad - \int_0^\infty\int_0^\infty \phi(\xi)K(\xi,\eta)n(\xi,t)n(\eta,t)d\eta d\xi \\
&= \frac{1}{2}\int_0^\infty\int_0^\infty \big(\phi(\xi+\eta)-\phi(\xi)-\phi(\eta)\big)K(\xi,\eta)n(\xi,t)n(\eta,t)d\xi d\eta. \quad (4)
\end{aligned}$$

Thus, choosing $\phi(\xi) = \xi$, we obtain on the formal level that $\frac{d}{dt}\int_0^\infty \xi n(\xi,t)d\xi = 0$. However, it is not clear, whether the integrals on the right hand side of (4) all exist. And in fact, it is by now well-known that mass-conservation fails to hold for all times if the kernel K grows faster than linearly at infinity. For the kernel $K(\xi,\eta) = \xi\eta$ this can be seen easily. Indeed, if one assumes that a nonnegative solution $n(\xi,t)$ exists, such that $\int_0^\infty \xi n(\xi,t)d\xi = 1$ say, then we obtain for the total number of clusters $N(t) = \int_0^\infty n(\xi,t)d\xi$ from (4)

$$\frac{d}{dt}N(t) = -\frac{1}{2}\left(\int_0^\infty \xi n(\xi,t)d\xi\right)^2 = -\frac{1}{2}.$$

Hence, N becomes negative after a finite time, which contradicts the existence of a mass conserving non-negative solution.

This phenomenon, that is the loss of mass at finite time, is known as gelation and is explained by the creation of infinitely large clusters at the time when the loss of mass starts, the so-called gelation time. In a first instance one would think of this kind of singularity in the mathematical model as unphysical, but is has its direct correspondence to physically relevant processes such as the formation of gels, that is large polymer networks, in polymerization processes.

For the discrete equation with $K(i,j) = ij$ an explicit solution has been given in [21] in terms of power series from which one can read of the blow-up of higher moments as the time approaches the gelation time. While the corresponding formulas fail to give a solution after the gelation time, since they are derived under the assumption of conserved mass, an extension of the solution has been provided in [20]. In subsequent work it has also been rigorously established by indirect arguments that for a large class of kernels that grow faster than linearly at infinity, gelation must necessarily occur at a finite time [8, 13].

In contrast, for kernels that grow at most linearly at infinity one can expect that given initial data with finite mass, there exists a unique global solution to (3) that conserves the mass for all times. This has in fact also been proved rigorously for a large range of kernels, see e.g. [2, 15, 16, 31]. In Norris [31], however, an example of a kernel is given for which there exist different mass-conserving solutions. Laurençot and Fournier [12] also proved the existence and uniqueness of solutions to (3) for a range of kernels with homogeneity $\gamma < 1$ under the rather weak assumption that the initial data have finite γ-th moment.

3 Self-similarity

We now turn to the main focus of this review article that is the so-called scaling hypothesis. It has its roots in the observation that in many experiments a certain mean or 'typical' size evolves after some transient time according to some simple power law in time and that most other macroscopic quantities of interest can be computed using this quantity.

In mathematical terms, the scaling hypothesis states that for homogeneous kernels, that is for kernels satisfying $K(ax,ay) = a^\gamma K(x,y)$ for some $\gamma \in \mathbb{R}$, solutions approach a unique self-similar profile for large times. Despite a significant range of results based on formal asymptotics (see in particular [18, 34]) mathematically rigorous results supporting this hypothesis are still rare except for the special case of solvable kernels.

In the following we will exclusively consider the case of homogeneity $\gamma < 1$. In the case $\gamma > 1$, at least to our knowledge apart from the solvable kernel $K(\xi,\eta) = \xi\eta$, absolutely nothing is known about the convergence to self-similar form when t approaches the gelation time. The case $\gamma = 1$ is special and also only understood for the solvable kernel $K(\xi,\eta) = \xi + \eta$. In this case the rescaling that is suggested by mass conservation leads to a self-similar solution that does not have finite mass. Instead one needs to consider a different scaling regime to obtain a meaningful limit. We recommend to the interested reader the extensive review article [18] that discusses all these aspects in detail.

Hence, from now on we will assume that the kernel K has homogeneity $\gamma < 1$. It is easily checked that if $n(\xi,t)$ is a solution to (3), then so is $m(\xi,t) = a^{\gamma+1}b^{-1}n(a\xi,bt)$ for arbitrary positive constants a,b. This scale invariance leads one to expect that there is a self-similar solution to (3), that is a special solution of the form

$$n(\xi,t) = \frac{1}{t^{\alpha}} f(x), \qquad x = \frac{\xi}{t^{\beta}}. \tag{5}$$

with $\alpha = 1 + (1+\gamma)\beta$ and with a so-called self-similar profile f that solves

$$\begin{aligned} -\big(1 + (1+\gamma)\beta\big) f - \beta x f'(x) = \frac{1}{2}\int_0^x K(x,y) f(x-y,t) f(y,t)\,dy \\ - f(x,t)\int_0^\infty K(x,y) f(y,t)\,dy. \end{aligned} \tag{6}$$

Here we encounter the first problem. A self-consistent scaling argument gives that for some kernels, e.g. $K(x,y) = (xy)^{\gamma/2}$, $\gamma \in (0,1)$, that $f(x) \sim x^{-(1+\gamma)}$ as $x \to 0$. Hence, the integrals on the right-hand side of (6) are not finite. For the equation to make sense we need to go over to a weaker formulation and for that it is convenient to rewrite the equation in the following way. Multiplying equation (6) by x and changing the order of integration on the right hand side we obtain

$$\beta\partial_x\big(x^2 f(x)\big) = \partial_x\left[\int_0^x \int_{x-y}^\infty y K(y,z) f(z) f(y)\,dz\,dy\right] + \big((1-\gamma)\beta - 1\big) x f(x). \tag{7}$$

Under the assumption that $\lim_{x\to 0} x^2 f(x) = 0$ one can integrate (7) to find that f satisfies

$$\beta x^2 f(x) = \int_0^x \int_{x-y}^\infty K(y,z) y f(y) f(z)\,dz\,dy + \big((1-\gamma)\beta - 1\big)\int_0^x y f(y)\,dy \tag{8}$$

for almost all $x > 0$.

In our computations the parameter β is still free. If we require in addition that our self-similar solution has conserved finite first moment, this fixes $\beta = 1/(1-\gamma)$ and

$$\int_0^\infty xf(x)dx = M. \tag{9}$$

Note also that in this case the second term on the right hand side of (7) vanishes. Otherwise, one might obtain, as we will see below, for a certain range of parameters β a solution f with power law decay at infinity that has infinite first moment.

3.1 The Constant Kernel

Let us now first describe what is known about self-similar solutions and the scaling hypothesis for the constant kernel which is the only solvable one with homogeneity smaller than 1. In this case one has an explicit self-similar solution with finite mass, given by the self-similar profile $f(x) = e^{-x}$. Convergence to this self-similar solution has been established under some assumptions on the initial data in several papers [4, 5, 14, 17]. A complete characterization of its domain of attraction has more recently been given in [23]. Moreover, it is also proved in [23] that there exists a family of self-similar solutions with infinite mass, so-called self-similar solutions with fat tails. More precisely, it was established that for any $\rho \in (0,1)$ there exists a self-similar profile with decay $x^{-(1+\rho)}$. Furthermore it is shown that a solution to the coagulation equation converges to the self-similar solution with decay behaviour $x^{-(1+\rho)}$ if and only if the integrated mass distribution is regularly varying with exponent $1-\rho$. The proof is simple and elegant, but relies on the use of the Laplace transform and the methods are not, at least not directly, applicable to non-solvable kernels.

3.2 Non-solvable Kernels: The Case of Finite Mass

3.2.1 Existence of Self-similar Profiles

The situation for non-solvable kernels is very different from the one described in the previous section. The analysis of self-similar solutions in this case started by considerations by van Dongen and Ernst [34]. Under the assumption that a self-similar solution with finite mass exists, they derived for different classes of kernels detailed results on the behaviour of self-similar profiles for $x \to 0$ and $x \to \infty$ respectively. However, one also finds in their paper the remark, that in principle it is possible that they investigate the empty set since the existence of such self-similar solutions is far from obvious.

The existence of self-similar solutions was first rigorously established in [9, 10] for a large class of kernels. The strategy in [10] is to discretize equation (7) (for $\beta = 1/(1-\gamma)$) in a suitable way to obtain a finite dimensional system of equations. Then they consider a corresponding time-dependent problem and show that the flow leaves the cone of nonnegative functions with given first moment invariant. Standard results from ordinary differential equations guarantee the existence of a nonnegative solution of the discretized problem with given first moment. The main task then is to obtain moment estimates and equiintegrability estimates that are sufficient to ensure that the set of solutions to the discretized problem is pre-compact. One can then extract a subsequence that converges weakly in L^1 to a nonnegative weak solution of (7) with given first moment. To obtain further regularity of such a profile is not straightforward since due to possible singularities near zero, one cannot use a usual bootstrap argument in (8). Nevertheless, for certain kernels it has later been established that any self-similar profile is indeed locally smooth [3, 11].

Having proved existence of self-similar profiles one would like to show that they are also unique (up to rescaling). However, to derive such a result turns out to be rather difficult. We will describe below in Sect. 3.2.2 a recent uniqueness result that has been obtained for kernels that are close to constant. Apart from this, the only result in this direction has been a conditional uniqueness result [3]. It is proved for kernels of the form $K(x,y) = x^{-a}y^b + x^b y^{-a}$ with $a \geq 0$ that two self-similar profiles are equal if they have the same moments of order a and b (if $a > 0$) or of order b if $a = 0$. However, it seems to be equally difficult to show that two solutions have the same moments. Nevertheless, if this is known then one can transform the uniqueness problem into a type of initial value problem, starting at $x = 0$ and then use the usual contraction principle to show uniqueness. In order to use this strategy more generally, several results have been obtained that prove rigorously a certain behaviour of self-similar solutions as $x \to 0$ [3, 7, 11, 22, 25]. The details of these results depend very sensitively on the behaviour of the kernel for small x, y. We do not want to give a complete overview of all cases, but rather illustrate the difference between different kernels by considering the two cases $K_1(x,y) = (xy)^{\gamma/2}$ and $K_2(x,y) = x^\gamma + y^\gamma$ with $\gamma \in (0,1)$. Assuming that a self-similar profile satisfies $f(x) \sim x^{-\alpha}$ as $x \to 0$, one easily finds for K_1 that $f(x) \sim c_0(\gamma)x^{-(1+\gamma)}$ with $c_0(\gamma) = \frac{\gamma}{2}(\int_0^1 s^{\gamma/2}(1-s)^{\gamma/2}ds)^{-1}$. Interestingly in [34] the authors remark that for such type of kernels they were not able to calculate on a formal basis the next order behaviour. Formally one finds that the next order behaviour is oscillatory and a solution of such a form has been constructed in [22], but a rigorous proof that every solution behaves in this way is presently not available. In comparison to the 'diagonal dominant' kernel K_1, self-similar profiles for K_2 behave more regularly. It has been made rigorous in [11] that any self-similar profile behaves as $f(x) = x^{-(2-M_\gamma)} + O(x^{-2(2-M_\gamma)})$, where M_γ is the γ-th moment of the profile. The value of M_γ is not known, but estimates are given in [11] and in fact $2 - M_\gamma < 1 + \gamma$, so that the profile is not as singular as in the case K_1.

If the kernel is singular for small x, such as for example Smoluchowski's kernel, the behaviour of self-similar profiles is very different. In fact, if $K(x,y) = x^{-a}y^b + y^{-a}x^b$ with $a > 0$, and $b > -a$, then the profiles behaves as $\frac{1}{x^2}e^{-c\frac{1}{x^a}}$ as $x \to 0$ (see e.g. [3]) where the constant c is related to the moment of order b.

However, despite all these results on the characterizations of self-similar profiles for small x, the derivation of a uniqueness result has been to no avail. In the next section we present a first uniqueness result for kernels that are close to constant. It turns out that the basis of this result is rather a detailed analysis of the decay behaviour of the profiles as $x \to \infty$.

3.2.2 Uniqueness

To our knowledge the following theorem establishes the first uniqueness result for self-similar profiles for any non-solvable kernel.

Theorem 3.1 [29] *Assume that the kernel K is homogeneous of degree zero and satisfies*

$$-\varepsilon \le K(x,y) - 2 \le \varepsilon\left(\left(\frac{x}{y}\right)^a + \left(\frac{y}{x}\right)^a\right) \quad \text{for all } x, y > 0 \tag{10}$$

and some $a \in [0,1)$, *as well as*

$$\left|\frac{\partial}{\partial x} K(x,y)\right| \le \frac{C\varepsilon}{x}\left(\left(\frac{x}{y}\right)^a + \left(\frac{y}{x}\right)^a\right) \quad \text{for all } x, y > 0. \tag{11}$$

Then, if ε *is sufficiently small, there exists at most one continuous self-similar profile, that is* $f \in L^1_{loc}(0,\infty) \cap C((0,\infty))$, $f \ge 0$ *and* f *satisfies* (9) *and* (8) *for all* $x > 0$.

The proof of this theorem is elementary, but still quite technical. It is significantly simpler in the case $a = 0$ and in the following we will present some details of the proof under this simplifying assumption, that is

$$\left|K(x,y) - 2\right| \le \varepsilon \quad \text{for all } x, y > 0. \tag{12}$$

A key idea in the proof is to consider the difference of two profiles in a suitable weak topology. More precisely, even though we do not deal with solvable kernels, we will as in [23] make use of the desingularized Laplace transform of f, given by

$$Q(q) = \int_0^\infty \left(1 - e^{-qx}\right) f(x)\,dx. \tag{13}$$

The function Q is defined for all $q \ge 0$ and due to (9) we have $Q(0) = 0$. Normalizing the mass $M = 1$ also implies that $Q'(0) = 1$. We will see later, see Lemma 3.4, that the function Q is defined on $(-\delta, \infty)$ for some $\delta > 0$. For the following we define

$$\mathcal{M}(f,f)(q) = \frac{1}{2}\int_0^\infty \int_0^\infty W(x,y) f(x) f(y)\left(1 - e^{-qx}\right)\left(1 - e^{-qy}\right) dx\,dy \tag{14}$$

with $W(x,y) := K(x,y) - 2$. To obtain further estimates we derive a differential equation for Q.

Lemma 3.2 *The function* Q *satisfies for all* q *with* $Q(q) < \infty$ *that*

$$-q Q'(q) = Q^2 - Q + \mathcal{M}(f,f)(q). \tag{15}$$

Proof Multiplying (8) by e^{-qx} and integrating we find, after changing the order of integration,

$$\begin{aligned} -Q''(q) &= \int_0^\infty x^2 f(x) e^{-qx} dx \\ &= \int_0^\infty \int_0^\infty K(y,z) y f(y) f(z) \int_y^{y+z} e^{-qx} dx dy dz \\ &= \int_0^\infty \int_0^\infty K(y,z) y f(y) f(z) \frac{1}{q} e^{-qy} \left(1 - e^{-qz}\right) dy dz \\ &= \frac{2}{q} Q'(q) Q(q) + \frac{1}{q} \mathcal{M}(f,f)'(q) \end{aligned}$$

and as a consequence we find

$$-\left(qQ'\right)' = \left(Q^2\right)' - Q' + \left(\mathcal{M}(f,f)\right)'.$$

By definition, we have $Q(0) = 0$ and $\mathcal{M}(f,f)(0) = 0$. Hence, integrating the previous identity we deduce the claim. □

In the following we denote by $\bar{Q}$ the desingularized Laplace transform for the case $K = 2$, that is

$$\bar{Q}(q) = \int_0^\infty e^{-x} \left(1 - e^{-qx}\right) dx = 1 - \frac{1}{1+q} = \frac{q}{1+q}. \tag{16}$$

In the following Lemma we derive some a-priori estimates for Q and $\mathcal{M}$ that are essential for our analysis and follow rather easily from the lower bound on K.

Lemma 3.3 *If $K(x,y) \geq c_0 > 0$ for all $x, y > 0$, then the following estimates hold.*

$$\lim_{q\to\infty} Q(q) < \infty \quad \textit{and hence} \int_0^\infty f(x) dx < \infty, \tag{17}$$

$$\sup_{q>0} \left|qQ'(q)\right| \leq C, \tag{18}$$

$$\int_0^\infty \int_0^\infty K(x,y) f(x) f(y) dx dy < \infty, \tag{19}$$

$$\lim_{q\to\infty} \mathcal{M}(f,f)(q) < \infty. \tag{20}$$

Proof With the assumption on K we can deduce from (15), written with K instead of W, that

$$\begin{aligned} -qQ'(q) &= -Q + \int_0^\infty \int_0^\infty K(x,y) f(x) f(y) \left(1 - e^{-qx}\right)\left(1 - e^{-qy}\right) dx dy \\ &\geq -Q + c_0 Q^2. \end{aligned} \tag{21}$$

Hence, by comparing with the solution of the corresponding ODE, the function Q is uniformly bounded. Since Q is increasing, statement (17) follows.

Next, we have

$$Q'(q) = \frac{1}{q} \int_0^\infty xqe^{-xq} f(x)dx \le \frac{C}{q} \int f(x)dx,$$

which together with (17) establishes (18).

Then it follows from (21) that

$$\int_0^\infty \int_0^\infty K(x,y) f(x) f(y) \big(1 - e^{-qx}\big)\big(1 - e^{-qy}\big) dxdy \le C$$

and by monotone convergence we find (19) in the limit $q \to \infty$. Denoting this limit by J we finally get that

$$\begin{aligned}\mathcal{M}(f,f)(q) &= \int_0^\infty \int_0^\infty W(x,y) f(x) f(y) \big(1 - e^{-qx}\big)\big(1 - e^{-qy}\big) dxdy \\ &= \int_0^\infty \int_0^\infty K(x,y) f(x) f(y) \big(1 - e^{-qx}\big)\big(1 - e^{-qy}\big) dxdy - Q(q)^2 \\ &\to J - Q(\infty)^2\end{aligned}$$

which proves (20). □

A key result for our analysis is the following decay estimate. If f is a solution of (8) and (9) then it decays exponentially fast. This fact is proved in more general situations in [27].

Lemma 3.4 *There exist constants $C, \alpha > 0$ such that any solution of (8), (9) satisfies*

$$f(x) \le Ce^{-\alpha x} \quad \textit{for all } x \ge 1.$$

Furthermore, let $f(x)$ be a solution to (8) and (9) such that $\int_0^\infty f(x)e^{\alpha x}dx < \infty$ for some $\alpha > 0$. Then there exists $\beta > 0$ such that $f(x)e^{\alpha x} \le Ce^{-\beta x}$ for all $x \ge 1$.

Proof We give here a proof in the simpler case (12). More precisely, we only use that K is uniformly bounded.

In the following we denote for $\gamma \ge 1$

$$M(\gamma) := \int_0^\infty x^\gamma f(x)dx. \tag{22}$$

Our goal is to show inductively that

$$M(\gamma) \le \gamma^\gamma e^{A\gamma} \tag{23}$$

for some (large) constant A.

To that aim we first multiply (8) by $x^{\gamma-2}$ with some $\gamma>1$ and after integrating we obtain

$$(\gamma-1)M(\gamma)=\frac{1}{2}\int_0^\infty\int_0^\infty K(x,y)f(x)f(y)\big((x+y)^\gamma-x^\gamma-y^\gamma\big)dxdy.$$

By symmetry we also find

$$\begin{aligned}M(\gamma)&=\frac{1}{\gamma-1}\int_0^\infty\int_0^x K(x,y)f(x)f(y)\big((x+y)^\gamma-x^\gamma\big)dydx\\&=\int_0^1 dx\int_0^x dy\cdots+\int_1^\infty dx\int_0^{x/\gamma}dy\cdots+\int_1^\infty dx\int_{x/\gamma}^x dy\cdots.\end{aligned}$$

Due to (9) we have

$$\begin{aligned}&\int_0^1\int_0^x K(x,y)f(x)f(y)\big((x+y)^\gamma-x^\gamma\big)dydx\\&\quad\le C\int_0^1\int_0^x K(x,y)f(x)f(y)x^\gamma dydx\\&\quad\le C\int_0^1 x^\gamma f(x)\int_0^x yf(y)dydx\le C.\end{aligned}\tag{24}$$

Using (9) and $(x+y)^\gamma-x^\gamma\le cx^{\gamma-1}y$ for $y\le\frac{x}{\gamma}$, we find that

$$\begin{aligned}\int_1^\infty\int_0^{x/\gamma}K(x,y)yx^{\gamma-1}f(x)f(y)dydx&\le C\int_1^\infty\int_0^{x/\gamma}x^{\gamma-1}yf(x)f(y)dydx\\&\le CM(\gamma-1),\end{aligned}\tag{25}$$

so that for the sum of both terms we can prove by induction that it is smaller than $1/2\gamma^\gamma e^{A\gamma}$. It remains to estimate

$$\begin{aligned}&\frac{C}{\gamma-1}\int_1^\infty dx\int_{x/\gamma}^x dyK(x,y)f(x)f(y)(x+y)^\gamma\\&\quad\le\frac{C}{\gamma}\int_1^\infty dx\int_{x/\gamma}^x dyf(x)f(y)(x+y)^\gamma:=(*).\end{aligned}$$

In the following $\{\zeta_n\}\subset(0,1]$, $\zeta_0=1$, will be a decreasing sequence of numbers that will be specified later. Then we define a corresponding sequence of numbers κ_n such that given a sequence $\{\theta_n\}\subset(0,1)$, also to be specified later, we have

$$(x+y)^\gamma\le\kappa_n^\gamma x^{\gamma(1-\theta_n)}y^{\gamma\theta_n}\quad\text{for }\frac{y}{x}\in[\zeta_{n+1},\zeta_n].\tag{26}$$

Equivalently we have

$$\kappa_n=\max_{\zeta\in[\zeta_{n+1},\zeta_n]}\left(\frac{1+\zeta}{\zeta^{\theta_n}}\right).\tag{27}$$

With these definitions we have

$$(*) \leq \frac{C}{\gamma} \sum_{n=0}^{n_0(\gamma)} \kappa_n^{\gamma} M\big(\gamma(1-\theta_n)\big) M(\gamma\theta_n),$$

where $n_0(\gamma)$ is such that $\zeta_{n_0(\gamma)} = \frac{1}{\gamma}$.

We now choose θ_n such that for $\psi_{\theta_n}(\zeta) := \log(1+\zeta) - \theta_n \log\zeta$ we have

$$\min_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta) = \log(1+\zeta_n) - \theta_n \log(\zeta_n).$$

This is equivalent to

$$\theta_n = \frac{\zeta_n}{1+\zeta_n}. \tag{28}$$

We want to prove now by induction over γ that $(*) \leq \frac{1}{2}\gamma^{\gamma} e^{A\gamma}$. Inserting the corresponding hypothesis, this reduces to showing that

$$\frac{C}{\gamma} \sum_{n=0}^{n_0} \exp\Big(\gamma\Big(\max_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta) + \theta_n \log\theta_n + (1-\theta_n)\log(1-\theta_n)\Big)\Big) \leq \frac{1}{2}.$$

By definition (28) we have

$$\begin{aligned}
&\max_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta) + \theta_n \log\theta_n + (1-\theta_n)\log(1-\theta_n) \\
&\quad = \min_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta) + (\max - \min)_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta) + \theta_n \log\theta_n \\
&\qquad + (1-\theta_n)\log(1-\theta_n) \\
&\quad = (\max - \min)_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta).
\end{aligned}$$

Thus we need to investigate

$$\begin{aligned}
(\max - \min)_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta) &= \psi_{\theta_n}(\zeta_{n+1}) - \psi_{\theta_n}(\zeta_n) \\
&= \log\left(\frac{1+\zeta_{n+1}}{1+\zeta_n}\right) - \theta_n \log\left(\frac{\zeta_{n+1}}{\zeta_n}\right) \\
&= \log\left(1 + \frac{\zeta_{n+1}-\zeta_n}{1+\zeta_n}\right) - \frac{\zeta_n}{1+\zeta_n} \log\left(1 + \frac{\zeta_{n+1}-\zeta_n}{\zeta_n}\right).
\end{aligned}$$

Expanding the nonlinear terms we find

$$W := (\max - \min)_{\zeta\in[\zeta_{n+1},\zeta_n]} \psi_{\theta_n}(\zeta) \leq C\left(|\zeta_{n+1}-\zeta_n|^2 + \frac{(\zeta_{n+1}-\zeta_n)^2}{\zeta_n}\right).$$

We define

$$\zeta_n = \left(1 + \frac{1}{\sqrt{\gamma}}\right)^{-n} \quad \text{for } n \leq N$$

where N is such that $\zeta_N \geq \frac{1}{\gamma}$, that is $N \sim \sqrt{\gamma}\log\gamma$. With these definitions we find

$$\left|\frac{\zeta_{n+1}-\zeta_n}{\zeta_n}\right| \leq \frac{C}{\sqrt{\gamma}} \quad \text{for all } 1 \leq n \leq N$$

and thus

$$\frac{1}{\gamma}\sum_{n=0}^{N}\exp(\gamma W) \leq \frac{CN}{\gamma} \sim \gamma^{-1)/2}\log\gamma \to 0 \quad \text{as } \gamma \to \infty,$$

which proves the desired estimate.

It remains to show that (23) implies the pointwise estimate for f. Indeed, (23) implies for $R > 0$ that

$$R^{\gamma}\int_R^{2R} f(x)dx \leq \int_R^{2R} x^{\gamma} f(x)dx \leq \gamma^{\gamma}e^{A\gamma}$$

and thus

$$\int_R^{2R} f(x)dx \leq \exp\big(\gamma\big(\log(\gamma)+\log(R)\big)+A\gamma\big).$$

For bounded kernels the pointwise bound follows now easily from this and (8).

Finally, the second statement follows from the observation that the function $g(x) = f(x)e^{\alpha x}$ satisfies the inequality

$$\begin{aligned} x^2 g(x) &= \int_0^x\int_{x-y}^{\infty} K(y,z)e^{\alpha(x-(y+z))}yg(y)g(z)dzdy \\ &\leq \int_0^x\int_{x-y}^{\infty} K(y,z)yg(y)g(z)dzdy. \end{aligned}$$

which is sufficient to apply the above reasoning to $g(x)$. □

Another key ingredient of the proof of Theorem 3.1 is an a priori estimate on a weighted norm of $Q - \bar{Q}$. Here we have to work quite a bit harder if K is not uniformly close to 2. If it is, then one has the estimate

$$\big|\mathcal{M}f, f(q)\big| \leq C\varepsilon Q^2$$

and thus Q satisfies the perturbed equation $-Q' = (1+O(\varepsilon))Q^2$. The following Lemma then basically follows from integrating the equation for $Q - \bar{Q}$.

Lemma 3.5 *Given* $\delta > 0$ *there exists* $\varepsilon > 0$ *such that there exists* q^* *with* $|q^*+1| \leq \delta$ *and* $\lim_{q\to q^*}|Q(q)| = \infty$.

Furthermore there exists $r > 0$ *such that*

$$\big|(q-q^*)Q(q)+1\big| \leq \delta \quad \text{for all } q \in \big(q^*, q^*+r\big). \tag{29}$$

From now on we rescale the solution such that the singularity of its desingularized Laplace transform Q is at $q = -1$. We denote the corresponding functions again by f and Q respectively.

Since all the transforms are defined on the interval $(-1, \infty)$ we can define the following norm, that is particularly suited for our uniqueness proof:

$$\|Q\| := \sup_{q>-1} \frac{1+q}{|q|} \big|Q(q)\big|. \tag{30}$$

As a corollary of Lemma 3.5 we obtain the following.

Lemma 3.6 *Given $\delta > 0$ there exists $\varepsilon > 0$ such that*

$$\|Q - \bar{Q}\| \leq \delta. \tag{31}$$

Our next goal is to derive a representation formula for $U := Q - \bar{Q}$. Then U satisfies the equation

$$-qU'(q) = (2\bar{Q} - 1)U + U^2 + \mathcal{M}(f, f)(q) \tag{32}$$

and $U = o(\frac{1}{1+q})$ as $q \to -1$.

Lemma 3.7 *The solution to (32) can be represented as*

$$U(q) = -\frac{q}{(1+q)^2} \int_{-1}^{q} \frac{(1+s)^2}{s^2} \int_0^s \psi(\eta) d\eta ds \quad \text{with } \psi = U^2 + \mathcal{M}(f, f). \tag{33}$$

Furthermore, if U_1 and U_2 are two such solutions, then

$$\begin{aligned} U_1(q) - U_2(q) = &-\frac{q}{(1+q)^2} \int_{-1}^{q} \frac{(1+s)^2}{s^2} \big(U_1(s)^2 - U_2(s)^2\big) ds \\ &- \frac{q}{(1+q)^2} \int_{-1}^{q} \frac{(1+s)^2}{s^2} \big(\mathcal{M}(f_1, f_1)(s) - \mathcal{M}(f_2, f_2)(s)\big) ds. \end{aligned} \tag{34}$$

Proof Integrating the equation

$$-qU'(q) = (2\bar{Q} - 1)U + \psi = \left(1 - \frac{2}{1+q}\right)U + \psi$$

gives

$$\left(\frac{(1+q)^2}{q} U\right)' = -\left(\frac{1+q}{q}\right)^2 \psi$$

and thus (33) follows. □

Proposition 3.8 *Let U_1 and U_2 be two solutions of (32) as in Lemma 3.7 then we have $U_1 = U_2$ if $\varepsilon > 0$ is sufficiently small.*

Proof We deduce from (34) that

$$\begin{aligned}\|U_1 - U_2\| \le & \sup_{q>-1} \frac{1}{q+1} \int_{-1}^{q} \frac{(1+s)^2}{s^2} \big|U_1(s)^2 - U_2(s)^2\big| ds \\ & + \sup_{q>-1} \frac{1}{q+1} \left| \int_{-1}^{q} \frac{(1+s)^2}{s^2} \big(\mathcal{M}(f_1, f_1)(s) - \mathcal{M}(f_2, f_2)(s)\big) ds \right| \\ =: & (I) + (II). \end{aligned} \tag{35}$$

The first term is easy to estimate. In fact, using (31), we find for sufficiently small ε that

$$\begin{aligned}\big|(I)\big| &\le \sup_{q>-1} \frac{1}{(1+q)} \int_{-1}^{q} \big(\|U_1\| + \|U_2\|\big) \|U_1 - U_2\| ds \\ &\le \big(\|U_1\| + \|U_2\|\big) \|U_1 - U_2\| \\ &\le \frac{1}{2} \|U_1 - U_2\|. \end{aligned} \tag{36}$$

The main task is to derive a similar bound on the second term in (35). We formulate this main result as a proposition.

Proposition 3.9 *For sufficiently small ε we have*

$$\sup_{q>-1} \frac{1}{1+q} \left| \int_{-1}^{q} \frac{(1+s)^2}{s^2} \big(\mathcal{M}(f_1, f_1)(s) - \mathcal{M}(f_2, f_2)(s)\big) ds \right| \le C\varepsilon \|U_1 - U_2\|. \tag{37}$$

With Proposition 3.9 the statement of the Proposition 3.8 follows. □

It remains to prove Proposition 3.9. We first notice that it suffices to prove it for $W(x, y)$ that satisfies (12) and (11) with $\varepsilon = 1$. The result then follows by scaling. For the proof of Proposition 3.9 we argue by contradiction. Suppose that (37) (with $\varepsilon = 1$) is not true. Then there exist sequences $\{W_n\}$, $\{f_{1,n}\}$, $\{f_{2,n}\}$ and $\{q_n\}$ such that, with $U_{i,n}$ denoting the corresponding functions as above,

$$\|U_{1,n} - U_{2,n}\| \to 0 \quad \text{as } n \to \infty \tag{38}$$

and

$$\frac{1}{q_n+1} \left| \int_{-1}^{q_n} \frac{(1+s)^2}{s^2} \big(\mathcal{M}(f_{1,n}, f_{1,n})(s) - \mathcal{M}(f_{2,n}, f_{2,n})(s)\big) ds \right| \ge 1. \tag{39}$$

By our regularity assumption (11) we can assume without loss of generality that there exists a function $W_* = W_*(x, y)$, satisfying (11) and (12) such that

$$W_n \to W_* \quad \text{locally uniformly on } (0, \infty)^2. \tag{40}$$

We now collect some a-priori estimates for solutions f.

Lemma 3.10 *Let f be a solution to* (8). *Then*

$$\int_0^\infty f(x)dx \le 2, \tag{41}$$

$$\left|\int_0^\infty \big(1-e^{-qx}\big) f(x)dx\right| \le \frac{2|q|}{1+q} \quad \textit{for all } q > -1, \tag{42}$$

$$\int_R^{2R} e^x f(x)dx \le 4R \quad \textit{for } R \ge \frac{1}{1-\log 2}, \tag{43}$$

$$\int_1^\infty \frac{e^x}{x^3} f(x)dx \le C. \tag{44}$$

Proof The first two estimates (41) and (42) follow from (31). We can now deduce (43) from (42). In fact, choosing $q < -\log 2$, we have

$$\int_0^\infty \big(e^{-qx}-1\big) f(x)dx = \left|\int_0^\infty \big(1-e^{-xq}\big) f(x)dx\right| \ge \frac{1}{2}\int_1^\infty e^{-qx} f(x)dx.$$

As a consequence we obtain

$$\int_1^\infty e^{|q|x} f(x)dx \le \frac{4}{1+q} \quad \text{for } q \in (-1, -\log 2).$$

Choosing now $1+q = \frac{1}{R}$ and $x \in (R, 2R)$ estimate (43) follows. Finally, estimate (44) follows from (43) via a dyadic argument, that is

$$\int_1^\infty \frac{e^x}{x^3} f(x)dx \le \sum_{n=0}^\infty \int_{2^n}^{2^{n+1}} \frac{e^x}{x^3} f(x)dx \le C \sum_{n=0}^\infty 2^{n+1} 2^{-3(n+1)} \le C.$$

□

We now write

$$\begin{aligned}
&\frac{1}{q+1}\int_{-1}^q \frac{(1+s)^2}{s^2}\big(\mathcal{M}(f_1,f_1)(s) - \mathcal{M}(f_2,f_2)(s)\big)ds \\
&\quad = \int_0^\infty \int_0^\infty W(x,y)\big(f_1(x)+f_2(x)\big)\big(f_1(y)-f_2(y)\big)H(q,x,y)dxdy
\end{aligned}$$

with

$$H(q,x,y) = \frac{1}{1+q}\int_{-1}^q \frac{(1+s)^2}{s^2}\big(1-e^{-sx}\big)\big(1-e^{-sy}\big)ds. \tag{45}$$

We assume first that $q_n \to q^* \in (-1, \infty]$. In this case we can use the following estimate for H, that is if $q > -1 + \frac{1}{L}$ we have

$$0 \le H(q,x,y) \le C_L \frac{\min(x,1)\min(y,1)}{1+(x+y)^3} e^{x+y}. \tag{46}$$

Furthermore, if $q_n \to q^* \in (-1, \infty)$, then it is obvious that $H(q_n, \cdot)$ converges locally uniformly in $(0,\infty)^2$ to $H(q^*, \cdot, \cdot)$. If $q_n \to \infty$, then $H(q_n, \cdot)$ converges locally uniformly to 2.

Then, if $q > -1 + \frac{1}{L}$ and if f is a solution to (8), we have, using (46), that for large R

$$\begin{aligned}
&\int_0^{1/R}\int_0^\infty W(x,y)f(x)f(y)H(q,x,y)dydx\\
&\quad\le C\int_0^{1/R}\int_0^\infty f(x)f(y)\frac{x\min(y,1)}{1+(x+y)^3}e^{x+y}dydx\\
&\quad\le C\int_0^{1/R} xf(x)dx\int_0^\infty \frac{\min(y,1)e^y}{1+y^3}f(y)dy \le \frac{C}{R},
\end{aligned}$$

where the last estimate follows from (41) and (44).

Furthermore, using also (43), we arrive similarly at

$$\begin{aligned}
&\int_R^\infty\int_0^\infty W(x,y)f(x)f(y)H(q,x,y)dydx\\
&\le C\int_R^\infty \frac{e^x}{1+x^3}f(x)dx\\
&\le C\sum_{n=0}^\infty\int_{R2^n}^{R2^{n+1}}\frac{e^x}{x^3}f(x)dx \le C\sum_{n=0}^\infty \big(R2^n\big)^{-3}R2^n \le \frac{C}{R^2}\sum_{n=0}^\infty 2^{-2n}\le \frac{C}{R^2}.
\end{aligned}$$

Hence, in order to arrive at a contradiction to (39), it remains to show that for large but fixed R

$$\int_{1/R}^R\int_{1/R}^R W_n(x,y)\big(f_{1,n}(x)+f_{2,n}(x)\big)\big(f_{1,n}(y)-f_{2,n}(y)\big)H(q_n,x,y)dxdy \to 0 \tag{47}$$

as $n\to\infty$. Since W_n and $H(q_n,\cdot,\cdot)$ converge locally uniformly to their respective limits and since assumption (38) in particular implies that $f_{1,n}-f_{2,n}\to 0$ locally in the sense of measures, we find that

$$F_n(x) := \int_{1/R}^R W_n(x,y)H(q_n,x,y)\big(f_{1,n}(y)-f_{2,n}(y)\big)dy \to 0 \quad \text{as } n\to\infty$$

locally uniformly in x. Hence, we can derive (47) and we have proved a contradiction in case $q_n \to q^* \in (-1,\infty]$.

The case $q_n \to -1$ is somewhat more difficult to treat. We introduce the rescaling

$$X = (1+q_n)x \quad \text{and} \quad g(X) = e^x f(x). \tag{48}$$

The integral (cf. (39)) for which we want to show that it converges to zero as $n \to \infty$ becomes

$$\int_0^\infty \int_0^\infty W_n(X,Y)\big(g_{1,n}(X)+g_{2,n}(X)\big)\big(g_{1,n}(Y)-g_{2,n}(Y)\big)\tilde{H}(q_n,X,Y)\,dX\,dY \tag{49}$$

with

$$\tilde{H}(q_n,X,Y)=\frac{e^{-(X+Y)/(1+q_n)}}{(1+q_n)^2}H\left(q_n,\frac{X}{1+q_n},\frac{Y}{1+q_n}\right). \tag{50}$$

The proof from now on follows similar lines as in the case $q_n \to q^* > -1$. We again refer for details to [29].

3.3 Non-solvable Kernels: Solutions with Fat Tails

We will now come back to the question of existence of self-similar solutions that do not have finite mass. The first such result for a non-solvable kernel has been obtained in [26] for the special case of the diagonal kernel with homogeneity $\gamma \in (-\infty, 1)$, that is $K(x,y)=x^{1+\gamma}\delta_{x-y}$. In this case the coagulation equation reduces to

$$\partial_t n(t,\xi)=\frac{1}{4}\left(\frac{\xi}{2}\right)^{1+\gamma} n^2\left(\frac{\xi}{2}\right)-\xi^{1+\gamma}n^2(\xi). \tag{51}$$

A specific property of (51) is that the evolution of $n(t,\xi)$ depends only on the value in $\xi/2$ but in particular not on values larger than ξ.

It has been established in [26] that for any $\rho \in (\gamma, 1)$ there exists a self-similar solution with a corresponding profile that decays as $x^{-(1+\rho)}$ as $x \to \infty$. It seems somewhat surprising that one obtains solutions that increase as $x \to \infty$. It is not clear to us whether this is another specialty of the diagonal kernel, at least so far all other existence results for fat tail solutions need $\rho > 0$ (see also [30]). The proof of existence of fat tail self-similar solutions to (51) is in fact rather simple and follows the approach of [19]. The idea is to reformulate the equation for the self-similar profile as an initial value problem starting at $x = 0$. For this to be uniquely solvable one needs to expand the solution up to second order around the expected power law behaviour as $x \to 0$. The standard contraction mapping theorem applied to this set of functions give a unique self-similar profile in a neighbourhood of $x \sim 0$. Certain monotonicity properties of Eq. (51) that are specific to the diagonal kernel allow to extend the solution to $(0, \infty)$ and to guarantee that it remains positive. Note that this result gives also a uniqueness result, but only in the class of functions with a certain power law behaviour at $x \to 0$. To prove that every solution behaves in this way does however not seem easy to establish.

For more general coagulation kernels one needs to develop another strategy. The main problem in any existence proof is to avoid the loss of mass at infinity and by this to avoid to obtain the trivial solution in the end. In [28] a fixed point method is used in a suitable subset of measures that encode the expected decay behaviour in a weak sense.

We continue by describing the main results and the ideas of the proof in [28]. For that purpose it is convenient to go over to the mass density function $h(x,t)=xf(x,t)$

and to introduce the parameter $\rho = \gamma + \frac{1}{\beta}$. Then, after rescaling, the time dependent version of equation (7) becomes

$$\partial_t h + \partial_x \left[\int_0^x \int_{x-y}^{\infty} \frac{K(y,z)}{z} h(z)h(y)dzdy \right] - \beta\big[\partial_x(xh) + (\rho-1)h\big] = 0, \tag{52}$$

with initial data

$$h(x,0) = h_0(x). \tag{53}$$

The kernel K is assumed to be continuous and to be homogeneous of degree $\gamma \in [0, 1)$, that is

$$K(ax, ay) = a^{\gamma} K(x, y) \quad \text{for all } x, y \in (0, \infty). \tag{54}$$

The key assumption in [26] is the following growth condition

$$K(x, y) \leq C\big(x^{\gamma} + y^{\gamma}\big) \quad \text{for all } x, y \in (0, \infty). \tag{55}$$

Theorem 3.11 ([28]) *Given $\gamma \in [0, 1)$ and a kernel K that satisfies assumptions (54)–(55), then for any $\rho \in (\gamma, 1)$ there exists a weak stationary solution h to (52). This solution is nonnegative, continuous and satisfies*

$$h(x) \sim (1-\rho)x^{-\rho} \quad \text{as } x \to \infty.$$

While this result covers a wide range of kernels, in particular for example the product kernel $K(x, y) = (xy)^{\gamma/2}$ and the general sum kernel $K(x, y) = x^{\alpha} y^{\gamma-\alpha} + y^{\alpha} x^{\gamma-\alpha}$ with $\alpha \geq 0$, it does not apply to singular kernels, such as Smolochowski's kernel $K(x, y) = (x^{1/3} + y^{1/3})(x^{-1/3} + y^{-1/3})$. The main reason is that for the type of kernels considered in [28] the global estimate (57) below suffices to prove that the nonlinear integral terms in (52) are well defined, but this is not sufficient for singular kernels. In this case one first has to regularize the kernel for small x and y and undertake another limit procedure in the end. The corresponding analysis is currently work in preparation [30].

It is also worth mentioning that Theorem 3.11 provides under rather minimal assumptions on the kernel K the existence of a weak continuous solution. One would expect that for a kernel that is smooth on $(0, \infty)^2$, the self-similar solution is smooth in $(0, \infty)$ as well. The proof of such a property is however not, as discussed earlier, a straightforward bootstrap argument due to the possibly singular behavior of solutions near $x = 0$, but we expect that methods that have been developed to establish smoothness of self-similar solutions with finite mass can be applied here as well.

The strategy to find a stationary solution to (52) will in principle be the following. One considers the corresponding evolution problem and proves that it preserves a convex set that is compact in the weak topology and contains functions with the expected decay behaviour. This allows to prove the existence of a fixed point using the following variant of Tykonov's fixed point theorem.

Theorem 3.12 (Theorem 1.2 in [8]) *Let X be a Banach space and $(S_t)_{t\geq 0}$ be a continuous semigroup on X. Assume that S_t is weakly sequentially continuous for any $t > 0$ and that there exists a subset $\mathcal{Y}$ of X that is nonempty, convex, weakly sequentially compact and invariant under the action of S_t. Then there exists $z_0 \in \mathcal{Y}$ which is stationary under the action of S_t.*

Unfortunately it is not so easy to prove well-posedness directly for (52)–(53) since one has to consider the well-posedness of the problem in a space of functions that are singular at the origin. Instead, one needs to consider first a family of regularized problems and prove that self-similar solutions for this regularized problem exist and satisfy uniform estimates that allow to pass to the limit in the corresponding equation. For details of this regularization process we refer to [28]

The key step in order to apply Theorem 3.12 is to identify suitable subsets of certain Banach spaces and to show that they are invariant under the evolution. In [28] the underlying space is the metric space $\mathcal{X}_\rho$ of nonnegative Radon measures, which is denoted by some abuse of notation by $h dx \in \mathcal{M}^+([0,\infty))$, satisfying the condition

$$\sup_{R\geq 0} \frac{\int_{[0,R]} h dx}{R^{1-\rho}} < \infty. \tag{56}$$

The set that will be shown to be invariant under the evolution induced by (52) will be the set $\mathcal{Y}$ of measures $h dx \in \mathcal{M}^+([0,\infty))$ satisfying

$$\int_{[0,R]} h dx \leq R^{1-\rho}, \quad R \geq 0 \tag{57}$$

$$\int_{[0,R]} h dx \geq R^{1-\rho}\left(1 - \frac{R_0^\delta}{R^\delta}\right)_+, \quad R \geq 0, \tag{58}$$

for a sufficiently large R_0 and a sufficiently small $\delta > 0$. It is straightforward to see that this set is convex and compact in the weak topology. The heart of the analysis in [28] is the proof of the invariance of (57) and (58) under the evolution (52). The upper bound (57) can be proved by analyzing a simple differential inequality that is satisfied by $\int_{[0,R]} h dx$. The proof of the invariance of (58) is more delicate and is based on the analysis of the corresponding dual problem.

To explain the main idea it is useful to comment first on the particular choice (58) for a lower bound. If one considers (52)–(53) with $K \equiv 0$ one obtains a linear transport equation that has the explicit solution $h(x,t) = e^{\rho\beta t} h_0(x e^{\beta t})$. If one takes data $h_0(x) = x^{-\rho}(1 - Cx^{-\delta})$ one finds

$$h(x,t) = x^{-\rho}\left(1 - \frac{Ce^{-\delta\beta t}}{x^\delta}\right) \sim x^{-\rho}\left(1 - \frac{C}{x^\delta}\right) + C\delta\beta t x^{-(\rho+\delta)}$$

and thus obtains an improved lower bound for positive times. The main task is then to show that the additional error terms induced by the nonlinear coagulation term can be absorbed into the positive term if δ is sufficiently small. It is easy to see, by testing (52)–(53) with a function $\psi = \psi(x,t)$, that after some rearrangements one obtains

$$\int \psi(x,t) h_t dx = \int h_0(x)\psi(x,0) h_0 dx$$

if ψ solves the associated dual problem

$$-\partial_s \psi(x,s) - \int_0^\infty \frac{K(x,z)}{z}\big[\psi(x+z,s) - \psi(x,s)\big] h_s dz$$
$$+ \beta x \partial_x \psi(x,s) - \beta(\rho-1)\psi(x,s) = 0. \tag{59}$$

Since we want to estimate $\int_{[0,R]} h_t dx$, we pose the initial condition

$$\psi(x,t) = \chi_{[0,R]}(x). \tag{60}$$

Thus, in order to estimate $\int_{[0,R]} h_t dx$ one needs to estimate $\psi(x,0)$ from below. With regard to an estimate of $\psi(x,0)$, it turns out that in the case of kernels satisfying (55) and measures h satisfying (57) and (58) one can construct a subsolution for ψ by replacing the term $\frac{K(x,z)}{z} h_s$ by a suitable power law. The equation for the subsolution has an explicit self-similar solution. Finally, it remains to work out that this subsolution is sufficiently good to show that (58) is preserved under the evolution. With all these ingredients one can then apply Theorem 3.12 to obtain the existence of a weak stationary solution to (52). The final step in the proof of Theorem 3.11 is to show that the weak solution has a density that is in fact also continuous on $(0,\infty)$ and has the desired decay behaviour.

Acknowledgements I am indebted to Philippe Laurençot and Juan Velázquez for numerous discussions and collaborations on the subject of this article. Support through the Cooperative Research Centre (CRC) 1060 "The Mathematics of Emergent Effects" at the University of Bonn is also gratefully acknowledged.

References

1. Aldous, D.J.: Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. Bernoulli **5**(1), 3–48 (1999)
2. Ball, J.M., Carr, J.: The discrete coagulation-fragmentation equations: existence, uniqueness, and density conservation. J. Stat. Phys. **61**(1–2), 203–234 (1990)
3. Cañizo, J., Mischler, S.: Regularity, asymptotic behavior and partial uniqueness for Smoluchowski's coagulation equation. Rev. Mat. Iberoam. **27**(3), 503–564 (2011)
4. Da Costa, F.P.: On the dynamic scaling behaviour of solutions to the discrete Smoluchowski equations. Proc. Edinb. Math. Soc. **39**(3), 547–559 (1996)
5. Deaconu, M., Tanré, E.: Smoluchowski's coagulation equation: probabilistic interpretation of solutions for constant, additive and multiplicative kernels. Ann. Sc. Norm. Super. Pisa, Cl. Sci. **29**(3), 549–579 (2000).
6. Drake, R.-L.: A general mathematical survey of the coagulation equation. In: Topics in Current Aerosol Research (Part 2), International Reviews in Aerosol Physics and Chemistry, pp. 203–376. Oxford University Press, Oxford (1972)

7. Escobedo, M., Mischler, S.: Dust and self-similarity for the Smoluchowski coagulation equation. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **23**(3), 331–362 (2006)
8. Escobedo, M., Mischler, S., Perthame, B.: Gelation in coagulation and fragmentation models. Commun. Math. Phys. **231**(1), 157–188 (2002)
9. Escobedo, M., Mischler, S., Rodriguez Ricard, M.: On self-similarity and stationary problem for fragmentation and coagulation models. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **22**(1), 99–125 (2005)
10. Fournier, N., Laurençot, P.: Existence of self-similar solutions to Smoluchowski's coagulation equation. Commun. Math. Phys. **256**(3), 589–609 (2005)
11. Fournier, N., Laurençot, P.: Local properties of self-similar solutions to Smoluchowski's coagulation equation with sum kernels. Proc. R. Soc. Edinb., Sect. A, Math. **136**(3), 485–508 (2006)
12. Fournier, N., Laurençot, P.: Well-posedness of Smoluchowski's coagulation equation for a class of homogeneous kernels. J. Funct. Anal. **233**(2), 351–379 (2006)
13. Jeon, I.: Existence of gelling solutions for coagulation-fragmentation equations. Commun. Math. Phys. **194**(3), 541–567 (1998)
14. Kreer, M., Penrose, O.: Proof of dynamical scaling in Smoluchowski's coagulation equation with constant kernel. J. Stat. Phys. **75**(3–4), 389–407 (1994)
15. Laurençot, P., Mischler, S.: From the discrete to the continuous coagulation-fragmentation equations. Proc. R. Soc. Edinb., Sect. A, Math. **132**(5), 1219–1248 (2002)
16. Laurençot, P., Mischler, S.: On coalescence equations and related models. In: Modeling and Computational Methods for Kinetic Equations. Model. Simul. Sci. Eng. Technol., pp. 321–356. Birkhäuser, Boston (2004)
17. Laurençot, P., Mischler, S.: Lyapunov functionals for Smoluchowski's coagulation equation and convergence to self-similarity. Monatshefte Math. **146**(2), 127–142 (2005)
18. Leyvraz, R.: Scaling theory and exactly solvable models in the kinetics of irreversible aggregation. Phys. Rep. **383**, 95–212 (2003)
19. Leyvraz, F.: Rigorous results in the scaling theory of irreversible aggregation kinetics. J. Nonlinear Math. Phys. **12**(suppl. 1), 449–465 (2005)
20. Leyvraz, F., Tschudi, H.R.: Singularities in the kinetics of coagulation processes. J. Phys. A **14**(12), 3389–3405 (1981)
21. McLeod, J.B.: On an infinite set of non-linear differential equations. Q. J. Math. **13**, 119–128 (1962)
22. McLeod, J.B., Niethammer, B., Velázquez, J.J.L.: Asymptotics of self-similar solutions to coagulation equations with product kernel. J. Stat. Phys. **144**, 76–100 (2011)
23. Menon, G., Pego, R.L.: Approach to self-similarity in Smoluchowski's coagulation equations. Commun. Pure Appl. Math. **57**(9), 1197–1232 (2004)
24. Mulholland, G.W., Samson, R.J., Mountain, R.D., Ernst, M.H.: Cluster size distribution for free molecular agglomeration. Energy Fuels **2**, 481–486 (1988)
25. Niethammer, B., Velázquez, J.J.L.: Optimal bounds for self-similar solutions to coagulation equations with product kernel. Commun. Partial Differ. Equ. **36**(12), 2049–2061 (2011)
26. Niethammer, B., Velázquez, J.J.L.: Self-similar solutions with fat tails for a coagulation equation with diagonal kernel. C. R. Math. Acad. Sci. Paris **349**(9–10), 559–562 (2011)
27. Niethammer, B., Velázquez, J.J.L.: Exponential decay of self-similar profiles to Smoluchowski's coagulation equation (2013). Preprint, arXiv:1310.4732
28. Niethammer, B., Velázquez, J.J.L.: Self-similar solutions with fat tails for Smoluchowski's coagulation equation with locally bounded kernels. Commun. Math. Phys. **318**, 505–532 (2013)
29. Niethammer, B., Velázquez, J.J.L.: Uniqueness of self-similar solutions to Smoluchowski's coagulation equations for kernels that are close to constant (2013). Preprint, arXiv:1309.4621
30. Niethammer, B., Throm, S., Velázquez, J.J.L.: Self-similar solutions with fat tails for Smoluchowski's coagulation equation with singular kernels (2013, in preparation)
31. Norris, J.R.: Smoluchowski's coagulation equation: uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent. Ann. Appl. Probab. **9**(1), 78–109 (1999)
32. Silk, J., Takamasa, T.: A statistical model for the initial stellar mass function. Astrophys. J. **229**, 242–256 (1979)
33. Smoluchowski, M.: Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. Phys. Z. **17**, 557–599 (1916)
34. van Dongen, P.G.J., Ernst, M.H.: Scaling solutions of Smoluchowski's coagulation equation. J. Stat. Phys. **50**(1–2), 295–329 (1988)

Barbara Niethammer has studied mathematics at the University of Bonn and held postdoctoral positions at the Courant Institute, the University of Bonn and the Max-Planck Institute for Mathematics in the Sciences in Leipzig. She was professor at the Humboldt University in Berlin and the University of Oxford before she returned to Bonn in 2012. Her research interests are in the analysis of mathematical models that arise in the kinetics of phase transitions and material science.

Jahresber Dtsch Math-Ver (2014) 116:67–71
DOI 10.1365/s13291-014-0082-x

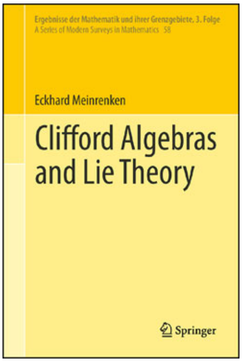
BOOK REVIEW

Eckhard Meinrenken: "Clifford algebras and Lie theory"

Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Vol. 58, Springer, 2013, 321 pp

Ilka Agricola

Published online: 29 January 2014



If mathematics is a landscape whose regions are its different research areas, this book lives in the algebra region, at the main junction between the highways to harmonic analysis and quantization, representation theory, global analysis and spin geometry, and from where smaller roads heading to topology (more precisely, the neighbourhoods of homology and cohomology), complex geometry, supergeometry, and index theorems depart. Thus, any motivating path to this book has the drawback to start somewhere outside its core topic, and may therefore present a subjective perspective on the book. Such an interdisciplinary approach, moreover, may mislead the travelling mathematician to believe that the book is 'only' a collection of algebraic tools needed for accessing these other areas.

Such point of view is wrong. Meinrenken's research monograph is a self-contained book about a new fascinating research topic in non-commutative algebra, that has a number of interesting connections to other cutting-edge subjects; this partially explains why this area is so rich and beautiful. The writing of such a text is a considerable challenge on its own. In the present case the author took the decision not to leave his area; rather, he chose to show the routes to other topics whenever they appeared, as marked by an impressive number of side-turns and bibliographic comments. Additionally, one of the work's strong points is that it contains abundant original material that appears for the first time in a textbook.

I. Agricola (✉)
Fachbereich Mathematik und Informatik, Philipps-Universität, Marburg, Germany
e-mail: agricola@mathematik.uni-marburg.de

One good way to understand the key achievements and the current status of a certain branch of pure mathematics is to follow its historical development. Therefore, I shall approach the topic not by starting in some area and following the highway to 'Clifford algebras and Lie theory', but rather by looking back in time when these highways were still meandering farm tracks, when many roads hadn't been yet constructed, and when areas that are nowadays well connected seemed distant and unrelated. In fact, this is presumably quite close to how the author (and his long-lasting collaborator Anton Alekseev) came to think about the questions that triggered the research work exposed in this book. Along the way I will comment on the methods and tools needed to advance in the theory and indicate where in the book they are treated.

In the 1960s Alexander Kirillov proposed the so-called 'orbit method' to describe unitary representations of Lie groups. This boils down to the fact that given a Lie group G with Lie algebra $\mathfrak{g}$, the orbits of the coadjoint action of G on the dual $\mathfrak{g}^*$ are symplectic manifolds for the Kostant-Souriau form, and fibre bundles over them carry unitary representations of G. While this worked neatly for nilpotent groups, the technical difficulties were tremendous in virtually any other case, which spawned a large amount of research activity by many mathematicians. For us, what is important is that Kirillov conjectured a 'universal character formula' for such representations, which roughly says the following: given a coadjoint orbit M with Liouville measure β, the character of the associated Hilbert space representation T should look like

$$\operatorname{tr} T(\exp X) = \int_M e^{iF(X)} J(X)^{-1/2}\, d\beta(F).$$

This is meant as an identity of distributions on a suitable neighbourhood of 0 inside $\mathfrak{g}$, X is an element of the Lie algebra $\mathfrak{g}$, hence indeed $\exp X \in G$; since F belongs to M, a subset of $\mathfrak{g}^*$, its action on X is well defined. The most interesting part is the occurence of the mysterious function $J(X)$,

$$J(X) = \det\bigl(j(\mathrm{ad}_X)\bigr), \quad j(z) = \frac{\sinh(z/2)}{z/2}.$$

This turns out to be a smooth map on all of $\mathfrak{g}$; recall that $\mathrm{ad}_X Y = [X, Y]$, so this is meant as a formal power series in the operators ad_X. This function and its many relatives have various manifestations in mathematics—for example, $j(z)$ is basically the generating function of the $\hat{A}$-genus, an integer number on spin manifolds that coincides with the index of the Dirac operator by the Atiyah–Singer index theorem. This already brings us close to Clifford algebras, even if the link is not transparent! Besides, mathematical experience tells us that 'good' character formulas can (and should) be understood as index formulas in disguise.

The orbit method, albeit more a philosophical principle than a rigorous method, inspired another famous result. For any Lie group G with Lie algebra $\mathfrak{g}$ let us consider the following two important algebras: the universal algebra $\mathcal{U}(\mathfrak{g})$—that is, the quotient of the tensor algebra $T(\mathfrak{g})$ by the relations $X \otimes Y - Y \otimes X = [X, Y]$)—and the symmetric algebra $S(\mathfrak{g})$, i.e. the polynomial ring on a basis of $\mathfrak{g}$ (Chap. 5; observe that this chapter contains an interesting alternative proof of the Poincaré-Birkhoff-Witt Theorem due to Emanuela Petracci, [8]). The inclusion of the symmetric algebra $S(\mathfrak{g})$ in $T(\mathfrak{g})$ followed by the quotient map gives an isomorphism $\mathrm{sym}: \ S(\mathfrak{g}) \to \mathcal{U}(\mathfrak{g})$, which, alas, is not compatible with the product structure of the two algebras. In 1977

Michel Duflo proved that the composition $\mathrm{sym} \circ J^{1/2}: S(\mathfrak{g}) \to \mathcal{U}(\mathfrak{g})$, if interpreted in the right way, restricts to an *algebra isomorphism* between the algebra of G-invariant polynomials $S(\mathfrak{g})^G$ and the center $\mathcal{U}(\mathfrak{g})^G$ of the universal enveloping algebra [3]. This looks like a purely algebraic statement at first sight; however, $\mathcal{U}(\mathfrak{g})$ can be identified in a natural way with the left-invariant differential operators on G, and $S(\mathfrak{g})$ with the differential operators on $\mathfrak{g}$ with constant coefficients. Duflo's theorem, therefore, yields an algebra isomorphism between *bi-invariant* differential operators on G and G*-invariant*, constant-coefficient differential operators on $\mathfrak{g}$, and both algebras are commutative. A set of generators of $\mathcal{U}(\mathfrak{g})^G \cong S(\mathfrak{g})^G$ is typically called a set of *Casimir operators*. If $\mathfrak{g}$ is semisimple, the number of Casimir operators is equal to the rank of $\mathfrak{g}$, and Duflo's map coincides with the Harish-Chandra isomorphism. For example, if $G = \mathrm{SU}(2) \cong S^3$ is the 3-dimensional Lie group of unitary 2×2 matrices with determinant 1, there exists exactly one Casimir operator C; as an element of $S(\mathfrak{g})^G$, it is a quadratic polynomial, and in the usual basis X_1, X_2, X_3 of $\mathfrak{g} = \mathfrak{su}(2)$ (the trace-free skew-hermitian 2×2 matrices), we can write it as $C = X_1^2 + X_2^2 + X_3^2$; viewed as bi-invariant differential operator on $G \cong S^3$ it coincides with the Laplacian, which is known to be invariant under unitary transformations. As one learns in any standard lecture on quantum mechanics, finite-dimensional representations of $G = \mathrm{SU}(2)$ can be characterized by the eigenvalues of the Casimir operator. Duflo's theorem is a beautiful generalization of this collection of classical results. Unfortunately, his original proof is very technical and sophisticated. Again, this theorem was the starting point for a wealth of subsequent developments that point at many different mathematical directions (see for example [4] for additional reading).

The starting point of Meinrenken's book is the observation that the situation can be drastically simplified if the Lie algebra $\mathfrak{g}$ carries a non-degenerate symmetric bilinear form $B : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ that is invariant under $\mathrm{Ad}\, G$, i.e. satisfying

$$B(\mathrm{Ad}_g X, \mathrm{Ad}_g Y) = B(X, Y) \quad \forall X, Y \in \mathfrak{g},\ \forall g \in G.$$

Recall that if G is a matrix group, the adjoint action of G on $\mathfrak{g}$ is just conjugation, $\mathrm{Ad}_g X = gXg^{-1}$, i.e. we require the scalar product to be constant on conjugacy classes of matrices; in the previous example $G = \mathrm{SU}(2)$ the form $B(X, Y) = \mathrm{tr}(XY)$ does the job. Lie algebras of this type are called *quadratic*, and they include many interesting instances (for example, all semisimple ones). Given the scalar product B, many further tools and results become available. One can consider a third algebra (besides the universal and symmetric algebras), namely, the Clifford algebra (Chap. 2). This is the quotient of the tensor algebra of $\mathfrak{g}$ by the relations $X \otimes Y - Y \otimes X - 2B(X, Y)$,

$$\mathrm{Cl}(\mathfrak{g}) := T(\mathfrak{g}) / \big\langle X \otimes Y - Y \otimes X - 2B(X, Y) : X, Y \in \mathfrak{g} \big\rangle.$$

Since the exterior algebra $\Lambda(\mathfrak{g})$ of $\mathfrak{g}$ can be embedded in $T(\mathfrak{g})$ as totally antisymmetric tensors, we have a vector space isomorphism

$$q : \Lambda(\mathfrak{g}) \longrightarrow \mathrm{Cl}(\mathfrak{g}),$$

called the *quantization map*. Roughly speaking, the Duflo factor $J^{1/2}(X)$ measures the difference between two possible ways to embed the exponential of skew-symmetric matrices into $\mathrm{Cl}(\mathfrak{g})$ (Chap. 4). In fancy argot, it can be interpreted as a

'Berezin integral' and is deeply linked to the quantization map q. The exterior and the symmetric algebra can be combined into a differential algebra $W(\mathfrak{g})$, the *Weil algebra*, which the factorwise map $\mathrm{sym} \otimes q$ maps to the *quantum Weil algebra* $\mathcal{W}(\mathfrak{g})$,

$$\mathrm{sym} \otimes q: \quad W(\mathfrak{g}) := S(\mathfrak{g}) \otimes \Lambda(\mathfrak{g}) \longrightarrow \mathcal{W}(\mathfrak{g}) := \mathcal{U}(\mathfrak{g}) \otimes \mathrm{Cl}(\mathfrak{g}).$$

The algebra $\mathcal{W}(\mathfrak{g})$ was defined and investigated in the article [2]. The classical Weil algebra $W(\mathfrak{g})$ was introduced by Henri Cartan (1950) as an algebraic model for the algebra of differential forms on the classifying space EG (for G compact). Both Weil algebras carry a lot of structure—graduations, differentials, Lie derivatives and contraction operators—all of which is described in great detail in the book using the language of abstract enveloping algebras, Hopf algebras, super spaces, and $\mathfrak{g}$-differential spaces (Chaps. 6 and 7). Not surprisingly, in the light of Duflo's result the map $\mathrm{sym} \otimes q$ is not compatible with this extra data. Alekseev and Meinrenken prove that there is a suitable generalization of $J^{1/2}$ (let us denote it by the same symbol for simplicity) such that $\mathcal{Q} := (\mathrm{sym} \otimes q) \circ J^{1/2} : W(\mathfrak{g}) \to \mathcal{W}(\mathfrak{g})$ *is* compatible with the additional structure of these algebras. Since Duflo's theorem can now be obtained by restriction to $S(\mathfrak{g}) \subset W(\mathfrak{g})$, this approach gives a much simpler proof of this classical result for quadratic Lie algebras (Sect. 7.3). This illustrates neatly the power of the methods exposed in this book.

There exists a natural element in the center of $\mathcal{W}(\mathfrak{g})$, which one should understand as the non-commutative relative of the quadratic Casimir operator. Observe that the scalar product B allows to define a closed, bi-invariant 3-form on G,

$$T(X, Y, Z) = -\frac{1}{6} B\big([X, Y], Z\big) \quad \forall\, X, Y, Z \in \mathfrak{g}.$$

Let $e_1, \ldots, e_n$ be an orthonormal basis of $\mathfrak{g}$ with respect to B, viewed both as a subset of $\mathcal{U}(\mathfrak{g})$ and of $\mathrm{Cl}(\mathfrak{g})$. We can then define 'Kostant's cubic Dirac operator' ([5], [7]; Sect. 7.2)

$$D := \sum_{i=1}^{n} e_i \otimes e_i + 1 \otimes T \in \mathcal{W}(\mathfrak{g}),$$

which has the remarkable property that D^2 lies in the center of $\mathcal{W}(\mathfrak{g})$. The kernel of the twisted cubic Dirac operator is described in terms of representations and used to prove a character formula that generalizes the classical Weyl character formula (Chap. 8). In fact, the element D can be understood as the symbol of a true Dirac operator on G. Its analogue on homogeneous spaces G/K (its algebraic counterpart is called 'relative Dirac operator' in the book) turns out to be the usual Dirac operator, although not with respect to the Levi-Civita connection, but rather for a metric connection with skew torsion induced by the K-principal fibre bundle $G \to G/K$ (Chap. 9; see also [1]). Again, one sees that the presence of a scalar product plays a crucial role.

Invariant elements in the different non-commutative algebras associated with a Lie algebra are one possible road map to this book. The last space of invariants discussed is $\Lambda(\mathfrak{g})^{\mathfrak{g}}$ (for $\mathfrak{g}$ reductive). The exterior algebra $\Lambda(\mathfrak{g})$ carries a differential map d that makes it a complex (the Chevally-Eilenberg complex), thus it induces a cohomology theory $H^*(\mathfrak{g})$. It is shown that $\Lambda(\mathfrak{g})^{\mathfrak{g}}$ and $H^*(\mathfrak{g})$ are isomorphic, and that they can then be identified with the exterior algebra of a certain subspace of $\Lambda(\mathfrak{g})$, the so-called space of *primitive elements* $P(\mathfrak{g})$ (Hopf-Koszul-Samelson Theorem). Primitivity is a notion coming from the theory of Hopf algebras that I will not explain here; I'll just observe that we already encountered a primitive element, namely the 3-form T. Remarkably, $P(\mathfrak{g})$ has a natural interpretation in terms of Weil algebras (Chap. 10). Furthermore, there is a Clifford analogue of this Theorem that identifies $\mathrm{Cl}(\mathfrak{g})^{\mathfrak{g}}$ with the Clifford algebra $\mathrm{Cl}(P(\mathfrak{g}))$, another result of Kostant ([6]; Chap. 11).

At this juncture our mathematical *tour d'horizon* comes to an end. Meinrenken's book is far from a standard textbook on a standard topic, meaning that the reader will find a profusion of material presented from many different viewpoints. Even if all topics are cultivated, and thrive, within the gates of the garden of algebra, some are likely to be new or at least unfamiliar. Although the book starts with a review of basic results (symmetric bilinear forms, Clifford algebras, Lie algebras ...), it assumes a good knowledge from these areas, for the exposition is elegantly short, with few motivating examples. The title might be slightly misleading in the sense that the text is not meant to be (nor include) a complete treatise on Clifford algebras and spin representations. The reader should be comfortable with the methods of abstract algebra, lest he feel lost already in the preparations to the most interesting results. But then he is rewarded with a concise account of a modern topic in non-commutative algebra, a fresh view over invariants and a large supply of pointers that lead directly to current research.

References

1. Agricola, I.: The Srní lectures on non-integrable geometries with torsion. Arch. Math. (Brno) **42**, 5–84 (2006). With an appendix by Mario Kassuba
2. Alekseev, A., Meinrenken, E.: The non-commutative Weil algebra. Invent. Math. **139**, 135–172 (2000)
3. Duflo, M.: Opérateurs différentiels bi-invariants sur un groupe de Lie. Ann. Sci. Éc. Norm. Super. **10**, 265–288 (1977)
4. Calaque, D., Rossi, C.A.: Lectures on Duflo Isomorphisms in Lie Algebra and Complex Geometry. EMS Series of Lectures in Mathematics. European Mathematical Society, Zürich (2011)
5. Kostant, B., Sternberg, S.: Symplectic reduction, BRS cohomology, and infinite-dimensional Clifford algebras. Ann. Phys. **176**, 49–113 (1987)
6. Kostant, B.: Clifford algebra analogue of the Hopf-Koszul-Samelson theorem, the ρ-decomposition $C(\mathfrak{g}) = \mathrm{End}V_\rho \otimes C(P)$, and the $\mathfrak{g}$-module structure of $\Lambda\mathfrak{g}$. Adv. Math. **125**, 275–350 (1997)
7. Kostant, B.: A cubic Dirac operator and the emergence of Euler number multiplets of representations for equal rank subgroups. Duke Math. J. **100**, 447–501 (1999)
8. Petracci, E.: Universal representations of Lie algebras by coderivations. Bull. Sci. Math. **127**, 439–465 (2003)

Jahresber Dtsch Math-Ver (2014) 116:73–77
DOI 10.1365/s13291-014-0081-y

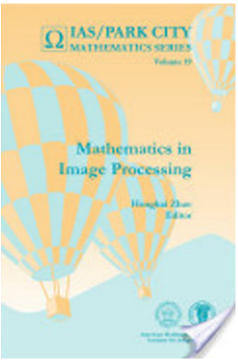
BOOK REVIEW

Hongkai Zhao (Editor): "Mathematics in Image Processing"

IAS/Park City Mathematics Series, Vol. 19. AMS 2013, 245 pp

Gabriele Steidl

Published online: 30 January 2014



Mathematical Signal and Image Processing is an innovative, rapidly developing area of Applied Mathematics which addresses also topics in Computer Science and Engineering. The demand of efficiently processing the overwhelming data from extensive, completely different imaging technologies is continuously growing and provides new scientific challenges. Current research involves various mathematical fields as Linear Algebra, Approximation Theory, Inverse Problems, Applied Harmonic Analysis, Geometric and Topological Techniques, Mathematical Morphology, Partial Differential Equations (PDE), Probabilistic and Statistical Approaches, Variational Methods and Optimization.

The book "Mathematics in Image Processing" edited by H. Zhao can clearly not cover at its 245 pages all the above topics. It focuses on *frame-based image processing*, *sparse and redundant representation modeling of images* and *the simulation of elastic tissues by PDE-FEM methods*. Due to the title of the book we like to point to some books covering subjects in Mathematical Image Processing which are not contained in the present book, namely PDE-based image analysis and calculus of variation [1, 6, 8, 11], morphological image processing [9, 10], probabilistic and statistical methods in imaging [4, 7, 12]. In the following we concentrate on the topics covered by the book of H. Zhao.

G. Steidl (✉)
Department of Mathematics, University of Kaiserslautern, Paul Ehrlich Str. 31, 67653 Kaiserslautern, Germany
e-mail: steidl@mathematik.uni-kl.de

Frame theory dates back to the paper of R.J. Duffin and A.C. Schaeffer 1952. The rich literature in particular on Gabor and wavelet frames shows a wide range of applications including time frequency analysis in signal processing, coherent states in quantum mechanics, filter bank design in electrical engineering, edge detection in image processing. Here we refer to the excellent book on frames [3]. After the introduction of the multiresolution analysis (MRA) concept by S. Mallat (1989) and the systematic construction of compactly supported orthonormal wavelets by I. Daubechies (1988) wavelet analysis became one of the most active research areas in Applied Mathematics. Wavelet frames are overcomplete systems which generalize the concept of orthonormal wavelets and allow more freedom in the representation of data. A systematic construction of wavelet frames was possible due to the unitary extension principle developed by A. Ron and Z. Shen (1997). Wavelet frames, in particular so-called tight frames, fulfill some Parseval-like property which allows the stable decomposition of data as linear combinations of frame elements. The underlying MRA structure guarantees for fast decomposition and reconstruction algorithms and makes wavelet frames well-suited for image processing tasks.

Sparse and redundant representations are closely related to the topic of wavelet frames which also provide sparse representations of certain data. Therefore it is not astonishing that many researchers from wavelet analysis have also contributed to the tremendous development of the "sparsity" field over the last decade. Sparsity expresses the idea that discrete natural images depend only on a number of degrees of freedom which is much smaller than their size. This has lead to a research direction called "compressed sensing" by D. Donoho. In other words the underlying assumption is that the data have a concise representation when expressed in a proper dictionary. Typical dictionaries are Gabor, cosine and wavelet (frame) systems. Task adapted dictionaries can by obtained via dictionary learning and updating. Finding a fixed number of dictionary elements that leads to a "best" approximation is an NP hard problem. Therefore "good" approximation strategies as greedy and relaxation methods must be applied and their approximation quality must be estimated. While the present book focuses on worst case estimates, analysis of average performance has also be done from a probabilistic point of view showing that up to some bounds the desired data may be recovered with a probability close to 1.

Convex analysis and optimization plays a role both in wavelet frame based and sparse and redundant dictionary based modeling. In general these models contain proper, convex and lower semi-continuous functionals which are not smooth. Simple and fast algorithms are required to find minimizers of the functionals. Recently, first order operator splitting methods became very popular in image processing. The origin of these methods goes back to linear operator splitting methods used in linear algebra to solve PDEs (e.g. Douglas-Rachford 1956). More than twenty years later these methods were generalized to inclusion problems and set-valued operators, e.g., by P.L. Lions and B. Mercier 1979. Only recently these methods became very popular in imaging. Pioneering work was done by P. Combettes and co-workers, see [2] and the references therein and by S. Osher and co-workers under the name "Bregman methods". Indeed it appears that in many applications at hand the Douglas-Rachford splitting method applied to the dual problem coincides with the well-known alternating direction method of multipliers and the alternating split Bregman algorithm.

Image processing tasks which can be handled by the above methods are versatile. In the present book they range from basic image restoration tasks as image denoising, (blind) deblurring and inpainting to image decomposition, e.g., into texture and structure, image segmentation and compression.

The book edited by H. Zhao is an outcome of the summer school on "Mathematics in Image Processing" organized at the IAS/Park City Mathematics Institute (PCMI) in 2010. The institute was founded in 1991 as part of the "Regional Geometry Institute" initiative of the National Science Foundation with institutional home at the Institute of Advanced Study (IAS) in Princeton. The annual summer schools focus on different research topics and corresponding Lecture Notes were published each year in the IAS/Park City Mathematical Series. Volume 19 of the series is devoted to mathematical image processing including mathematical theory, computation algorithms and applications. The 2010 summer school was composed of nine lectures given by well-recognized speakers. Three of these lecture of quite different length are contained in this book, namely

- I: Bin Dong and Zuowei Shen: MRA-Based Wavelet Frames and Applications (pp. 7–158),
- II: Michael Elad: Five Lectures on Sparse and Redundant Representations, Modeling of Images (pp. 159–208),
- III: J.M. Teran, J.L. Hellrung, Jr. and J. Hegemann: Simulation of Elasticity, Biomechanics, and Virtual Surgery (with Supplemental Material) (pp. 209–245).

Part I of B. Dong and Z. Shen includes five lectures. The first three lectures give an excellent, self-contained introduction to MRA-based wavelet frames. The lectures start with an approximation theoretic introduction of MRAs. The first lecture contains in particular concise proofs of classical MRA results of C. De Boor, R. deVore and A. Ron (1993/1994). The key point of lecture two is the unitary extension principle (UEP) developed by A. Ron and Z. Shen which allows to construct tight wavelet MRA-frames in a systematic way. The MRA structure guarantees the existence of fast decomposition and reconstruction algorithms. Lecture three deals with a general class of refinable functions, so-called *pseudo-splines*, which include B-splines, Daubechies' orthogonal scaling functions and interpolatory scaling functions. The corresponding analysis is provided. Many additional cross references and annotations are given. In lecture four tight wavelet frames were applied within image restoration and analysis models. The considered restoration tasks cover denoising, deblurring and inpainting which can be considered also under the point of view of inverse problems and image decomposition (e.g., into texture and structure part). More precisely, a convex, in general non-smooth functional containing wavelet frames is adapted to the problem so that a minimizer of the functional approximates the original image. Minimizing such functionals is a topic of convex analysis and optimization. The authors touch first order minimization algorithms from these fields as the forward-backward splitting algorithm and some of its accelerations. They focus mainly on the "Bregman point of view". Lecture five addresses other interesting applications of wavelet frames in optimization models, namely blind deblurring, where the blur kernel has to be estimated too so that the problem is no longer convex, image segmentation which plays an important role in practical image analysis, and scene reconstruction from range data. In the segmentation part the choice of a so-called edge indicator function appears to be crucial. The book chapter contains many important contributions of the authors to the field.

Part II of M. Elad with five lectures gives a concise overview on practical sparse and redundant representation modeling. The first lecture introduces the NP-hard basic "sparsity" problem

$$(P_0) \quad \operatorname{argmin}_\alpha \|\alpha\|_0 \quad \text{subject to } D\alpha = x,$$

where $D \in \mathbb{R}^{m,n}$, $m \ll n$ contains for example the dictionary elements and $\|\alpha\|_0$ counts the number of non-zero entries in α. Its noisy counterpart, where the constraint is replaced by $\|D\alpha - x\|_2 \le \varepsilon$, is denoted by (P_0^ε). The two main approaches to solve the problems approximately are provided, namely greedy algorithms as, e.g., (orthogonal) matching pursuit (OMP), and relaxation methods, e.g., by replacing the ℓ_0-seminorm by the ℓ_1-norm which is known as basis pursuit (BP). Lecture two recalls conditions ensuring that (P_0) has a unique solution and (P_0^ε) a stable one and analyses the performance of OMP and BP. This requires the introduction of the *matrix spark*, *mutual coherence* (D. Donoho and M. Elad 2003) and the *restricted isometry property* (RIP) (E.J. Candés and T. Tao 2005) as a key property. In lecture three several dictionary-based models for general inverse problems, compressed sensing and morphological component analysis, i.e., image decomposition are discussed. The basic "Sparseland" (notation of the author) assumption is that the data can be sparsely expressed as linear combination of some dictionary elements. Lecture four deals with the concrete tasks of denoising, deblurring and inpainting. An important step goes from global to local processing where only image patches are assumed to have a sparse dictionary representation. Finally the question of dictionary learning is addressed. The two basic methods, namely updating the dictionary by the method of optimal directions (MOD) and the K-SVD algorithm of M. Aharon, M. Elad and A.M. Bruckstein (2006) are presented. Finally, lecture five adds further application of the "Sparseland" philosophy: image scale up with a pair of dictionaries and image compression using sparse representation, in particular a facial image compression algorithm.

Part III of J.M. Teran, J.L. Hellrung, Jr. and J. Hegemann consists of three lectures on the simulation of elastic materials characteristics of biological soft tissues. To provide an interactive environment in a surgical simulator this simulation must be *extremely fast* and *very robust*. This is a hard challenge with enormous potential benefit for surgical simulations.

The first lecture gives a tutorial on the basic simulation of elasticity. Initial boundary value problems (IBVPs) for deformable objects are derived from continuum mechanics with an emphasis on those details necessary for finite element method (FEM) type simulation. The following lectures deal with the implementation of numerical solvers for the equations of elasticity based on finite element method (FEM) type discretization of the weak formulation of the PDE. It begins with a review of FEM for solving Poisson's equation, continues with Neo-Hookian elasticity with quasistatic evolution in 1D and extend this to the Neo-Hookian elasticity with backward Euler evolution in 2D. The lectures contain pseudocodes in a syntax similar to Matlab or Octave. Finally supplemental material on special implementation issues (inversion via diagonalization, constitutive models for muscles, positive definiteness of linear systems in Newton iterations) is provided which is a short version of three papers of J.M. Teran and co-workers.

In summary the book gives nice and concise overviews on wavelet-frame based image processing, sparse and redundant representation models and simulation of elastic materials and can be recommended to students and lecturers. The first three lectures of B. Dong and Z. Shen can serve very well as course material on wavelets and wavelet frames with focus on applications in image and signal processing. The "Sparseland" part of M. Elad sketches after a brief review of the basic methodologies various practical applications. For a more extensive coverage of the area the author refers to his book [5]. The final third part is a tutorial on FEM simulation of elasticity and is a good starting point to learn more about the topic.

References

1. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations. Applied Mathematical Sciences, vol. 147. Springer, Berlin (2006)
2. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
3. Christensen, O.: An Introduction to Frames and Riesz Bases. Birkhäuser, Basel (2003)
4. Desolneux, A., Moisan, L., Morel, J.-P.: From Gestalt Theory to Image Analysis. Springer, Berlin (2008)
5. Elad, M.: Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer, Berlin (2010)
6. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, New York (2004)
7. Ohser, J., Schladitz, K.: 3D Images of Materials Structures—Processing and Analysis. Wiley-VCH, New York (2009)
8. Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F.: Variational Methods in Imaging. Springer, Berlin (2009)
9. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press, London (1982)
10. Soille, P.: Morphological Image Analysis. Springer, Berlin (1999)
11. Weickert, J.: Anisotropic Diffusion in Image Processing. Teubner, Stuttgart (1998)
12. Winkler, G.: Image Analysis, Random Fields and Dynamic Monte Carlo Methods. Applications of Mathematics, vol. 27. Springer, Berlin (1995)

Jahresber. Dtsch. Math. Ver. (2014) 116:79–84
DOI 10.1365/s13291-013-0072-4

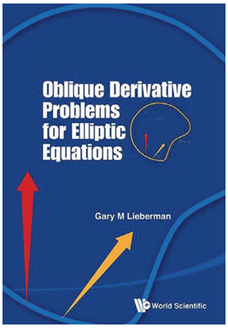
BOOK REVIEW

Gary M. Lieberman: "Oblique Derivative Problems for Elliptic Equations"

World Scientific, 2013, 509 pp

Dian K. Palagachev

Published online: 3 December 2013



Oblique derivative problems are those in which the boundary condition has the form $b(x, u, Du) = 0$ with b depending on the gradient Du of the unknown function u in an appropriate way. The most common particular case is that of Neumann boundary condition

$$\overrightarrow{\gamma} \cdot Du := \frac{\partial u}{\partial \overrightarrow{\gamma}} = \varphi(x) \tag{1}$$

with $\overrightarrow{\gamma}$ standing for the *inner* normal to the boundary $\partial\Omega$ of a given domain Ω where the problem is studied. In the case when the function $b(x, z, p)$ is *linear* with respect to $z \in \mathbb{R}$ and $p \in \mathbb{R}^n$, then the boundary condition is *linear* and has the form

$$\overrightarrow{\beta}(x) \cdot Du + \beta^0(x)u = \varphi(x) \tag{2}$$

with a unit vector field $\overrightarrow{\beta}(x)$ defined on $\partial\Omega$.

The boundary condition (2) is called *oblique* if $\overrightarrow{\beta}(x) \cdot \overrightarrow{\gamma}(x) > 0$ for each $x \in \partial\Omega$ which simply means that the vector field $\overrightarrow{\beta}$ is *strictly transversal* to $\partial\Omega$ and therefore $\overrightarrow{\beta}(x)$ is pointing *inward* Ω at each $x \in \partial\Omega$. The general *nonlinear* boundary condition

$$b(x, u, Du) = 0 \tag{3}$$

is *oblique* if

$$\frac{\partial b}{\partial p}(x, z, p) \cdot \overrightarrow{\gamma}(x) > 0 \qquad \forall (x, z, p) \in \partial\Omega \times \mathbb{R} \times \mathbb{R}^n.$$

D. K. Palagachev (✉)
Bari, Italy
e-mail: palaga@poliba.it

Indeed, these definitions make sense when Ω and $b(x, z, p)$ are smooth enough but the concept of obliqueness can be introduced in a relevant way also in a more general situation. Precisely, the field $\overrightarrow{\beta}$ is *oblique* at a point $x \in \partial\Omega$ if there is a cone of finite height, vertex at x and axis parallel to $\overrightarrow{\beta}$ that lies inside Ω. Further on, the nonlinear boundary condition (3) is *oblique* if the function $\tau \to b(x, z, p + \tau\overrightarrow{\beta})$ is *increasing* for some oblique vector field $\overrightarrow{\beta}$.

Apart from the purely theoretical interest to boundary value problems for PDEs with boundary conditions of the form (2) or (3), these are related to important problems of Mathematical Physics and Applied Mathematics in general. For example, the Neumann condition (1) prescribes the flux along the boundary of an electrostatic potential assigned in a domain Ω. Moving to the theory of Markov processes (e.g. Brownian motion), the first term on the left in (2) corresponds to reflection of the process along $\overrightarrow{\beta}$, while the second one describes absorption phenomena. However, the theory of PDEs with oblique boundary conditions as (2) or (3) goes beyond these classical examples and finds important applications in the mechanics of celestial bodies, stochastic control theory, capillary free surfaces, reflected shocks in transonic flows and so on.

Gary Lieberman's book under review deals with oblique derivative problems for second order elliptic equations. The author provides a systematic and comprehensive exposition of the theory developed over the last three decades, thanks also to his own ideas and results. As for spirit, concept and performance, the book is similar to the classics by Ladyzhenskaya and Ural'tseva [2] and Gilbarg and Trudinger [1] which mainly deal with Dirichlet problems for elliptic PDEs. Although the theory of oblique derivative problems is more technical, the author has fully reached his goal to put it in the same systematic shape as done in [1,2] for the Dirichlet problem. The reader will find that the oblique derivative problems possess many fascinating properties in common with the Dirichlet problems. Actually, these are regular problems obeying the elliptic regularization property, that is, the solution gains two derivatives from the right-hand side of the equation and one derivative from the boundary data. Optimal *a priori estimates* in various functional scales not only imply that regularizing property but allow also to built relevant solvability theories in classical, weak, strong or viscosity settings for the oblique derivative problems and to show these are of Fredholm type.

The hypothesis of *obliqueness* is crucial in order to have all these results. Consider for a moment a more general situation when the product $\overrightarrow{\beta}(x) \cdot \overrightarrow{\gamma}(x)$ is *non-negative* or even *changes its sign* on $\partial\Omega$ so that the boundary condition is no more oblique in the above sense and, therefore, there exists a non-empty subset of the boundary where the field $\overrightarrow{\beta}$ becomes *tangential* to $\partial\Omega$. It was Poincaré the first to arrive at a boundary value problem with such a *tangential-oblique* condition when studying the tides on the Earth (cf. [5]) and what is interesting is that the theory of the *tangential* oblique derivative problems has various features not present neither in the oblique case studied in the book under review, nor in the Dirichlet case (see [4] and [6]). Precisely, the corresponding solution need not be as smooth as the data of the problem would suggest; that *loss of regularity* depends on the order of contact between $\overrightarrow{\beta}$ and $\partial\Omega$;

the index of the problem need not be zero or even finite in the case of more than two dimensions.

Even if Lieberman follows [1,2] and his previous book [3] as models, his personal choice in the exposition is not to duplicate their contents but rather to emphasize on ideas and elements which are not immediate consequences of the corresponding theory of Dirichlet problems. The author provides an alternative development of classical theories of elliptic PDEs seen from modern perspective. Various facts and properties from functional analysis and measure theory available in [1–3] are only reported on, while the author includes detailed proofs of results from real and functional analysis whenever they are not in [2,3] or in standard graduate texts on the topic. This way the readers, already familiar with the classical texts [1,2], will have no difficulties to follow the exposition. Notes at the end of each chapter give the history of the topic treated and suggestions for further reading, while the problems attached to the chapters will be useful exercises for the readers.

Moving on to the details, the book under review consists of 12 chapters, a relatively comprehensive bibliography and a subject index.

The exposition starts with the study of *general linear* oblique derivative problems of the form

$$\begin{aligned} a^{ij}(x)D_{ij}u + b^i(x)D_i u + c(x)u &= f(x) \quad \text{in } \Omega, \\ \overrightarrow{\beta}(x)\cdot Du + \beta^0(x)u &= \varphi(x) \quad \text{on } \partial\Omega \end{aligned}$$

and Chapter 1 deals with the basic pointwise estimates for solutions of these problems, the analytic heart of which is the *maximum principle.* This is a very powerful tool which allows, in its classical version, to estimate the solution of a given boundary value problem in terms of simple properties of the coefficients of the differential operators involved and the maximum of the nonhomogeneous terms of the equation and boundary condition. Although quite straightforward for the Dirichlet problem, the maximum principle needs more attention in the case of oblique boundary conditions. It is here the point where the author defines the basic notion of *obliqueness* in the cases of domains with smooth or Lipschitz continuous boundaries, and proposes a detailed study of the classical maximum principle in the situation considered. The reader will find also a maximum principle of Aleksandrov–Bakel'man–Pucci type which estimates the maximum of the solution in terms of the Lebesgue L^n-norm of the right-hand side $f(x)$ of the equation with n being the number of the spatial dimensions. As consequence, a *modulus of continuity* estimate follows for the solution, which is a starting point for the following study of a large variety of nonlinear problems.

The classical Schauder theory of linear elliptic equations is presented in Chapter 2. This is a key part of the theory of the Dirichlet problem which has two elements:

1) It provides estimates for the Hölder norms of any solution of the problem considered and for the first and second derivatives of this solution;
2) it gives existence results under appropriate conditions on the coefficients of the elliptic operator.

An excellent description of the Schauder approach, based on *potential theory*, is given in Chapter 6 of the book [1] of Gilbarg and Trudinger. Here the author proposes a *modern perspective* to Schauder's theory, using the *integral characterization* of the Hölder spaces first introduced by S. Campanato and later improved by Caffarelli and Safonov. A particular advantage of this approach is the possibility to treat oblique derivative problems over domains with less regular boundaries. As a groundwork for the estimates in the later chapters, the author considers also mixed boundary value problems in which Dirichlet data are prescribed on a part of the boundary and an oblique derivative condition is given on the remainder. Estimates of Hölder norms of solutions to such problems are derived and it is proved that these are Fredholm problems with index zero.

Chapter 3 is devoted to the construction of various supersolutions for both the Dirichlet and oblique derivative problems in domains with Lipschitz continuous boundaries. Involving solutions of singular ordinary differential equations, the author employs the ideas of K. Miller to construct suitable comparison functions in a cone.

Hölder estimates for the first and second derivatives of solutions to linear oblique derivative problems are given in Chapter 4. It is shown in particular that the solutions to linear oblique derivative problems are typically smoother than the underlying domain. Thus, solutions in Lipschitz domains have Hölder continuous first derivatives, that is, these are $C^{1,\alpha}$ smooth with some $\alpha > 0$ determined by the domain and the differential equation but not by the boundary condition, while the solutions are $C^{2,\alpha}$ smooth in $C^{1,\alpha}$-domains.

The theory of *weak solutions* to the *conormal* derivative problem for divergence form, linear elliptic equations is elaborated in Chapter 5. The problem under consideration now is

$$\begin{aligned} D_i\left(a^{ij}(x)D_ju + b^i(x)u\right) + c^i(x)D_iu + c^0(x)u &= D_if^i(x) + g(x) \quad \text{in } \Omega,\\ a^{ij}(x)\gamma_i(x)D_ju + \overrightarrow{b}\cdot\overrightarrow{\gamma}\,u + \beta^0(x)u &= \varphi(x) + \overrightarrow{f}\cdot\overrightarrow{\gamma} \quad \text{on } \partial\Omega, \end{aligned}$$

where $\overrightarrow{\gamma} = (\gamma_1, \ldots, \gamma_n)$ is the *inner* normal to $\partial\Omega$ as above, while $\overrightarrow{b}$ and $\overrightarrow{f}$ stand for the vector fields $\left[b^1(x), \ldots, b^n(x)\right]$ and $\left[f^1(x), \ldots, f^n(x)\right]$, respectively. As in the case of the Dirichlet problem, one multiplies the equation and the boundary condition by a suitable test function and then, thanks to the special divergence structure of the differential operators involved, integrates formally by parts. This leads to an equivalent, *integral* formulation of the problem considered and thus to the natural concept of *weak solution* which is only required to lie in the *Sobolev space* $W^{1,2}(\Omega)$, that is, to have first derivatives in L^2 in contrast to the classical solutions of general oblique derivatives problems which have to be at least C^2-smooth inside Ω. This way, the corresponding *weak* solvability follows almost for free from the powerful results of linear functional analysis. Apart from weak existence results, the author studies in this chapter finer properties of weak solutions such as their higher Sobolev regularity, global boundedness, local maximum principles and so on. Indeed, under suitable hypotheses on the data, a De Giorgi type result holds true for weak solutions to the conormal problem: Indeed, these are *Hölder continuous*. The author derives also

Hölder estimates for derivatives of weak solutions under stronger assumptions on the coefficients than just boundedness.

Strong solutions of general linear elliptic equations represent an intermediate situation between the concepts of *weak* solutions which need to be only once weakly differentiable and that of solutions in *classical* sense which must be at least twice continuously differentiable. Moreover, the notion of weak solution depends on the particular *divergence form* structure of the elliptic equation studied. A *strong solution* of an oblique derivative problem is a *twice weakly differentiable* function which satisfies the given *general form* linear elliptic equation *almost everywhere* in Ω and for which the oblique derivative boundary condition holds in the sense of trace or in classical sense on $\partial\Omega$. The *"linear"* part of Lieberman's book closes with Chapter 6 where the theory of strong solutions is developed. The main analytic tools employed here are the Aleksandrov–Bakel'man–Pucci maximum principle, the Hardy Littlewood maximal function and a Vitali type covering lemma. As a result, the author builds a regularity and strong solvability theory for the linear oblique derivative problem similar to that of Calderón and Zygmund for the Dirichlet problem.

In the rest of the book Gary Lieberman exposes the *nonlinear* theory of the oblique derivative problems. Chapter 7 provides a brief introduction to *viscosity solutions* of fully nonlinear elliptic equations with fully nonlinear oblique boundary conditions

$$F(x, u, Du, D^2u) = 0 \quad \text{in } \Omega, \qquad b(x, u, Du) = 0 \quad \text{on } \partial\Omega. \tag{4}$$

The machinery here relies on comparison principles for viscosity sub- and supersolutions, suitable test functions and the Perron process which lead to existence and uniqueness results. Although of auxiliary character, the full proof of the deep result by Aleksandrov that every convex function is twice differentiable almost everywhere, will be of particular interest for the readers.

Chapter 8 deals with various pointwise bounds (maximum estimates, Harnack inequalities, Hölder estimates) for solutions of *quasilinear* oblique derivative problems

$$\mathcal{Q}u = 0 \quad \text{in } \Omega, \qquad \mathcal{N}u = 0 \quad \text{on } \partial\Omega$$

with a general elliptic operator $\mathcal{Q}u := a^{ij}(x, u, Du)D_{ij}u + a(x, u, Du)$ and $\mathcal{N}u := b(x, u, Du)$, or with a *divergence form* operator $\mathcal{Q}u := \operatorname{div} A(x, u, Du) + B(x, u, Du)$ and a *conormal* boundary condition $\mathcal{N}u := A(x, u, Du) \cdot \overrightarrow{\gamma} + \psi(x, u)$.

Gradient estimates for solutions to general form *quasilinear* oblique derivative problems

$$\mathcal{Q}u := a^{ij}(x, u, Du)D_{ij}u + a(x, u, Du) \quad \text{in } \Omega, \qquad b(x, u, Du) = 0 \quad \text{on } \partial\Omega$$

are derived in Chapter 9 for special type decomposition of the principal coefficients of the operator $\mathcal{Q}$ including both the false mean curvature equation and capillary-type problems, while Chapter 10 deals with similar estimates for the quasilinear conormal derivative problem.

In Chapter 11, the author derives Hölder estimates for the first and second derivatives of the solutions to quasilinear oblique derivative problems in terms of the maximum modulus of the gradient. These are further combined with the gradient estimates from Chapters 9 and 10 to prove a number of existence results through the Leray–Schauder fixed point principle. Special attention is paid to the particular examples considered in the previous Chapters, such as the capillary problem, the uniformly elliptic conormal problem and the false mean curvature problem.

The final Chapter 12 deals with the *fully nonlinear* oblique derivative problem (4). The new element here, in comparison with the quasilinear case, is the necessity to derive Hölder estimates for the second derivatives of the solution. These are obtained in the both cases of smooth and low regular functions involved, after that the author proves selected existence theorems.

In conclusion, the book of Gary Lieberman is a comprehensive presentation of the theory of oblique derivative problems for elliptic equations and, without any doubts, makes an important contribution to the general theory of boundary value problems for elliptic PDEs. Although addressed mainly to post-graduate students, it is accessible and will be useful to a broad spectrum of professional mathematicians working in the areas of partial differential equations, operator theory, stochastic processes, etc.

References

1. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Classics in Mathematics. Springer-Verlag, Berlin (2001). (Reprint of the 1998 edn.)
2. Ladyzhenskaya, O.A., Ural'tseva, N.N.: Linear and Quasilinear Elliptic Equations. Mathematics in Science and Engineering, vol. 46. Academic Press, New York (1968). Translated from the Russian by Scripta Technica, Inc. Translation editor: Leon Ehrenpreis
3. Lieberman, G.M.: Second Order Parabolic Differential Equations. World Scientific, Singapore (1996)
4. Paneah, B.P.: The Oblique Derivative Problem. The Poincaré Problem. Mathematical Topics, vol. 17. Wiley-VCH-Verlag, Berlin (2000)
5. Poincaré, H.: Leçons de mécanique céleste professées a la Sorbonne. Tome III: Théorie des Marées. Gauthier-Villars, Paris (1910)
6. Popivanov, P.R., Palagachev, D.K.: The Degenerate Oblique Derivative Problem for Elliptic and Parabolic Equations. Mathematical Research, vol. 93. Akademie-Verlag, Berlin (1997)