Classical and quantum algorithms for algebraic problems

Thesis for the degree
"Doctor of the Hungarian Academy of Sciences"

Gábor Ivanyos
Computer and Automation Research Institue
of the
Hungarian Academy of Sciences

2007

Acknowledgments

I am grateful to my scientific collaborators, in particular to Alexander Chistov, Kati Friedl, Marek Karpinski, Klaus Lux, Fred Magniez, Miklos Santha and Pranab Shen who were coauthors of the papers this thesis is based on. I am also indebted to Arjeh Cohen, Jan Draisma, Willem de Graaf, Lajos Rónyai and Csaba Schneider for invaluable discussions inspiring research behind this work.

My thanks also go to my colleagues at the Computer and Automation Research Institute of the Hungarian Academy of Sciences, in particular to the collectives led by Lajos Rónyai and János Demetrovics for ensuring support and the excellent working atmosphere making my research possible. I am also indebted to my friends in the Discrete Mathematics group at the Department of Mathematics at Eindhoven University of Technology as well as those in the group ”Algorithms and Complexity” of the Laboratoire de Recherche en Informatique of CNRS and Université de Paris-Sud for the warm and helpful hospitality during my visits. My research has also benefited from teaching at the Institute for Mathematics at the Budapest University of Technology and Economics and from collaborating with the team at the Department of Algebra therein.

I acknowledge supplementary financial support provided by several grants of the Hungarian Research Fund (OTKA) and of the Dutch Science Organization (NWO), by IST-FET grants of the EC and by the Bolyai Fellowship of the Hungarian Academy of Sciences.

Finally, but not least I would like to thank my family, for their sacrifice and the warm, loving atmosphere they provided.

Contents

1 Introduction 1

2 Preliminaries 7

2.1 Fields, matrices and polynomials 7

2.2 Algebras 7

2.2.1 Structure of algebras 8

2.2.2 Extending scalars 8

2.2.3 Idempotents and the primary decomposition 9

2.2.4 Separability and the Wedderburn–Malcev theorem 9

2.2.5 Tori 10

2.3 Polycyclic presentations of finite solvable groups 10

2.4 Randomized algorithms 11

2.5 Quantum computing 12

2.5.1 Quantum circuits 13

2.5.2 Cleaning up 14

2.5.3 Classical computation as quantum computation 15

2.5.4 Numerical vs. probabilistic errors 16

2.5.5 State sampling 16

2.5.6 The hidden subgroup problem 18

2.5.7 The quantum Fourier transform 18

3 Finiteness of matrix semigroups over function fields over finite fields 20

4 Finding the radical of matrix algebras using Fitting decompositions 23

4.1 Tori and maximal tori 25

4.1.1 Centralizers of tori and Fitting decompositions 27

4.2 Decomposition with respect to a maximal torus 28

4.3 A reduction to the commutative case 31

4.4 Computing Fitting decomposition with respect to a semisimple matrix . . 32

4.5 A Monte Carlo method for finding the radical 34

4.5.1 Jordan decomposition 34

4.5.2 Finding a maximal torus 34

4.5.3 Calculating C 35

4.5.4 Generating elements of N 37

4.5.5 Generating elements of $\mathrm{Rad}(H)$ 37

4.5.6 Computing $\mathrm{Rad}(A)$ 39

4.6 Remarks 39

5 Treating the exceptional cases of the MeatAxe 40

5.1 The exceptional algebras . . . 41

5.2 The algorithm . . . 44

5.3 Probability of success . . . 45

6 Computing the radical of matrix algebras over finite fields 48

6.1 Wedderburn complements in local algebras . . . 51

6.2 Using a complete system of primitive idempotents . . . 54

6.3 Verifying correctness . . . 58

7 Constructing module isomorphisms 61

7.1 Finding free submodules over semisimple algebras . . . 62

7.2 Finding a single generator . . . 64

7.3 The general conjugacy problem . . . 65

7.4 Remarks . . . 65

8 Deciding universality of quantum gates 66

8.1 Completeness . . . 67

8.2 Universality . . . 69

8.2.1 The ideal of a gate set . . . 70

8.2.2 The proof of Theorem 8.1 . . . 70

8.3 Remarks . . . 71

9 A quantum algorithm for finding hidden subgroups in a class of solvable groups 73

9.1 Abelian quantum stabilizer . . . 77

9.2 Constructive orbit membership test in abelian groups . . . 80

9.3 Solving random systems of linear disequations . . . 82

9.3.1 Reductions . . . 85

9.3.2 An algorithm for p-groups . . . 87

9.4 A recursion for computing stabilizers . . . 90

9.5 Remarks . . . 92

10 Efficient Testing of Groups 94

10.1 Approximate group homomorphisms . . . 97

10.2 Growing subgroups . . . 101

10.3 The tester . . . 103

Bibliography 105

Chapter 1

Introduction

In this thesis we present some results regarding algorithmic aspects of certain algebraic problems. A substantial part of the problems concerns computations in matrix algebras and modules while the other major part addresses efficient quantum algorithms and related probabilistic methods for problems from group theory. In this chapter we give a brief and somewhat informal summary of the most important results presented in the thesis. Chapter 2 is devoted to the definitions, basic facts, techniques, computational models used later. For convenience of readers not familiar with quantum computing we give a rather detailed description of a simple model of quantum computation.

The first group of the results presented in this thesis concerns algorithmic problems in (associative) matrix algebras. Algorithms for matrix algebras and modules play an important role in several branches of computational mathematics, including computational (modular) representation theory of finite groups, computing with finite or infinite Lie algebras and also some computational aspects of differential algebra. L. E. Dickson [25] already in 1923 proved a theorem which characterizes the Jacobson radical in a computational flavor. The first systematic collection of *polynomial time* methods for finite dimensional associative algebras can be found in the paper [38] by K. Friedl and L. Rónyai from 1983. Since then the collection has grown substantially, now polynomial time algorithms are known for several problems regarding the structure of matrix algebras over various ground fields.

One of the important structural invariants of a matrix algebra is its Jacobson radical, the largest nilpotent ideal contained. Algorithms based on solving systems of (semi-)linear equations arising from extensions of Dickson's characterization to positive characteristic have been proposed over various families of ground fields. See Rónyai's method [85] and the slightly more efficient algorithm of W. Eberly [29] over finite fields, the method in [60] over function fields and finally the procedure given in [21] which works over a wide class of fields of positive characteristic. In the brief Chapter 3 we give an application of computation of the radical to deciding finiteness of certain matrix (semi-)groups. We show the following.

- There is a deterministic polynomial time algorithm which decides finiteness of a matrix semigroup generated by a set of matrices with entries from a function field with constant number of variables over a finite field, see Corollary 3.4.

To begin the description of the contents of Chapter 4, we note that the method presented in [21] ultimately relies on an assumption which is, intuitively, similar to that pth roots can be efficiently taken where p is the characteristic. Taking roots is not possible using merely the field operations, actually existence of roots is an undecidable problem over general fields. Based on this observation W. Eberly showed in [28] that there is no general algorithm based on merely the four field operations which determines the Jacobson radical of a commutative matrix algebra over a general field of positive characteristic. He also conjectured that there are no more obstacles in the noncommutative case. Although the algorithm presented in [21] settles this conjecture in affirmative, a strong version of it – namely, *polynomial time* reducibility to a *unique* instance of computing the radical of a commutative algebra – was one of the sources of motivation for investigating alternative approaches to computing the radical.

Simultaneously W. A. de Graaf was developing an algorithm for computing the solvable radical of a Lie algebra of characteristic zero, see [44]. The method was based on computing certain subalgebras (so called Cartan subalgebras) first; and in practice his algorithm outperformed former methods which were based on applications of Dickson's theorem to certain related associative algebras. This drew our attention to considering methods which use certain commutative semisimple subalgebras (maximal tori) and techniques similar to weight decompositions. Note that the centralizers of these subalgebras are the associative counterparts of Cartan subalgebras. The first result using the new approach gives the desired reduction:

- Computing the Jacobson radical of a finite dimensional associative algebra A can be reduced to computing the radical of a subfactor of A, see Theorem 4.1.

The input is a set of matrices which generate the algebra and the output is a set of matrices which generate the radical as an ideal. Note however, that linear bases from such generating sets can be computed in polynomial time. By a subfactor we mean a factor of a subalgebra of the original algebra. We also remark that the reduction requires only a polynomial number of field operations. To give an example of algebraic theorems that support algorithms in this thesis, we also mention that the structure theorems behind the reduction (Proposition 4.9 and Theorem 4.11) state that the radical of an algebra A can be written as the sum of the ideal of A generated by the radical of the centralizer of a maximal torus T and the commutator subspace $[A, C]$ of A with certain subalgebra C of T. The sum is direct sum of vector spaces. The subalgebra C consists of the elements of T which are central modulo $\mathrm{Rad}(A)$. In this introduction we shall refer to C as the semi-central part of T and the subspace $[A, C]$ as the "commutator part" of the radical. Of course, for computing $\mathrm{Rad}(A)$ one needs an alternative characterization of the semi-central part. For algebras over general ground fields regarded in Chapter 4, this is given in Theorem 4.12.

Following the lines of the reduction algorithm discussed above, we developed a randomized algorithm for computing the radical of a matrix algebra over a perfect field. We changed the model of the input, namely we assumed that besides having the algebra generators for A, we are supported by an oracle for drawing "sufficiently random" elements of the matrix algebra A. Here randomness is understood in an algebraic sense: zeros of polynomial functions of moderate degree on A are assumed to be avoided with a good chance. Such an oracle can be easily implemented if a linear basis of A is given. Also, there are heuristic methods for producing random elements of algebras given by generators, like the one used in the MeatAxe [52]. Note that in the assumption on algebraic randomness it is implicit that the ground field is sufficiently large. The output (which, in previous algorithms was a linear basis of $\mathrm{Rad}(A)$) is a set which generate $\mathrm{Rad}(A)$ as an ideal of A. We stress again that from such a set one can produce a linear basis in polynomial time.

However in certain applications it is sufficient to have an ideal generating set. Below is a rough description of the result for the important case where number of generators is constant.

- Assume that K is a perfect field and $A \leq M_n(K)$ is given by a m generators and an oracle for producing "random" elements and m is constant. Then matrices which generate $\mathrm{Rad}(A)$ with high probability can be found in time roughly $O(n^4)$ by a randomized algorithm of Monte Carlo type. See Theorem 4.3 for a more precise statement of the result.

Recall that a Monte Carlo type randomized algorithm may return a wrong answer with probability which can be made exponentially small by independent repetitions. In the rough complexity estimate $O(n^4)$ we ignored multiplicative factors of polylogarithmic order and additive terms involving the cost of drawing several "random" elements of A as well as terms involving the complexity of computing the squarefree part of polynomials over K of degree n. The complexity of the latter task can be interpreted as how effective is the perfectness of the field K. However, over finite fields or fields of characteristic zero the number of field operations required for computing the squarefree part of a polynomial is nearly linear.

Unlike the "algebraic randomness" assumption above, in the context of algebras over finite fields – including both the radical computation algorithm and the analysis of the MeatAxe discussed below – we assume ability of drawing uniformly random elements of the algebra.

We observed that for purposes of module problems, like the principal subtask of the MeatAxe, even exhibiting a single nontrivial element of $\mathrm{Rad}(A)$ is often sufficient. The MeatAxe is a widely used collection of procedures performing various tasks in modules for finite dimensional algebras over finite fields. The most important subtask is finding a nontrivial submodule if exists. The original approach of R. Parker [79] performed well in practice over small ground fields. D. F. Holt and S. Rees developed a randomized extension [52] which worked efficiently over large ground fields as well, except in certain special classes of algebras. We noticed that the bad cases were closely related to the "commutator part" of the radical mentioned in the structure theorems above. Furthermore, in these bad cases a good replacement for the semi-central part of a maximal torus is available with high probability in the form of a primitive idempotent. This resulted in the following (see Chapter 5).

- There is a simple extension of the Holt-Rees MeatAxe procedure which – without essential loss in speed – finds a submodule with high probability even in the exceptional cases. See Section 5.2 for the description of the extension and Proposition 5.4 for a lower bound on success probability.

Together with the extension, with probability larger than a positive constant, the MeatAxe either finds a submodule or finds a proof of irreducibility. Therefore it can be considered as a Las Vegas type method, which never returns a wrong answer but may fail with a probability which can be made arbitrarily small by repetitions. Our extension has been first implemented in the C-MeatAxe package by M. Ringe and later in the computer algebra systems GAP [39] (by A. Hulpke) and MAGMA [18] (by J. Cannon and C. Leedham-Green).

Returning to the probabilistic algorithm for finding the radical of a matrix algebra, we saw that the bottleneck of the method was determining the semi-central part of the maximal torus. Independently W. Eberly and M. Giesbrecht developed a fast randomized algorithm for computing the simple components of a semisimple matrix algebra over a finite field [31]. They considered an input model similar to the one used in the MeatAxe: assumed access to random elements of the algebra. They posed the question whether there are methods of the same complexity for certain tasks in non-semisimple algebras. The key ingredient of the method in [31] is building a maximal *split* torus, or, equivalently, a complete system of primitive idempotents. It turns out that for such a torus the semi-central part can be found quickly. The method we present in Chapter 6 does substantially more than computing the radical: it can also be used to construct a Wedderburn complement of the algebra A. Recall that a Wedderburn complement is a subalgebra isomorphic to the factor $A/\mathrm{Rad}(A)$. For a constant number of generators the result sounds as follows.

- Assume that K is a finite field and $A \leq M_n(K)$ is given by m generators and an oracle for producing "random" elements and m is constant. Then a collection of matrices which generate a Wedderburn complement in A and a set of matrices which generate $\mathrm{Rad}(A)$ as an ideal can be found in time roughly $O(n^3)$ by a randomized algorithm of Las Vegas type. See Corollary 6.5 for a more precise statement of the result.

We remark that the complexity of a Monte Carlo version of the method (that is, a version where correctness is not tested) is actually less than $O(n^3)$, it is roughly proportional to the cost of a matrix multiplication. This is also the complexity of the algorithm of Eberly and Giesbrecht in the semisimple case.

We conclude the part concerning computational representation theory with a brief chapter on a deterministic polynomial time solution to a certain simple task. In Chapter 7 we address the problem of finding explicit isomorphisms between modules. The task (at least over sufficiently large ground fields) admits a straightforward efficient randomized solution by the Schwartz-Zippel lemma (see Section 2.4). We show the following.

- Assume that K is a field admitting a deterministic polynomial time method for computing the Jacobson radical of finite dimensional algebras over K. Then there is a *deterministic* polynomial time algorithm for deciding whether two finite dimensional modules over an algebra are isomorphic and for computing explicit isomorphism between isomorphic modules (see Corollary 7.3).

Chapter 8 connects the part related to computational representation theory to the part whose main topic is quantum computing. The problem addressed there is actually related to the physical realization of quantum computers.

Quantum circuits are built from so-called quantum gates. An n-qubit quantum gate is an unitary transformation of the complex Euclidean space $\mathbb{C}^{2^n}$ capturing the possible states of n qubits. For $N \geq n$ there are $N(N-1)\cdots(N-n+1)$ ways to wire an n-qubit gate to a system consisting of N qubits. A wired n-qubit gate acts on the space $\mathbb{C}^{2^N}$ corresponding to the possible states of the N-qubit system as the tensor product of the unitary operation acting on the 2^n-dimensional space of the selected n qubits with the identity on the 2^{N-n}-dimensional space corresponding to the rest of the qubits. A circuit on an N-qubit system built from a fixed set Γ of gates is just a sequence of wired elements from Γ. The operation implemented by the circuit is just the product of the unitary transformations corresponding to the members.

If experimental physicists come up with realization of a specific set of gates it is natural to ask how powerful circuits can be built from the collection. An n-qubit gate set Γ is said to be N*-universal* if every unitary transformation of the space $\mathbb{C}^{2^N}$ can be approximated with arbitrary precision by a circuit built from the elements of Γ. Mathematically, the $N(N_1)\cdots(N-n+1)|\Gamma|$ unitary transformations corresponding to the wired gates should generate a dense subgroup of the whole unitary group U_{2^N}. (More accurately, as scalar multiples of a vector represent the same quantum state, density must be understood projectively, i.e., modulo scalar matrices.) If $n > 1$ then N-universality of a fixed gate set is monotone in N: for $N' > N \geq n$, if Γ is N-universal then it is N'-universal as well. Based on this, we say that a set Γ of n-qubit gates is universal if it is N-universal form some $N \geq n$. This notion expresses certain *ultimate usefulness* of the gate set Γ.

In Chapter 8 we use a combination of recent result from representation theory of finite groups and bounds on commutative algebra to show that universality of gate sets is algorithmically decidable. Actually, N-universality for a fixed N can be decided using the Zariski closure algorithm of H. Derksen, E. Jeandel and P. Koiran. Unless there is an effective bound on the smallest N such that a universal n-qubit gate set is already N-universal, decidability of the weaker notion does not follow immediately. However, we can show the following.

- If an n-qubit gate set is universal then it is already N-universal for some $N \leq 255n$, see Theorem 8.1. As a consequence, universality is an algorithmically decidable property.

It turns out that N-universality can be tested by solving a system of $m \cdot 2^{O(N)}$ homogeneous linear equations in $2^{O(N)}$ variables where m is the number of gates in the collection. Note that if the input is given as an array consisting of all the $m \cdot 2^{2n}$ entries of the 2^n by 2^n matrices, then for $N = 255n$, the quantity $m \cdot 2^{O(N)}$ is still polynomial in the input size.

Chapter 9 is devoted to a polynomial time solution of the hidden subgroup problem in a class of solvable groups. The hidden subgroup paradigm generalizes computing multiplicative orders of numbers modulo composite numbers as well as the discrete logarithm problem in various groups. P. Shor's polynomial time solutions to these two problems [90] are the most remarkable achievements in the history of quantum algorithms. Shor's method generalizes to a polynomial time solution (in the logarithm of the group size) if the hidden subgroup problem over finite abelian groups. Extensions to noncommutative groups are subject of active research. We show the following.

- The hidden subgroup problem can be solved in polynomial time over solvable groups of constant derived length whose commutator subgroups have constant exponent. See Theorem 9.1 and Corollary 9.2 for more precise statements.

We remark that in 2003 when our paper [36] was published the class above included almost all cases of groups for which polynomial time hidden subgroup algorithms were known. Currently the most important groups with polynomial time hidden subgroup algorithms outside this class are certain nilpotent groups of larger exponent having derived length 2, see [8, 61, 62].

We conclude this thesis with Chapter 10. It can be considered as a quantum vs. classical counterpart of Chapter 7, where we gave a deterministic polynomial time algorithm solving a problem which had an easy randomized solution. In Chapter 10 we consider testing multiplication tables of abelian groups. There is a relatively easy quantum algorithm solving this problem in time polynomial in the table size. We substitute the power of quantum computers by the assumption that a multiple of the exponent of the group is given. We show the following.

- Given a table for a binary operation, it can be tested in time polynomial in the *logarithm* of the size of the table whether the table corresponds to an abelian group whose exponent is a divisor of a given number. The test always accepts tables corresponding to such groups and rejects tables ”far away” from such group multiplication tables with high probability. See Theorem 10.1 for a more precise statement.

We remark that the best previously known methods for related problems were slightly sublinear in the table size.

Chapter 2

Preliminaries

This chapter is devoted to fixing notation and terminology used throughout the thesis and to briefly recalling basic definitions and facts related to the problems addressed later on. The main computational models we work with are also discussed here.

2.1 Fields, matrices and polynomials

We assume that the reader is familiar with the basic notions and facts regarding fields and vector spaces over various fields. In this part we give a brief overview of the general computational model for fields we use throughout the thesis. We also discuss complexity of basic linear algebra tasks and introduce related notation.

In the most general computational model for fields it is merely assumed that the ground field admits effective procedures for the field operations and equality tests. Alternatively, one can assume that we are equipped with oracles (”black boxes”) for performing these tasks. The complexity of an algorithm is measured by the number of operations and equality tests required by the algorithm in the worst case. Elementary tasks of linear algebra (matrix multiplication, computing determinants, solving systems of linear equations, etc.) admit efficient solutions in this model, cf. [14]. Let $MM(n) := MM_K(n)$ be a function on n such that $MM(n)$ arithmetical operations are sufficient to calculate the product of two n by n matrices over K. We assume that $MM(n) \geq n^2$. The standard method shows that one can take $MM(n) = O(n^3)$. Using the asymptotically fastest known (but not very practical) multiplication algorithm one achieves $MM(n) = O(n^{2.376})$. The complexity of all the linear algebra tasks mentioned above is $O(MM(n))$.

2.2 Algebras

In this section we give some definitions and basic facts concerning the structure of associative algebras. We assume that the reader is familiar with the basic ring theoretic notions for associative algebras over fields (subalgebras, homomorphisms, ideals, factor algebras, nilpotency, modules, direct sums, tensor products, etc.). Throughout the thesis by an algebra we understand a finite dimensional associative algebra over the field K. Unless otherwise stated, we also assume that the algebra has an identity denoted by 1_A (if the algebra is A) or briefly by 1. Modules are assumed to be finite dimensional unital left A-modules. (The A-module U is called unital if $1_A u = u$ for every $u \in U$.) Let A be an algebra and let M be an A-module (which can be A itself). For subsets $B \subseteq A$ and $C \subseteq M$

we denote be BC the K-linear span of $\{bc|b \in B,\ c \in C\}$. For $b, c \in A$ we denote by $[b, c]$ the additive commutator $bc - cb$ of b and c. For $B, C \subseteq A$ we use the notation $[B, C]$ for the linear span of $\{[b, c]|b \in B,\ c \in C\}$. For a subset $B \subseteq A$, $C_A(B)$ stands for the centralizer of B in A: $C_A(B) = \{x \in A|[x, b] = 0 \text{ for every } b \in B\}$. The center $C_A(A)$ of A is denoted by $Z(A)$. If K is a field then the algebra of n by n matrices with entries from K is denoted by $M_n(K)$. By a matrix algebra over K we mean a subalgebra of $M_n(K)$ containing the identity matrix for some integer n.

2.2.1 Structure of algebras

We recommend the reader familiar with the basic structure theory of algebras to skip this part. Here we briefly recall Wedderburn's theorems on the structure of algebras. In every finite dimensional algebra A there exists a largest nilpotent ideal $\mathrm{Rad}(A)$, called the Jacobson radical (or just radical) of A. A is called semisimple if $\mathrm{Rad}(A) = (0)$. The factor algebra $A/\mathrm{Rad}(A)$ of an arbitrary algebra is semisimple. A is called simple if A contains no proper nonzero ideals. A semisimple algebra A can be decomposed into the direct sum of its minimal ideals $A_1, \ldots, A_r$. We refer to the simple algebras A_i as the simple components of A. A simple algebra A is isomorphic to the algebra $M_d(D)$ of d by d matrices with entries from D, where D is a division algebra (or skew field) over A. By this we mean that D contains no zero divisors. If Z is a subfield of $Z(A)$ containing the identity of A then it is possible (and often convenient) to consider A as an algebra over Z. A is called central over K if $Z(A) = K$ (more precisely, $Z(A) = K1_A$). The dimension of a central simple K-algebra is always a square.

A module U over the semisimple algebra A can be decomposed as a direct sum of simple A-modules (modules with no proper nonzero submodules). If A is a simple algebra then there is only one isomorphism class of simple A-modules. By A^{op} we denote the algebra opposite to A. A^{op} has the same vector space structure as A but the multiplication is reversed. A can be considered as an $A \otimes_K A^{op}$-module by the multiplication law $(a \otimes b)c = acb$. The ideal structure of A coincides with the $A \otimes_K A^{op}$-submodule structure of A. If A is a central simple K-algebra then $A \otimes_K A^{op} \cong M_{d^2}(K)$ (where $d^2 = \dim_K A$) and every simple $A \otimes_K A^{op}$-module is isomorphic to A with the module structure given above (cf. Corollary 12.3 and Proposition 12.4b in [80]).

Let U be a module for a finite dimensional arbitrary algebra A. By $\mathrm{Rad}(U)$ we denote the radical of U which is the intersection of its proper maximal submodules. It is known that $\mathrm{Rad}(U) = \mathrm{Rad}(A)U$.

2.2.2 Extending scalars

It is sometimes useful to consider the K'-algebra $K' \otimes_K A$ where K' is a field extension K. We refer to this construction as extending scalars. (For example if $A \leq M_n(K)$ is the matrix algebra generated by matrices $g_1, \ldots, g_m$ then we can think of $K' \otimes_K A$ as the subalgebra of $M_n(K')$ generated by the same matrices $g_1, \ldots, g_m$ considered as matrices over K'.) For a subspace B of A we consider $K' \otimes_K B$ embedded into $K' \otimes_K A$ in the natural way. Many constructions such as products and commutators of complexes and even centralizers behave well with respect to extension of scalars. For example, $[K' \otimes_K B, K' \otimes_K C] = K' \otimes_K [B, C]$ and $C_{K' \otimes_K A}(K' \otimes_K B) = K' \otimes_K C_A(B)$.

2.2.3 Idempotents and the primary decomposition

An idempotent of A is a nonzero element $e \in A$ with $e^2 = e$. Two idempotents e and f are called orthogonal if $ef = fe = 0$. An idempotent is called primitive if it cannot be decomposed as a sum of two orthogonal idempotents. A system $e_1, \ldots, e_r$ of pairwise orthogonal idempotents is called complete if their sum is the identity of A. Primitive idempotents of the center of A are called primitive central idempotents. The primitive central idempotents are pairwise orthogonal and form a complete system in $Z(A)$.

Idempotents can be lifted from the semisimple part $A/\mathrm{Rad}(A)$. That is, if $\widetilde{e}$ is an idempotent in $A/\mathrm{Rad}(A)$ then there exists an idempotent e of A such that $e \in \widetilde{e}$. Even complete systems of pairwise orthogonal idempotents can be lifted: assume that $\widetilde{e}_1, \ldots, \widetilde{e}_r$ are pairwise orthogonal idempotents of $A/\mathrm{Rad}(A)$ such that $\widetilde{e}_1 + \ldots + \widetilde{e}_r = 1_{A/\mathrm{Rad}(A)}$. Then there exist $e_i \in \widetilde{e}_i$ $(i = 1, \ldots, r)$ such that $e_1, \ldots, e_r$ are pairwise orthogonal idempotents of A and $e_1 + \ldots + e_r = 1_A$.

If we lift a complete system $\widetilde{e}_1, \ldots, \widetilde{e}_r$ of pairwise orthogonal primitive central idempotents of $\mathrm{Rad}(A)$ as above, the we obtain a decomposition

$$A = e_1Ae_1 + \ldots + e_rAe_r + N',$$

as a direct sum of vector spaces, where N' is a subspace of $\mathrm{Rad}(A)$ and for every $i \in \{1, \ldots, r\}$, the subspace $A_i = e_iAe_i$ is a subalgebra of A with identity element e_i. Furthermore, A_i are primary algebras. (An algebra B is primary if $B/\mathrm{Rad}(B)$ is simple.) The decomposition above is called the primary decomposition of A, see Theorem 49.1 of [67].

2.2.4 Separability and the Wedderburn–Malcev theorem

It is obvious that $K' \otimes_K \mathrm{Rad}(A)$ is a nilpotent ideal of $K' \otimes_K A$. However, there are cases where $\mathrm{Rad}(K' \otimes_K A)$ can be bigger than $K' \otimes_K \mathrm{Rad}(A)$. A general sufficient condition for $\mathrm{Rad}(K' \otimes_K A) = K' \otimes_K \mathrm{Rad}(A)$ is that K' is a (not necessarily finite) separable extension of K. We say that A is separable over K if for every field extension K' of K the K'-algebra $K' \otimes_K A$ is semisimple. (Note that in Chapter 10 of [80] a more general definition of separable algebras over an arbitrary ring is given. The simple definition given here for algebras over a field is equivalent to the general one, see Corollary 10.6 of [80]). Separability of finite dimensional algebras generalizes the notion of separability of finite field extensions: by Proposition 10.7 of [80], A is separable iff the centers of the simple components of A are separable extensions of K. From this characterization it follows immediately that A is separable over K if and only if $K' \otimes_K A$ is semisimple where K' denotes the algebraic closure of K. Obviously, over a perfect ground field K the notion of separability coincides with semisimplicity. It is immediate that if A is separable then $K' \otimes A$ is separable as well for an arbitrary field extension K' of K. Direct sums, homomorphic images and tensor products of separable algebras are separable as well (cf. Section 10.5 of [80]).

An extremely useful result where separability plays a role is the Wedderburn–Malcev Principal Theorem (See Section 11.6 of [80] for a general form): Assume that $A/\mathrm{Rad}(A)$ is separable. Then there exists a subalgebra $D \leq A$ such that $D \cong A/\mathrm{Rad}(A)$ and $A = D + \mathrm{Rad}(A)$ (direct sum of vector spaces). Furthermore, if D_1 is another subalgebra such that $D_1 \cong A/\mathrm{Rad}(A)$ then there exists an element $w \in \mathrm{Rad}(A)$ such that $D_1 = (1 + w)^{-1}D(1 + w)$. We shall refer to such subalgebras as Wedderburn complements in A.

We shall make use of the following consequence of the Principal Theorem. It states that separable subalgebras of $A/\mathrm{Rad}(A)$ can be lifted to A.

Corollary 2.1. *Let A be a finite dimensional K-algebra and $B \leq A$ be a subalgebra of A which is separable over K and assume that $\widetilde{D}$ is a separable subalgebra of $A/Rad(A)$ containing $B + Rad(A)$. Then there exists a subalgebra D of A such that $B \leq D$ and $D \cong \widetilde{D}$.*

Proof. Working in the pre-image of $\widetilde{D}$ at the natural projection $A \to A/\mathrm{Rad}(A)$ we may assume that $\widetilde{D} = A/\mathrm{Rad}(A)$. Then, by the first part of the principal theorem there exists a subalgebra $D_1 \leq A$ such that $D_1 \cong \widetilde{D}$ and $A = D_1 + \mathrm{Rad}(A)$. Let π be the projection of A onto D_1 corresponding to this decomposition and $B_1 = \pi(B + \mathrm{Rad}(A))$. By comparing dimensions it is clear that $B_1 + \mathrm{Rad}(A) = B + \mathrm{Rad}(A)$. By the second part of the principal theorem, applied to the algebra $B + \mathrm{Rad}(A)$, there exists an element $w \in \mathrm{Rad}(A)$ such that $(1-w)^{-1}B(1-w) = B_1$. Now the subalgebra $D = (1-w)D_1(1-w)^{-1}$ has the required property. □

2.2.5 Tori

A toral K-algebra or torus over K is a finite dimensional commutative K-algebra which is separable over K. Let K' stand for the algebraic closure of K. Then T is a torus if and only if $K' \otimes T$ is isomorphic to the direct sum of copies of K'. Let $T \leq M_n(K)$ be a commutative matrix algebra. Then T is a torus if and only if the matrices in T can be simultaneously diagonalized over K'. By this we mean that there exists a matrix $b \in M_n(K')$ such that $b^{-1}Tb \subseteq Diag_n(K')$, where $Diag_n(K')$ is the matrix algebra consisting of the diagonal n by n matrices. (The diagonalization can be obtained by decomposing $K' \otimes V$ into a direct sum of irreducible $K' \otimes T$-modules.) By a maximal torus of the algebra A we mean a torus which is not properly contained in any other toral subalgebra of A. Note that by Corollary 2.1, maximal tori of $A/\mathrm{Rad}(A)$ can be lifted to maximal tori in A.

2.3 Polycyclic presentations of finite solvable groups

In Chapter 9 we present quantum algorithms for certain problems related to finite solvable groups of constant derived length. In order to simplify discussion therein, we need to fix a simple way to represent elements of such groups. We have chosen the so-called *refined polycyclic presentations* (discussed later on). Using that representation, multiplication in finite solvable groups of constant derived length can be accomplished efficiently, there is a unique description for subgroups which can be found quickly from systems of generators, and data structures supporting computations in subgroups and factor groups can be obtained easily.

We denote the commutator subgroup of a finite group G by G'. The *derived series* of G consists of $G, G', G'' = (G')'$, etc. Recall that G is solvable if this sequence reaches the trivial subgroup $\{1\} = \{1_G\}$. The number of steps required to reach the trivial subgroup is the derived length of G. We assume that the groups we encounter in Chapter 9 are presented in terms of so-called *refined polycyclic presentations* [50]. Such a presentation of a finite solvable group G is based on a sequence $G = G_1 > \ldots > G_{m+1} = 1$ where for each $1 \leq i \leq m$ the subgroup G_{i+1} is a normal subgroup of G_i and the factor group G_i/G_{i+1} is cyclic of prime order r_i. For each $i \leq m$ an element $g_i \in G_i \setminus G_{i+1}$ is chosen. Then

$g_i^{r_i} \in G_{i+1}$. Every element g of G can be uniquely represented as a product of the form $g_1^{e_1} \cdots g_m^{e_m}$, called the normal word for g, where $0 \leq e_i < r_i$.

Note that in the abstract presentation the generators are $g_1, \ldots, g_m$ and the for each index $1 \leq i \leq m$ the following relations are included:

- $g_i^{r_i} = u_i$, where $u_i = g_{i+1}^{a_{i,i+1}} \cdots g_m^{a_{i,m}}$ is the normal word for $g^{r_i} \in G_{i+1}$
- $g_i^{-1} g_j g_i = w_{ij}$ for every index $j > i$, where $w_{ij} = g_{i+1}^{b_{i,j,i+1}} \cdots g_m^{b_{i,j,m}}$ is the normal word for $g_i^{-1} g_j g_i \in G_{i+1}$.

We assume that elements of G are encoded by normal words (actually by the row vectors consisting of the exponents e_i as above) and there is a (in $\log |G|$) polynomial time algorithm – so called collection procedure – which computes normal words representing products. This is the case for groups of constant derived length, see [51]. If there is an efficient collection procedure then polycyclic presentations for subgroups (given by generators) and factor groups can be obtained in polynomial time, cf. [50]. The major notable subgroups including Sylow subgroups, the center, and the members of the derived series can also be computed in polynomial time.

The usual way to compute polycyclic presentations of subgroups can be used to obtain a unique encoding of subgroups. A sequence $h_1, \ldots, h_r$ of elements of a subgroup H of G is called an *induced polycyclic series* for H if there is a sequence of numbers $j_1 < j_2 < \ldots < j_r$ between 1 and m such that for every $i \in \{1, \ldots, r\}$,

- $H \cap G_{j_i}$ is generated by $h_i, h_{i+1}, \ldots, h_r$.

An induced polycyclic series is in *reduced echelon form* if, in addition, for every $i \in \{1, \ldots, r\}$,

- $h_i \in g_{j_i} G_{j_i+1}$,
- for every i' with $i < i' \leq r$, the exponent of $g_{j_{i'}}$ in the normal word for h_i is zero.

From an arbitrary system of generators for H such a series can be obtained in polynomial time (using the efficient collection procedure) by a noncommutative analogue of Gaussian elimination, combined with conjugation steps, see [50, 89]. By induction on the length, it can be seen that different reduced row echelon form sequences generate different subgroups.

We remark that this choice of model for computing in groups is just for simplifying presentation of our result. At the cost of introducing additional definitions and making some explanations somewhat longer, one could use other – more general – models, such as black box groups. Note however, that, using a quantum implementation [59] of an algorithm of R. Beals and L. Babai [11], refined polycyclic presentation for a solvable black box group can be computed in polynomial time.

2.4 Randomized algorithms

Recall that a classical randomized algorithm can be defined as a Turing machine (or a family Boolean circuits) where the the original input string x is supplemented by a further string z, called the *random source*. Assume that the machine computes the function $f(x, z)$

on input x, z. If we assume that the random source consists of r bits then the randomized algorithm computes on input x the value y with probability

$$\mathbf{Pr}_{z \in \{0,1\}^r} \left[f(x,z) = y \right].$$

The related complexity class is BPP (for Bounded error Probabilistic Polynomial time). This is the class of languages L such that L can be recognized by a polynomial time randomized algorithm with error probability at most 1/3. (That means that the output is just one bit and if $x \in L$ then the output is 0 with probability at most 1/3 and if $y \notin L$ then the output is 1 with probability at most 1/3.) Note that with independent iterations and taking the majority of answers the error probability can be made exponentially small.

Randomized algorithms are often referred to as *Monte Carlo* algorithms. In this thesis we shall use the term Monte Carlo for distinguishing algorithms with possible incorrect answers from so-called Las Vegas methods (the name has been introduced by L. Babai) which may fail with probability at most 1/3 but never return an incorrect answer. Maybe the most important practical advantage of a method of this type is that an unknown deviation of the random source from the uniform distribution can be compensated by iterations until a successful outcome of the procedure.

One of the most important tools used in randomized algorithms of this thesis is the Schwartz-Zippel Lemma [87, 96].

Fact 2.2 (Schwartz-Zippel Lemma). *Let F be a field, let Ω be a non-empty finite subset of F and let $f \in F[x_1, \ldots, x_m]$ be a nonzero polynomial of total degree d. Then*

$$\mathbf{Pr}_{a_1,\ldots,a_m \in \Omega} \left[f(a_1, \ldots, a_m) = 0 \right] \leq \frac{d}{|\Omega|}.$$

We remark that refinements of the Schwartz-Zippel Lemma in important subcases where F is a finite field and $\Omega = F$ are known from algebraic coding theory as theorems on the relative distance of generalized Reed-Muller codes, see [2].

2.5 Quantum computing

In the description of quantum algorithms we use a simple, restricted model of quantum computers which is – up to polynomial slowdown – equivalent to many others, including the quantum Turing machine introduced by E. Bernstein and U. Vazirani in [13]. The model we use is the quantum circuit model with one- and two-qubit gates. This model is very close to the model introduced by D. Deutsch in [24], and whose computational power was investigated by A. Yao [95]. The main difference is that while Deutsch and Yao consider gates acting on 3 qubits, we – taking more recent developments regarding universality into account – restrict ourselves to one- and two-qubit gates. (Note that in Chapter 8 of this thesis we present some results regarding universality of gate sets acting on more than two qubits, or even on several qu*d*its. However for describing quantum *algorithms* it will be convenient to stick to a very simple model.) In this section we describe this simple model, and – for convenience of readers not familiar with quantum computations – give details of certain basic techniques that we shall use in Chapter 8. An excellent introduction to quantum computing written for pure mathematicians can be found in [10]. Its preprint is available on the Internet.

2.5.1 Quantum circuits

A *qubit* corresponds to the complex Euclidean space $\mathbb{C}^2$. We fix an orthonormal basis v_0, v_1 of $\mathbb{C}^2$. We call this basis the *computational basis*. The *state* of the qubit is a unit vector $a_0 v_0 + a_1 v_1$ in $\mathbb{C}^2$. A system consisting of n qubits, or an n-qubit *register* corresponds to the complex Euclidean space $\mathbb{C}^{2^n}$. It is instructive to consider $\mathbb{C}^{2^n}$ as the tensor product of n copies of $\mathbb{C}^2$. In view of this decomposition, the register can be considered as if it is composed of n qubits. The computational basis consists of just the products of the computational basis elements corresponding to the qubits. These products correspond to bit strings of length n. It is common to use the notation $|s\rangle$ for a computational basis vector corresponding to the string s. Thus $|0\rangle$ stands for v_0, $|1\rangle$ for v_1, and $|01\rangle$ for $v_0 \otimes v_1$, etc. A possible state of an n-qubit register is a unit vector from C^{2^n}. It can be written as a linear combination (*superposition* according to the quantum computing literature)

$$\underline{\psi} = \sum_{s \in \{0,1\}^n} a_s |s\rangle. \tag{2.1}$$

In this thesis we use the $|\,\rangle$-notation exclusively for computational basis elements, emphasizing that they are states from a very specific collection. (Note that in the literature Dirac's $|\,\rangle$-notation is also used to simplify some constructions in linear/tensor algebra therefore $|\,\rangle$ is used to denote arbitrary vectors. Here we do not use those constructions extensively, therefore general states will be typically denoted by underlined Greek letters.)

To every state of an n-qubit register, that is to every unit vector $\underline{\psi}$ of the form (2.1) there belongs a probability distribution over $\{0,1\}^n$ where the probability of the string s is $|a_s|^2$. Intuitively, the register in state $\underline{\psi}$ is considered as if it were simultaneously in all the states $|s\rangle$ with "weight" (*amplitude*) a_s and if we *measure* (or *observe*) the register then we obtain $|s\rangle$ with probability $|a_s|^2$.

A one-qubit gate is just a unitary transformation on $\mathbb{C}^2$ and a two-qubit gate is a unitary transformation on $\mathbb{C}^4$. For instance, the linear extension of the NOT or bit flip operation ($|0\rangle \leftrightarrow |1\rangle$) is a one-qubit gate. Another important one-qubit gate is the so-called *Hadamard gate*. Its matrix in the computation basis is

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The linear extension of the *exclusive or* (also known as *conditional not*) operation which maps $|b_1\rangle \otimes |b_2\rangle$ to $|b_1 \oplus b_2\rangle \otimes |b_2\rangle$ is a two-qubit gate. Another example for a two-qubit gate is the so-called controlled phase shift

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \omega \end{pmatrix},$$

where ω is a complex number with $|\omega| = 1$.

Given a one-qubit gate U and $i \in \{1, \ldots, n\}$ then we can let U act on the ith qubit of an n-qubit register as the operation $U \otimes I_{2^{n-1}}$, where we consider the decomposition of $\mathbb{C}^{2^n}$ as the tensor product of $\mathbb{C}^2$ corresponding to the ith qubit with $\mathbb{C}^{2^{n-1}}$ corresponding to the rest. Thus, a one-qubit gate can be *wired* in n ways to an n-qubit system. Similarly, let $i \neq j \in \{1, \ldots, n\}$. Then we can let a two-qubit gate U act on $\mathbb{C}^{2^n}$ as $U \otimes I_{2^{n-2}}$, where U

acts on the space corresponding to the ith and jth qubits (in this order) and $I_{2^{n-2}}$ stands for the identity of $\mathbb{C}^{2^{n-2}}$ corresponding to the rest. Thus a two-qubit gate can be wired to an n-qubit systems in $n(n-1)$ ways.

We consider quantum circuits built from one- and two-qubit gates. This is a finite sequence of one- and two-qubit gates, each wired to an n-qubit system. The circuit implements the unitary transformation on $\mathbb{C}^{2^n}$ which is just the product of the individual wired gates. The *size* of the circuit (or its *running time*) is just the length of the sequence.

We assume that a part of the n qubits is designated to the input, a disjoint part is designated to the output and the rest are for temporary storage, the so-called workspace. Thus we assume that $n = n_i + n_o + n_w$ and $\mathbb{C}^{2^n}$ is decomposed as a tensor product of $\mathbb{C}^{2^{n_i}} \otimes \mathbb{C}^{2^{n_o}} \otimes \mathbb{C}^{2^{n_w}}$. Intuitively, we have an input register, an output register, and possibly some further registers for the workspace. (We use the term *register* also for a subset of $\{1, \ldots, n\}$ representing any meaningful piece of the system we are working with.) By inputting x to the circuit we mean that we let the unitary transformation implemented by the circuit act on $|x\rangle \otimes |\underline{0}\rangle \otimes |\underline{0}\rangle$. Assume that the result is the vector (or state)

$$\underline{\psi}(x) = \sum_{(s_i,s_o,s_w)\in\{0,1\}^{n_i+n_o+n_w}} a_{s_i,s_o,s_w}(x)|s_i\rangle \otimes |s_o\rangle \otimes |s_w\rangle.$$

We say that the probability of that on input x the circuit computes y is the probability according the distribution corresponding to $\underline{\psi}(x)$ of that $s_o = y$ in the triple (s_i, s_o, s_w), that is

$$\sum_{s_i\in\{0,1\}^{n_i}, s_w\in\{0,1\}^{n_w}} |a_{s_i,y,s_w}(x)|^2.$$

Intuitively, this notion of computation corresponds to that a computational phase which consists of application of a quantum circuit is followed by *measuring* the result. The quantum analogue of the class BPP is BQP. It consists of the languages recognized by a quantum circuit in polynomial time with error probability at most 1/3. Note that using independent iteration, the error probability can be made exponentially small.

2.5.2 Cleaning up

As constituents of other circuits, we shall encounter a restrictive class of quantum circuits which compute functions. Let X be subset of $\{0,1\}^{n_i}$ and let f be a function $f : X \to \{0,1\}^{n_o}$. We consider a quantum circuit which, on every $x \in X$, computes the value $f(x)$ with probability 1. This means that on input $|x\rangle|\underline{0}\rangle|\underline{0}\rangle$, the result is the tensor product of $|f(x)\rangle$ with some unit vector $\underline{\psi}'(x) \in \mathbb{C}^{2^{n_i+n_w}}$. With a constant slowdown, we can modify such a circuit to obtain a circuit which, for every $x \in X$, transforms the state

$$|x\rangle \otimes |\underline{0}\rangle \otimes |\underline{0}\rangle \text{ to } |x\rangle \otimes |f(x)\rangle \otimes |\underline{0}\rangle.$$

This is done using the following standard cleanup trick. We extend the workspace by space for a second copy of the output register, that is, we increase n_w by n_o. We perform the original circuit with this copy of output register and leave the original output register intact. Notice that by the assumption, after executing the circuit, we have the state

$$\underline{\psi}'(x) \otimes |\underline{0}\rangle \otimes |f(x)\rangle,$$

where $\underline{\psi}'(x)$ is some unit vector from $C^{2^{n_i+n_w}}$. Next we copy the contents of the second output register to the original output register. This can be done by computing the bit-wise exclusive or of the two registers into the original output register. Now we undo (perform the inverse of) the circuit, leaving the current contents of the output register intact. The result is the state

$$|x\rangle \otimes |f(x)\rangle \otimes |\underline{0}\rangle,$$

as required. With some sloppiness, we will refer to such a circuit as a circuit which implements

$$|x\rangle \otimes |\underline{0}\rangle \mapsto |x\rangle \otimes |f(x)\rangle.$$

Thus we omit the qubits in the workspace, which are zero both initially and finally. (The starting state is actually the tensor product of $|x\rangle \otimes |\underline{0}\rangle$ with $|\underline{0}\rangle$ in the workspace and, similarly, the result is the product of $|x\rangle \otimes |f(x)\rangle$ with $|\underline{0}\rangle$.) As we are not concerned with the accurate space complexity of our algorithms, ignoring such "cleaned up" workspace will lead neither to confusion nor to loss of information. (Note that a circuit of size ℓ can actually use at most 2ℓ auxiliary qubits, so a circuit of polynomial size can be actually implemented on a polynomial number of qubits.)

The same trick applies to certain more general situations which we encounter as ingredients of larger circuits. Assume that the domain of the function f is a subset Ψ of $\mathbb{C}^{2^{n_i}}$ but the range is still from $\{0,1\}^{n_o}$, that is, we allow f to be defined on certain "quantum" states but its value is always "classical". (See the *swap test* discussed later in this section for an important example of such a function.) Assume further that we have a circuit which transforms the state $\underline{\psi} \otimes |\underline{0}\rangle \otimes |\underline{0}\rangle$ to the tensor product of $|f(\underline{\psi})\rangle$ with a vector from $\mathbb{C}^{2^{i+w}}$ (depending on $\underline{\psi}$) for every $\underline{\psi} \in \Psi$. Then the with the same trick as above, with a constant slowdown we can construct a circuit which transforms $\underline{\psi} \otimes |\underline{0}\rangle \otimes |\underline{0}\rangle$ to $\underline{\psi} \otimes |f(\underline{\psi})\rangle \otimes |\underline{0}\rangle$ for every $\underline{\psi} \in \Psi$. In this context we can also ignore designation of the workspace.

2.5.3 Classical computation as quantum computation

Although every quantum circuit implements a unitary (and hence invertible) operation, with some care it can be seen that quantum circuits can simulate deterministic computations with a constant slowdown. To be more precise, if the function $x \mapsto f(x)$ (from $X \subseteq \{0,1\}^{n_i}$ to $\{0,1\}^{n_o}$) can be implemented by a Boolean circuit of size ℓ then there is a quantum circuit of size $O(\ell)$ which implements $|x\rangle \otimes |\underline{0}\rangle \mapsto |x\rangle \otimes |f(x)\rangle$ for $x \in X$. This follows from the results of the theory of reversible computations, see [91]. (We remark that the procedure may actually use some workspace which is cleaned up.)

A randomized algorithm can be simulated on a quantum computer as follows. Assume that a Boolean circuit computes the function $(x,z) \mapsto f(x,z)$ where $z \in \{0,1\}^r$ corresponds to the random bit string. Given

$$|x\rangle \otimes |\underline{0}\rangle \otimes |\underline{0}\rangle$$

we first apply the Hadamard gate to each of the qubits of the second register and obtain the state

$$\frac{1}{\sqrt{2^r}} \sum_{z \in \{0,1\}^r} |x\rangle \otimes |z\rangle \otimes |\underline{0}\rangle$$

We next apply the quantum implementation of $|x\rangle \otimes |z\rangle \otimes |\underline{0}\rangle \mapsto |x\rangle \otimes |z\rangle \otimes |f(x,z)\rangle$ and obtain

$$\underline{\psi}(x) = \frac{1}{\sqrt{2^r}} \sum_{z \in \{0,1\}^r} |x\rangle \otimes |z\rangle \otimes |f(x,z)\rangle.$$

It is immediate that, according to the distribution corresponding to $\underline{\psi}(x)$, the probability that the third register contains y is the usual probability that $f(x,z)=y$ where z is drawn uniformly from $\{0,1\}^r$. This argument shows that $BPP \subseteq BQP$.

2.5.4 Numerical vs. probabilistic errors

Very often we are satisfied with sufficiently good approximate implementations of quantum circuits. The error is the Euclidean distance from the correct outcome. If a compound circuit consists of ℓ smaller circuits then in order to obtain error at most ϵ it is sufficient to have approximations of the constituents which work with error at most ϵ/ℓ for each meaningful input state.

Assume for example that f is a function from $X \subseteq \{0,1\}^{n_i}$ to $\{0,1\}^{n_o}$ and we have a circuit, which, for every $x \in X$, transforms the state $|x\rangle \otimes |\underline{0}\rangle$ to a unit vector

$$\underline{\psi}'(x) = \sum_{s\in\{0,1\}^{n_i+n_o}} a_s(x)|s\rangle$$

at distance form $|x\rangle \otimes |f(x)\rangle$ at most ϵ. Then, taking the square of the distance we have

$$\sum_{\substack{s \in \{0,1\}^{n_i+n_o} \\ s \neq (x, f(x))}} |a_s(x)|^2 \leq \epsilon^2.$$

Notice that the left hand side is the probability of $s \neq (x, f(x))$ according to the distribution corresponding to the state $\underline{\psi}'(x)$. Thus, a numerical error ϵ in a quantum circuit which computes the function f results in a probabilistic error ϵ^2. This argument shows that the class BQP is robust against taking sufficiently good approximation of ingredients of quantum algorithms.

One of the most important applications of this fact is the following. Instead of allowing arbitrary one- and two-qubit gates as building blocks of quantum circuits, one can take a fixed finite set which generate a dense subgroup of the unitary group U_4. Then, by the Solovay–Kitaev Theorem [69], for every $\epsilon > 0$, an arbitrary unitary operation in U_4 can be approximated with error at most ϵ by a product of $\frac{1}{\epsilon^{O(1)}}$ operations from the fixed set (the implicit constant $O(1)$ depends on the gate set). Thus a circuit of length ℓ built from arbitrary gates can be approximated with error 0.01 by a circuit of length $\ell \cdot (\log \ell)^{O(1)}$ with a set of restricted gates.

2.5.5 State sampling

As certain building blocks of our algorithms we shall use classical algorithms performing a sort of statistical distribution analysis in the following context. Assume that we have K copies of a state of the form

$$\underline{\psi} = \sum_{s\in\{0,1\}^n} a_s|s\rangle \otimes \underline{\psi}_s,$$

where for every $s \in \{0,1\}^n$, $\underline{\psi}_s$ is a unit vector from $\mathbb{C}^{2^{n_1}}$ and we want to evaluate a function f at $\underline{\psi}$ using these K copies. (We assume that the domain of f is $\Psi \subseteq \mathbb{C}^{2^{n+n_1}}$, its range is a subset of $\{0,1\}^{n_o}$, and $\underline{\psi} \in \Psi$.) Having K copies of $\underline{\psi}$ means that we are actually given the tensor power $\underline{\psi}^{\otimes K}$. By expanding the tensor power we see that

$$\underline{\psi}^{\otimes K} = \sum_{s_1,\ldots,s_K \in \{0,1\}^n} a_{s_1} \cdots a_{s_K} |s_1\rangle \otimes \underline{\psi}_{s_1} \otimes \cdots \otimes |s_K\rangle \otimes \underline{\psi}_{s_K}.$$

We pass this tensor power to a classical algorithm which computes the value $f'(s_1,\ldots,s_K)$ as an estimate for $f(\underline{\psi})$, where f' is a function from $\{0,1\}^n$ to $\{0,1\}^{n_o}$.

To be more precise, the initial sate (ignoring workspace) is actually $\underline{\psi}^K \otimes |\underline{0}\rangle$, the result of the estimating procedure is

$$\underline{\psi}^{\otimes K} = \sum_{s_1,\ldots,s_K \in \{0,1\}^n} a_{s_1} \cdots a_{s_K} |s_1\rangle \otimes \underline{\psi}_{s_1} \otimes \cdots \otimes |s_K\rangle \otimes \underline{\psi}_{s_K} \otimes |f'(s_1,\ldots,s_K)\rangle,$$

and we are interested in the distance of it from the desired result $\underline{\psi}^K \otimes |f(\underline{\psi})\rangle$. Observe that the square of the distance is at most

$$2 \cdot \sum_{\substack{s_1,\ldots,s_K \in \{0,1\}^n \\ f'(s_1,\ldots,s_K) \neq f(\underline{\psi})}} |a_{s_1}|^2 \cdots |a_{s_K}|^2.$$

Notice that this is 2 times the probability of that $f'(s_1,\ldots,s_K) \neq f(\underline{\psi})$, where $s_1,\ldots,s_K$ are drawn independently according to the distribution (on $\{0,1\}^n$) corresponding to $\underline{\psi}$. Thus the error is related to the statistical error of the classical procedure applied.

To see a specific example, we consider the following problem. Assume that we want to decide whether two states $\underline{\psi}_1$ and $\underline{\psi}_2$ from $\mathbb{C}^{2^{n/2}}$ are identical under the promise that they are either identical or orthogonal and we are given K copies of both states. Thus the initial state is $(\underline{\psi}_1 \otimes \underline{\psi}_2)^{\otimes K} \otimes |0\rangle$ and the desired outcome is $(\underline{\psi}_1 \otimes \underline{\psi}_2)^{\otimes K} \otimes |0\rangle$ if $\underline{\psi}_1 \perp \underline{\psi}_2$, and it is $(\underline{\psi}_1 \otimes \underline{\psi}_2)^{\otimes K} \otimes |1\rangle$ if $\underline{\psi}_1 = \underline{\psi}_2$.

This task is accomplished by the swap test [17] which we outline below. We take a workspace consisting of K auxiliary qubits, one for each pair. So our initial state is actually

$$(\underline{\psi}_1 \otimes \underline{\psi}_2 \otimes |0\rangle)^{\otimes K} \otimes |0\rangle.$$

We apply the Hadamard gate to each qubit of the workspace and obtain the state

$$\left(\frac{1}{\sqrt{2}} \underline{\psi}_1 \otimes \underline{\psi}_2 \otimes (|0\rangle + |1\rangle)\right)^{\otimes K} \otimes |0\rangle.$$

Now we swap (exchange bit by bit) each copy of the pair if the corresponding auxiliary qubit contains 1. If the qubit contains zero we do nothing. The result is

$$\left(\frac{1}{\sqrt{2}} \left(\underline{\psi}_1 \otimes \underline{\psi}_2 \otimes |0\rangle + \underline{\psi}_2 \otimes \underline{\psi}_1 \otimes |1\rangle\right)\right)^{\otimes K} \otimes |0\rangle.$$

Next we apply the Hadamard gate again to the auxiliary qubits and obtain the state

$$\left(\frac{1}{2} \left(\left(\underline{\psi}_1 \otimes \underline{\psi}_2 + \underline{\psi}_2 \otimes \underline{\psi}_1\right) \otimes |0\rangle + \left(\underline{\psi}_1 \otimes \underline{\psi}_2 - \underline{\psi}_2 \otimes \underline{\psi}_1\right) \otimes |1\rangle\right)\right)^{\otimes K} \otimes |0\rangle.$$

If $\underline{\psi}_2 = \underline{\psi}_1$ then put $\underline{\psi}'_0 = \underline{\psi}'_1 = \underline{\psi}_1$ and $\underline{\psi} = \underline{\psi}'_0 \otimes |0\rangle + 0 \cdot \underline{\psi}'_1 \otimes |1\rangle$. If $\underline{\psi}_2 \perp \underline{\psi}_1$ then put $\underline{\psi}'_0 = \frac{1}{\sqrt{2}}(\underline{\psi}_1 \otimes \underline{\psi}_2 + \underline{\psi}_2 \otimes \underline{\psi}_1)$, $\underline{\psi}'_1 = \frac{1}{\sqrt{2}}(\underline{\psi}_1 \otimes \underline{\psi}_2 - \underline{\psi}_2 \otimes \underline{\psi}_1)$ and $\underline{\psi} = \frac{1}{\sqrt{2}}(\underline{\psi}_0 \otimes |0\rangle + \underline{\psi}_1 |1\rangle)$. We have the state $\underline{\psi}^{\otimes K} \otimes |0\rangle$ and apply the scheme described above in this context. Here $f(\underline{\psi}) = 1$ if $\underline{\psi}_1 = \underline{\psi}_2$ and $f(\underline{\psi}) = 0$ if $\underline{\psi}_1 \perp \underline{\psi}_2$. In the first case, the probability of 1 according to the distribution on $\{0,1\}$ corresponding to $\underline{\psi}$ is zero, while in the second case both 0 and 1 have probability 1/2. We take $f'(\underline{0}) = 1$ and $f'(s) = 0$ if $s \neq \underline{0}$, that is, we return 1 if and only if all the K bits we see are zero. The probability of that we make a wrong decision is 0 in the first case while it is $\frac{1}{2^K}$ in the second case. Thus, after this simple statistical distribution analysis, the distance from the desired state $\underline{\psi}^{\otimes K} \otimes |f(\underline{\psi})\rangle$ is exponentially small in K. If we perform the usual cleanup technique, the distance from the desired final outcome $(\underline{\psi}_1 \otimes \underline{\psi}_2)^{\otimes K}|\text{answer}\rangle$ will remain the same.

2.5.6 The hidden subgroup problem

Almost all the computational problems in which quantum algorithms have an exponential advantage over the known classical methods are related to the *hidden subgroup problem* (HSP) which is the following. Let G be a finite group, and let H be a subgroup of G. Let f be a function mapping G into a finite set, say $\{0,1\}^s$. We say that f *hides* the subgroup $H \leq G$ if f is constant on every left coset of H but it takes different values on distinct cosets. (Equivalently, $f(x) = f(y)$ if and only if $xH = yH$.) We assume that f is given by a quantum oracle (that is, a unitary operation mapping states of type $|x\rangle \otimes |\underline{0}\rangle$ $(x \in G)$ to $|x\rangle \otimes |f(x)\rangle$. The task is finding H, say, by means of generators. (In this thesis we are concerned with a restricted version where the output is required to be a well defined unique description of H.)

In the most important applications the oracle is implemented by polynomial time algorithms. For example, in the discrete logarithm problem in an abelian group we have $G = \mathbb{Z}_m \oplus \mathbb{Z}_m$ and $f(u,v) = a^u b^{-v}$ (computed using fast exponentiation) where we want to compute $\log_a b$ and – to simplify discussion – m is assumed to be the order of both a and b. Then $H = \{((\log_a b)v, v) | v \in \mathbb{Z}_m\}$ and from any system of generators for H one can compute the desired logarithm easily. Similarly, if we want to compute the order of a in an abelian group of exponent m then $G = \mathbb{Z}_m$, $f(u) = a^u$, and $H = o(a)\mathbb{Z}_m$. Again, from a system of generators of H one obtains $o(a)$ easily. (We remark that Shor's order finding and discrete logarithm algorithms do not need knowledge of the exponent. These methods should be rather interpreted as hidden subgroup algorithms for the infinite groups $\mathbb{Z} \oplus \mathbb{Z}$ and $\mathbb{Z}$, respectively.) For computing automorphism groups of graphs (testing graph isomorphism can be reduced to this problem) the group G is the symmetric group (acting on the vertex set) and the values of the function are the "permuted" versions of the graph. Then the hidden subgroup is just the automorphism group of the graph.

2.5.7 The quantum Fourier transform

The quantum Fourier transform is one of the most important tools in the existing hidden subgroup algorithms. The quantum Fourier transform of an abelian group G is the unitary transformation which maps vectors of the form $|x\rangle$ where $x \in G$ to

$$\frac{1}{\sqrt{|G|}} \sum_{\chi \in \widehat{G}} \chi(x) |\chi\rangle,$$

where by $\widehat{G}$ we denote the set of the (linear) characters of the group G. Recall that a character of a finite abelian group is a homomorphisms from G to the multiplicative group of $\mathbb{C}$. With the point-wise multiplication $\widehat{G}$ is an abelian group isomorphic to G. In particular, $\left|\widehat{G}\right| = |G|$, so we can use a bijection between G and $\widehat{G}$ so that the Fourier transform is a transformation of a space onto itself. Note that it maps $|0\rangle$ to the uniform superposition of characters, which, by the bijection above, is the uniform superposition of the element of G. (The uniform superposition of a set $S \subseteq \{0,1\}^s$ is just the vector $\frac{1}{\sqrt{|S|}}\sum_{s\in S}|s\rangle$.) For an arbitrary finite abelian group G, its Fourier transform can be approximated in time polynomial in $\log |G|$. More precisely, for any abelian group G and for any ϵ, there is a quantum circuit of size $\log |G|^{O(1)} \log \frac{1}{\epsilon}$ which approximates the Fourier transform of G with precision ϵ, see [68]. The precision is understood in the operator norm. That is, when we apply the approximation to a unit vector then the distance from the precise Fourier transform will be at most ϵ.

Chapter 3

Finiteness of matrix semigroups over function fields over finite fields

This brief chapter is based on the note [55]. Here we present an application of computing the radical of a matrix algebra to deciding finiteness of a matrix semigroup given by generators whose entries are from a function field over a finite field.

In [4], L. Babai, R. Beals and D. Rockmore proposed an ingenious polynomial time algorithm for deciding finiteness of groups generated by matrices with entries from algebraic number fields. We remark that this algorithm is an ingredient of Jeandel's polynomial method for deciding N-universality of a set of quantum gates defined over a number field, see Chapter 8 for more details.

In [83] D. Rockmore, K.-S. Tan and R. Beals considered testing finiteness of matrix groups over function fields $F(t)$. Using a reduction to the case solved in [4], they showed that the problem is soluble in polynomial time if the base field F is a number field. They also showed that the problem is algorithmically solvable (in exponential time) in the case where F is a finite field. In this chapter we improve and extend this latter result: we present a deterministic polynomial time algorithm for testing finiteness of a *semigroup* S generated by matrices with entries from function fields over finite fields with a *constant number of variables.*

Finiteness of matrix groups over a univariate function field over a finite field was shown to be algorithmically soluble in [83] by giving a sharp exponential upper bound on the dimension of the matrix algebra generated by S over the field of constants. One of the exponential time algorithms proposed in [83] was expected to be improvable. The polynomial time method presented in this chapter actually combines the ideas of that algorithm with a procedure (e.g. from [60]) for calculating the radical.

The following simple observation from [83] reduces our task to deciding finiteness of certain matrix rings.

Observation 3.1. *Let F be a finite field, K be an arbitrary extension field of F, and S be a multiplicative subsemigroup of $M_n(K)$ generated by the finite set $\{a_1, \ldots, a_s\}$ of n by n matrices. Then S is finite if and only if the F-subalgebra A of $M_n(K)$ generated by $a_1, \ldots, a_s$ is finite.*

It is pointed out in Theorem 3.16 of [83] that $\dim_F A$ can be exponential in n. The next statement enables us to avoid computing a basis of the whole algebra.

Lemma 3.2. *Let A be a finitely generated algebra over an arbitrary field F and assume that I is a nilpotent ideal of A such that A/I is finite dimensional. Then A is finite dimensional as well.*

Proof. Assume that $a_1, \ldots, a_s$ generate A as an F-algebra. Choose elements $b_1, \ldots, b_r$ from A such that $b_1 + I, \ldots, b_r + I$ form a basis of A/I. Write

$$a_l = \sum_{k=1}^{r} \alpha_{lk} b_k + c_l$$

and

$$b_i b_j = \sum_{k=1}^{r} \beta_{ijk} b_k + d_{ij}$$

with $\alpha_{lk}, \beta_{ijk} \in F$ and $c_l, d_{ij} \in I$.

We claim that the elements c_l $(l = 1, \ldots, s)$ and d_{ij} $(i, j = 1, \ldots, r)$ generate I as an ideal of A. Indeed, let J be the ideal generated by these elements. Obviously $J \leq I$. On the other hand, $b_1 + J, \ldots, b_r + J$ span a subalgebra of A/J complementary to I/J. Also, it contains all the generators $a_i + J$ for A/J. Hence $I/J = (0)$, which means that $I = J$. From the claim we infer that I/I^2 is a finitely generated A/I-module and hence it is finite dimensional over F. Thus A/I^2 is finite dimensional as well and the proof can be completed by induction on the nilpotency class of I. □

The next result states that modulo the radical of KA, the dimension of A is small.

Theorem 3.3. *Let F be an arbitrary field. Assume that K is an extension field of F such that for every finite extension E of F the algebra $K \otimes_F E$ is a field (e.g., K is purely transcendental over F). Let A be a finitely generated F-subalgebra of the finite dimensional K-algebra B. Let ϕ stand for the natural projection $B \to B/Rad(B)$. Then the following assertions are equivalent.*

(1) $\dim_F A$ *is finite.*

(2) $\dim_F \phi(A)$ *is finite.*

(3) $\phi(KA)$ *and* $K \otimes_F \phi(A)$ *are isomorphic K-algebras.*

(4) $\dim_F \phi(A) \leq \dim_K(B)$.

Proof. The implications (1)⇒(2), (3)⇒(4), and (4) ⇒ (2) are obvious.

To shorten notation in the rest of the proof we put $J = \text{Rad}(B)$. Since J is nilpotent we have $J^{\dim_K J} = (0)$. This also implies that $\ker \phi_{|A} = A \cap J$ is a nilpotent ideal of A. The implication (2)⇒(1) follows from this observation and the lemma.

To see that (2) implies (3) assume that $\dim_F \phi(A)$ is finite. Without loss of generality we may assume that $B = KA$. By going over the factors $A/(A \cap J)$ and $B/K(A \cap J)$, we may further assume that $A \cap J = (0)$. Then, since the K-linear span of every nilpotent ideal of A is a nilpotent ideal of B, A is either semisimple or zero. If A is zero then (3) trivially holds. Otherwise let $A_1, \ldots, A_r$ be the minimal nonzero ideals of A. Then $A = A_1 + \ldots + A_r$ and the sum is direct. Also, $K \otimes_F A$ is the direct sum of the ideals $K \otimes_F A_i$.

We claim that for every index i the K-algebra $K \otimes_F A_i$ is simple. Indeed, let C_i be the center of A_i. Then C_i is a finite extension field of F. By the assumption on K, $K \otimes_F C_i$ is again a field. On the other hand, it is easy to see that the center of $K \otimes_F A_i$ is $K \otimes_F C_i$. Thus $K \otimes_F A_i$ is a simple K-algebra, as claimed.

Next we observe the natural map $K \otimes A \to B$ induced by the multiplication of elements of A by scalars from K is a K-algebra epimorphism. But this map is a monomorphism as well because it is nonzero on any simple component $K \otimes_K A_i$ of $K \otimes_K A$. This concludes the proof of the theorem. □

The algorithm proposed in [60] is directly applicable in the context of the theorem above. It is a deterministic method which computes the radical of a matrix algebra $A \leq M_n(F_q(x_1, \ldots, x_m))$. The running time is $(n + s + d + \log q)^{O(m)}$, where s is the number of generators and d is the maximum degree among all the numerators and denominators of the entries appearing in the generators for A. We remark that a more general – and more transparent – approach to computing the radical of algebras of positive characteristic can be found in [21]. Even the more efficient method discussed in the next chapter of this thesis could also be used. However, as we are merely interested in polynomial time methods and as the time complexity of the algorithm of [60] is explicitly given, it is easiest to refer to that paper.

Using the algorithm of [60], we can find generators for the algebra $\phi(A)$ defined in the theorem in time $(n + s + d + \log q)^{O(m)}$. Then we proceed with collecting F_q-linearly independent elements of $\phi(A)$ as products of generators. We either find a basis of $\phi(A)$ in $O(n^2)$ rounds, or stop with the conclusion that our semigroup is infinite. We remark that the method can be considered as an improved and generalized version of the algorithm proposed in Subsection 3.4.1 of [83]. We obtained the following.

Corollary 3.4. *There is a deterministic algorithm, which, in time $(n + s + d + \log q)^{O(m)}$ decides whether the semigroup generated by a finite set of matrices with entries from the function field $F_q(x_1, \ldots, x_m)$ is finite. Here, s is the number of generators and d stands for the maximum among the degrees of all the numerators and denominators of the entries in the generators. In particular, the algorithm runs in polynomial time for constant m.*

Chapter 4

Finding the radical of matrix algebras using Fitting decompositions

The material of this chapter is based on the paper [53]. Here we present an approach to calculating the Jacobson radical of a matrix algebra based on the Fitting decomposition with respect to the simultaneous adjoint action of maximal tori. This idea results in a reduction to finding the radical of a Lie nilpotent subalgebra or a commutative factor thereof. We also describe a probabilistic version for computing elements which generate the radical as an ideal.

We assume that the input is a (usually small) finite set of matrices which generate A as an algebra and the output is expected to be a set of matrices which generate $\mathrm{Rad}(A)$ as an ideal. We sketch a deterministic algorithm which works over an arbitrary field with effective arithmetic and reduces the problem of calculating $\mathrm{Rad}(A)$ to finding $\mathrm{Rad}(B)$ for a commutative algebra B which is a factor of a subalgebra A. The algorithm performs $n^{O(1)}$ operations and is of theoretical interest as the task of computing the radical is known to be unsolvable by an algorithm using merely the field operations. This result can be interpreted as all the obstacles regarding computability of the radical of an algebra are already there in a commutative subfactor.

We also present a probabilistic algorithm of Monte Carlo type which works over a sufficiently large perfect ground field where square-free factorization of polynomials can be carried out efficiently. (Examples of such fields are fields of characteristic zero and finite fields). This appears to be the first attempt to make use of randomization in computing the radical. Provided that the number of generators is small and random elements of A can be generated efficiently the method performs about $O(n^4)$ operations in K.

All the known methods for computing the radical are based on solving systems of linear (or semilinear) equations (cf. [38, 85, 28, 21]). The coefficients are the traces (and other invariants in positive characteristic) of the products b_ib_j where $b_1, \ldots, b_s$ is a basis of A. Unfortunately it is not known how to determine the coefficients in a way more efficient than computing the diagonal elements of the product b_ib_j for $O(s^2)$ pairs b_i, b_j. Since s can be as large as n^2, all the previously known algorithms require $\Omega(n^6)$ operations.

The approach presented here is different. The key idea parallels the method used by de Graaf [44] in a practical algorithm for computing the nilradical of Lie algebras. Here, using the Fitting decomposition with respect to the adjoint actions of appropriate subalgebras, we reduce the task to computing the radical of a subalgebra which is nilpotent as a Lie algebra. Factoring by the commutator ideal this leads to a reduction to the commutative case. The probabilistic version is based on the same ideas combined with methods for generating random elements of centralizers of certain subalgebras rather than computing the whole centralizers by solving systems of linear equations.

This chapter is structured as follows. In the rest of this introductory part we give a brief description of the computational models we work with and discuss assumptions on random generators for the probabilistic algorithm. The main algorithmic results are also stated here. Section 4.1 is devoted to a summary known facts related to tori in associative algebras. The theoretical background of the algorithms is presented in Section 4.2. The reduction algorithm which works over an arbitrary field can be found in Section 4.3. Our basic computational tool, an efficient algorithm for generating elements of the centralizer of a single semisimple matrix is presented in Section 4.4. We conclude with Section 4.5, where we describe the probabilistic method for finding the radical.

We assume that the field K admits effective procedures for performing the field operations as well as equality tests. For such a general field we obtain the following.

Theorem 4.1. *There is a deterministic algorithm which reduces the problem of computing the radical of A to the problem of calculating the radical of a commutative algebra B which is a factor of a subalgebra of A. The algorithm performs $n^{O(1)}$ operations in K.*

It is known [88] that there is no algorithm based merely on the field operations which finds the irreducible factors of a polynomial $f(x) \in K[x]$ for a general field K. Even the weaker problem of finding the square-free part of $f(x)$ (the product of the irreducible factors of $f(x)$) appears to be unsolvable in this model (over a field of positive characteristic). In Section 4.5 we shall restrict ourselves to fields where this latter task can be effectively solved. For such a field K we denote by $SF_K(n)$ the number of arithmetical operations required to calculate the square-free part of a polynomial of degree n. If K is an arbitrary field of zero characteristic then the square-free part of $f(x)$ is simply $f(x)/gcd(f(x), f'(x))$ and hence $SF_K(n) = n^{1+o(1)}$ (cf. [14]). If K is a finite field then $SF_K(n) = n^{1+o(1)} + O(n \log |K|)$ (cf. [72]).

The probabilistic algorithm of Section 4.5 also assumes the presence of an auxiliary procedure which selects random elements from the algebra A independently. The distribution of the elements is typically concentrated on an appropriate finite subset of A. We do not require uniformity. Instead, we assume randomness in an algebraic sense explained below.

Let U be a finite dimensional vector space over K and let K' be an algebraic closure of K. Let $0 < \delta < 1$ and D be an integer. We say that a probability distribution on U satisfies condition $AlgRand(U, D, \delta)$ if for every nonzero polynomial function $f : K' \otimes_K U \to K'$ of degree at most D the probability of $f(u) = 0$ is at most δ. A possible way to obtain a distribution with $AlgRand(U, D, \delta)$ is the following. Assume that $u_1, \ldots, u_s$ form a K-linear generating system of U. Let Ω be a finite subset of K with $|\Omega| \geq \frac{D}{\delta}$. We take $u = \alpha_1 u_1 + \ldots + \alpha_s u_s$ where the coefficients $\alpha_1, \ldots, \alpha_s$ are drawn uniformly and independently from Ω. Then, by the Schwartz–Zippel Lemma (see Section 2.4), the probability of $f(u) = 0$ is at most $\frac{D}{|\Omega|} \leq \delta$.

We shall make use of the following lemma. The proof can be carried out by a simple induction on k. We omit the details of the proof which follows the lines of the most common proof of the Schwartz–Zippel Lemma.

Lemma 4.2. *Let $0 < \epsilon < 1$ be a real number. Let $f : U'^k \to K'$ be a nonzero polynomial function of degree at most D. Let $h = \lceil (\log k + \log \frac{1}{\epsilon}) / \log \frac{1}{\delta} \rceil$ and assume that the elements*

$u_{11}, \dots, u_{1h}, \dots, u_{k1}, \dots, u_{kh} \in U$ are chosen independently according to a probability distribution satisfying $AlgRand(U, D, \delta)$. Then with probability at least $1 - \epsilon$ there exist indices $j_1, \dots, j_k \in \{1, \dots, h\}$ such that $f(u_{1j_1}, \dots, u_{kj_k}) \neq 0$.

The probabilistic algorithm of Section 4.5 requires that the random elements of A are chosen according to a probability distribution satisfying condition $AlgRand(A, n^2, \delta)$ for a constant $0 < \delta < 1$, say $\delta = \frac{1}{2}$. Of course, in our construction above it is implicit that the ground field K is sufficiently large (namely $|K| \geq n^2/\delta$). The cost of selecting a single random element is denoted by $R(A)$. For a matrix algebra $A \leq M_n(K)$ given by algebra generators unfortunately no mathematically rigorous efficient random generator is known which satisfies the requirement unless we have a K-linear generating system $b_1, \dots, b_s$ of A and take a random linear combination of $b_1, \dots, b_s$. Then the cost $R(A)$ is $O(sn^2)$. However, there are heuristic random generators (e.g., the one used in the Meataxe procedure [52] for finding composition series of modules over finite algebras) appear to work well in practice for similar problems.

Theorem 4.3. *Let $A \leq M_n(K)$ be given by m generators and $0 < \epsilon < 1$. Then a system of matrices which generate $Rad(A)$ with probability at least $1 - \epsilon$ as an ideal of A can be computed by a probabilistic algorithm which performs $O((n + n^{\frac{1}{2}}m)(MM(n) + SF_K(n) + R(A))\mathrm{polylog}\, n \log \frac{1}{\epsilon})$ operations in K. If $A/Rad(A)$ is commutative then the algorithm performs $O(m(MM(n) + SF_K(n) + R(A))\mathrm{polylog}\, n \log \frac{1}{\epsilon})$ operations.*

Recall that $MM(n)$ stands for the cost of matrix multiplication, see Section 2.1. We stress that our probabilistic method is of Monte Carlo type: the algorithm may fail or even produce an incorrect output within a prescribed error probability ϵ.

4.1 Tori and maximal tori

The material of this section consists of an easy combination of known more or less elementary facts. However, we are unable to propose a single textbook where all the facts we need in the subsequent parts are stated. Therefore we formulate the less trivial facts in lemmas and give some hints to the proofs.

Let T_1 and T_2 be tori in A (see Section 2.2.5 for the definition) such that $T_1 \leq C_A(T_2)$. Then by Proposition 10.5c of [80], the subalgebra T generated by $T_1 \cup T_2$ is a torus as well. In particular, a commutative algebra contains a unique maximal torus. Furthermore, a maximal torus of A must contain the maximal torus of $Z(A)$. We call an element $a \in A$ semisimple (or separable) if a is contained in a torus $T \leq A$. This is equivalent to that the subalgebra $K[a]$ generated by a and 1_A is a torus. As $K[a] \cong K[x]/f(x)$ where $f(x)$ is the minimal polynomial of a this is further equivalent to that $f(x)$ is a separable polynomial, i.e., $gcd(f(x), f'(x)) = 1$. Assume that A is a commutative algebra. Then the unique maximal torus in A consists of the semisimple elements of A. Hence if A is a field, then the maximal torus of A is the separable closure of K in A.

Lemma 4.4. *Let A be a finite dimensional K-algebra and $T \leq A$ be a torus. Let $\phi : A \rightarrow Rad(A)$ stand for the natural projection. Assume further that A is a direct sum of ideals $A_1, \dots, A_r$ and $Z \geq K1_A$ is a subfield of $Z(A)$.*

(o) T is a maximal torus in A.

(i) $\phi(T)$ is a maximal torus of $A/Rad(A)$.

(ii) *$T \cap A_i$ is a maximal torus of A_i for $i = 1, \ldots, r$.*

(iii) *TZ, considered as a Z-algebra, is a maximal Z-torus of A and Z is a purely inseparable extension of $Z \cap T$.*

(iv) *T is a maximal torus of its centralizer $C_A(T)$.*

Proof. We only give proofs that both the conditions (i) and (iii) are equivalent to (o). The rest is easy and we leave the details to the reader. Assume that T is a maximal torus and $\widetilde{U}$ is a torus of $A/\mathrm{Rad}(A)$ containing $\phi(T)$. Then by Corollary 2.1, there exists a subalgebra $U \leq A$ isomorphic to $\widetilde{U}$ which contains T. Since T is a maximal torus we have $U = T$ whence $\widetilde{U} = \phi(T)$. Thus condition (i) is necessary. Sufficiency of (i) is obvious.

Concerning condition (iii), let Z_0 be the separable closure of K in Z. Then Z_0 is the unique maximal K-torus of Z and for every maximal K-torus T of A we have $Z \cap T = Z_0$. Let T be a K-torus of A containing Z_0 and let $T_1, \ldots, T_s$ be the simple components of T. Then the simple components of ZT are $ZT_1, \ldots, ZT_s$ and by Proposition 2.5.13 of [9], each ZT_i is a purely inseparable field extension of T_i of degree $\dim_{Z_0} Z = \dim_K Z / \dim_K Z_0$ as well as a separable extension of the field $Z \cap ZT_i \cong Z$. It follows that ZT is a torus over Z and $\dim_K ZT / \dim_K T = \dim_K Z / \dim_K Z_0$, a ratio independent of T. From this it is immediate that if T is not a maximal K-torus then ZT is not a maximal Z-torus either.

To prove the reverse implication, let U be a Z-torus of A containing TZ and let $U_1, \ldots, U_t$ be the simple components of U. Then each U_i is a separable field extension of Z. Also, the unique maximal K-torus W of U is the sum of $W_1, \ldots, W_t$ where W_i is the separable closure of K in U_i. By Proposition 2.5.13 of [9], $\dim_K U_i / \dim_K W_i = \dim_K Z / \dim_K Z_0$. From this we infer $\dim_K U / \dim_K W = \dim_K Z / \dim_K Z_0 = \dim_K ZT / \dim_K T$. Hence if T is maximal then $T = W$ and $\dim_K U = \dim_K ZT$ whence $U = ZT$ for every Z-torus U containing ZT. □

Lemma 4.5. *Assume that T is a maximal torus of A. Then $C_A(T)/Rad(C_A(T))$ is commutative.*

Proof. By Lemma 4.4, (i) and (iv), $T + \mathrm{Rad}(C_A(T))$ is a maximal torus of the factor $C_A(T)/\mathrm{Rad}(C_A(T))$. Replacing A with $C_A(T)/\mathrm{Rad}(C_A(T))$ we have to show that if A is semisimple and $T \leq Z(A)$ is a maximal torus of A then A is commutative. In view of Lemma 4.4, (ii), we may further assume that A is simple. Assume that A is not a division algebra. Then there exists an idempotent $e \in A$ such that $e \notin Z(A)$. The subalgebra B generated by T and e is a torus properly containing T (because $B \cong T[x]/(x^2 - x)$), a contradiction. It remains to eliminate the case when A is a noncommutative division algebra. Then by Lemma 13.5 of [80], there exists a subfield $L \leq A$ which is a proper separable extension of $Z(A)$. Then the separable closure of K in L is a torus properly containing T, a contradiction. □

Lemma 4.6. *Assume that T is a maximal K-torus of A and let K' be an arbitrary field extension of K. Then $K' \otimes_K T$ is a maximal K'-torus of $K' \otimes_K A$.*

Proof. Let $A' = K \otimes_K A$, $T' = K' \otimes_K T$, $H = C_A(T)$, and $H' = C_{A'}(T')$. It is obvious that that $H' = K' \otimes_K H$. Let $I' = K' \otimes_K \mathrm{Rad}(H)$. Then I' is a nilpotent ideal of H'.

We claim that it is sufficient to show that $T' + I'$ is a maximal torus in H'/I'. Indeed, if $U' \geq T'$ is a torus of A' then $U' + I'$ is a torus of H'/I' whence by the maximality of $T' + I'$ we have $U' + I' \leq T' + I'$. On the other hand $U' \cap I'$ is a nilpotent ideal of U' which must be zero as U' is separable. From this it is immediate that $U' \leq T'$.

By the claim, we can work with $H/\mathrm{Rad}(H)$ in place of H, i.e., we may assume that H is a commutative semisimple algebra. If $\operatorname{char} K = 0$ then $H = T$ and $H' = T'$ therefore the assertion is obvious. Assume that $\operatorname{char} K = p > 0$. Let $H_1, \ldots, H_r$ be the simple components of T. Then T is the sum of the separable closures of K is H_i. In particular, by Corollary 2.5.14 of [9], T is the linear span over K of $\{a^{p^l} | a \in H\}$ where l is a sufficiently large integer. By the commutativity of H' the subalgebra T' is the linear span over K' of $\{a^{p^l} | a \in H'\}$. Assume that U' is a torus of H'. By Lemma 4.4(ii) and again by Corollary 2.5.14 of [9], $\{u^{p^l} | u \in U'\}$ span U' over K'. We obtained $U' \leq T'$. □

Lemma 4.7. *Assume that T is a maximal torus in a semisimple algebra A. Then $C_A(T) = TZ(A)$. Furthermore, if A is a central simple K-algebra then* $\dim_K T = \sqrt{\dim_K A}$.

Proof. We only give a proof of the first statement. The second assertion can be proved in a similar fashion. It is obvious that $C_A(T) \geq TZ(A)$. In view of Lemma 4.4(ii) it is sufficient to give a proof of the statement in the special case where A is simple. Then $TZ(A)$, considered as a $Z(A)$-algebra is a maximal $Z(A)$-torus of A. Replacing K with $Z(A)$ we assume that A is a central simple K-algebra. Let K' be the algebraic closure of K. By Lemma 4.6, $T' = K' \otimes_K T$ is a maximal K'-torus in $A' = K' \otimes_K A$. Obviously $C'_A(T') = K' \otimes_K C_A(T)$. On the other hand, $A' \cong M_d(K)$ for some integer d. In $M_d(K)$ every maximal torus is conjugate to $Diag_d(K)$, the subalgebra of $d \times d$ diagonal matrices and it is straightforward to verify the assertion for $Diag_d(K)$. □

4.1.1 Centralizers of tori and Fitting decompositions

Let T be a torus over K with $d = \dim_K T$. We recall some facts from Section 10.2 of [80], specialized to the context of tori. Let T be a torus over K. The map $\mu : T \otimes_K T \to T$ given by the law $\mu(a \otimes b) = ab$ is a K-algebra epimorphism. The kernel of μ is the ideal of $T \otimes_K T$ generated by the elements $b \otimes 1 - 1 \otimes b$ where b runs over T (or, equivalently, on a basis of T). Let $I = \{u \in T | u \ker \mu = 0\}$ be the ideal of $T \otimes_K T$ complementary to $\ker \mu$. Then $T \otimes_K T = I \oplus \ker \mu$ and the restriction of μ establishes an algebra isomorphism $I \cong T$. Let Φ_T stand for the identity element of I. Note that Φ_T is characterized by the properties $\mu(\Phi_T) = 1$ and $(1 \otimes b)\Phi_T = (b \otimes 1)\Phi_T$ for every $b \in T$. (We remark that this is the definition of a separating idempotent. Separating idempotent exists for an arbitrary separable algebra. However, in the noncommutative case it is not necessarily unique.)

Let U be a $T \otimes_K T$-module. Then for every $u \in U$ we have $u = \Phi_T u + (1 - \Phi_T)u$. This gives rise to a decomposition of U as the direct sum of submodules $U_0 = IU = \Phi_T U$ and $U_1 = (\ker \mu)U = (1 - \Phi_T)U$. Then Φ_T and $1 - \Phi_T$ act as the projections of U to U_0 and U_1 with respect to the decomposition $U = U_0 \oplus U_1$. We have $U_0 = \{u \in U | (\ker \mu)U = (0)\} = \{u \in U | (b \otimes 1 - 1 \otimes b)u = 0 \text{ for every } b \in T\}$ and $U_1 = (\ker \mu)U = \{b \otimes 1 - 1 \otimes b | b \in T\}(T \otimes_K T)U = \{b \otimes 1 - 1 \otimes b | b \in T\}U$. We refer to U_0 as the Fitting null component and to U_1 as the Fitting one component. This terminology is justified by the following. The adjoint action of $b \in T$ on U is defined as the linear transformation $\mathrm{ad}b : u \mapsto (b \otimes 1 - 1 \otimes b)u$. This gives rise to a representation of T considered as an abelian (and hence nilpotent) Lie algebra and the decomposition $U = U_0 + U_1$ appears to be the same as the Fitting decomposition of U given in Theorem II.4 of [63].

Now assume that the torus T is a subalgebra of the algebra A. We consider A as a $T \otimes_K T$-module in the natural way (multiplication from both sides). Then the Fitting null component A_0 is $\{a \in A | (b \otimes 1 - 1 \otimes b)a = [b, a] = 0 \text{ for every } b \in T\} = C_A(T)$ while the Fitting one component A_1 is the linear span of the elements of the form $(b \otimes 1 - 1 \otimes b)a = [b, a]$

$(a \in A, b \in T)$. Thus $A = C_A(T) + [T, A]$, a direct sum of vector spaces and the projections of A corresponding to this decomposition are Φ_T and the map $a \mapsto a - \Phi_T(a)$.

Similarly, let U and W be T-modules. For convenience we consider U and W as right T-modules. For $u \in U$, $c \in Hom_K(U, V)$ and $a, b \in T$ let $((a \otimes b)c)u := acbu$. The linear extension of this rule to $T \otimes T$ makes $Hom_K(U, V)$ a $T \otimes T$-module. Then the Fitting null component of $Hom_K(U, V)$ is $Hom_T(U, V)$.

We also need an explicit representation of Φ_T in terms of rank one tensors. This appears to be extremely useful for computational purposes. We use a construction which can be extended to the more general context of Frobenius algebras, cf. Theorem 62.11 of [22]. For an application in computational group theory we refer the reader to [5, 35]. Since the formulations appearing in the literature are slightly different from that we need, we give some hints to an easy proof of correctness of the construction in the special case of tori.

For $a \in T$ let $Tr(a)$ stand for the trace of the linear transformation of T given as $b \mapsto ab$. By Proposition 3.8.7 of [9], the separability of T implies that the linear function $Tr : T \to K$ is not identically zero. As a consequence, the the bilinear trace form $(,)$ on T given as $(a, b) = Tr(ab)$ is a non-degenerate bilinear form on T. Let $b_1, \dots, b_d$ be an arbitrary basis of T and $b'_1 \dots, b'_d$ be the dual basis with respect to the form $(,)$. We claim that

$$\Phi_T = \sum_{i=1}^{d} b_i \otimes b'_i. \tag{4.1}$$

To see this we note first that it is straightforward to verify that the element $f = \sum_{i=1}^d b_i \otimes b'_i \in T \otimes_K T$ does not depend on the particular choice of the basis $b_1, \dots, b_d$. Let K' be the algebraic closure of K and $T' = K' \otimes_K T$. We think of T as embedded in T' in the natural way. It is obvious that Φ_T satisfies the conditions for $\Phi_{T'}$, and hence $\Phi_{T'} = \Phi_T$. On the other hand, we know that T' is isomorphic to the direct sum of d copies of K'. Let $e_1, \dots, e_d$ be the identity elements of the simple components of T'. Then $e_1, \dots, e_d$ form a self-dual basis of T' with respect to the bilinear trace form and hence $f = \sum_{i=1}^d e_i \otimes e_i$. One easily verifies that $\sum_{i=1}^d e_i \otimes e_i$ also satisfies the properties characterizing $\Phi_{T'}$. Thus $\Phi_T = \Phi_{T'} = \sum_{i=1}^d e_i \otimes e_i = f$.

4.2 Decomposition with respect to a maximal torus

In this section we develop a structure theory which serves as a theoretical foundation for the subsequent algorithms. First we fix some notation. Let K be an arbitrary field. We denote by ϕ the natural projection $A \to A/\text{Rad}(A)$. Let $\widetilde{C}$ be the set of those central elements of $A/\text{Rad}(A)$ which are separable over K. Obviously, $\widetilde{C}$ is the unique maximal torus of $Z(A/\text{Rad}(A))$. Let T be a fixed maximal torus of A and let the set $C \subseteq T$ consist of those elements of T which are central modulo the radical:

$$C = \{x \in T | \phi(x) \in Z(A/\text{Rad}(A))\}.$$

C is a subalgebra of T as C is the intersection of T and the subalgebra $\phi^{-1}(Z(A/\text{Rad}(A)))$. By Lemma 4.4, $\phi(T)$ is a maximal torus of $A/\text{Rad}(A)$ and hence

$$\phi(C) = \phi(T) \cap Z(A/\text{Rad}(A)) = \widetilde{C}.$$

In view of Subsection 4.1.1,

$$A = S + N, \tag{4.2}$$

where $S = C_A(C)$ and $N = [C, A]$. We remark that, by Wedderburn–Malcev, applied to the algebra $\phi^{-1}(\widetilde{C})$, the subalgebra C (and hence S) is determined up to conjugation by a unit in A. Therefore the structural properties of S and N are independent of the particular choice of T.

Proposition 4.8. *N is an S-invariant subspace of Rad(A), i.e., $SN \subseteq N$, $NS \subseteq N$, and $N \subseteq$ Rad(A),*

Proof. The inclusions $SN \subseteq N$ and $NS \subseteq N$ follow from $s[x, y] = sxy - syx = xsy - syx = [x, sy]$ and $[x, y]s = xys - yxs = xys - ysx = [x, ys]$, respectively $(s \in S, x \in C, y \in A)$. To prove the remaining inclusion, observe that $\phi(C) = \widetilde{C}$ is in the center of $A/\mathrm{Rad}(A)$. From this we immediately obtain that $\phi(N) = [\phi(C), \phi(A)] = (0)$, whence $N \subseteq \mathrm{Rad}(A)$. □

The radical inherits the decomposition (4.2) of A in the following sense.

Proposition 4.9. *Rad(A) = Rad$(S) + N$.*

Proof. $A\mathrm{Rad}(S) = (S + N)\mathrm{Rad}(S) = \mathrm{Rad}(S) + N\mathrm{Rad}(S) \subseteq \mathrm{Rad}(S) + N \subseteq \mathrm{Rad}(S) + \mathrm{Rad}(A)$, hence the element as is nilpotent for every $a \in A$ and $s \in \mathrm{Rad}(S)$. This implies the inclusion $\mathrm{Rad}(S) + N \subseteq \mathrm{Rad}(A)$. To prove the reverse inclusion let $a \in \mathrm{Rad}(A)$. Then $a = s + n$ for some $s \in S$ and $n \in N$. We have $s = a - n \in (\mathrm{Rad}(A) + N) \cap S = \mathrm{Rad}(A) \cap S \subseteq \mathrm{Rad}(S)$, whence $a \in \mathrm{Rad}(S) + N$. □

A part of the next statement asserts that the radical part of the primary decomposition of S is zero. Actually, the subspace N' in the primary decomposition (Subsection 2.2.3) is a subspace of N in (4.2) and S is a subalgebra of the sum of the primary components of A.

Proposition 4.10. *Let $C_1, \ldots, C_r$ be the simple components of C. Then S is the direct sum $S = S_1 + \ldots + S_r$ of ideals $S_1 = C_1 S, \ldots, S_r = C_r S$. For every $i \in \{1, \ldots, r\}$ the factor algebra $S_i/$Rad(S_i) is a simple algebra. Furthermore, if we consider S_i as a C_i-algebra in the natural way then $Z(S_i/$Rad$(S_i))$ is a purely inseparable extension of C_i.*

Proof. To see the first two assertions, we use the the primary decomposition of S, see Subsection 2.2.3. Let $e_i \in C_i$ $(i = 1, \ldots, r)$ be the primitive idempotents of C and put $S_i = e_i S$. As $e_i S = S e_i$, we have $e_i S e_i = S_i$ and $\sum_{i \neq j} e_i S s_j = 0$.

To see the last assertion, let $Z_1, \ldots, Z_s$ be the simple components of $S/\mathrm{Rad}(S)$. Then the simple components of $\phi(C)$, the image of C at the the natural projection $\phi : S \rightarrow \mathrm{Rad}(S)$, are $Z_i \cap \phi(C)$. It follows that (after re-indexing) $Z_i \cap \phi(C) = \phi(C_i)$. Hence $\phi(C_i)S = \phi(C_i S) = Z_i \phi(S)$. Since $Z_i \phi(S)$ are the simple components of $S/\mathrm{Rad}(S)$ we obtained that $S_i/\mathrm{Rad}(S_i) \cong \phi(S_i)$ are simple. Also, as $\phi(C_i)$ is the set of elements of Z_i which are separable over K, $Z_i = Z(Z_i \phi(S))$ is purely inseparable over $\phi(C_i)$. □

Let $H = C_A(T)$, the centralizer of T. Obviously, H is a subalgebra of S. We remark that, by Theorem. 4.4.8 of [94], H is a Cartan subalgebra of A (considered as a Lie algebra).

Theorem 4.11. *Keeping the notation introduced above, Rad(S) is the ideal of S generated by Rad(H), that is, Rad$(S) = S$Rad$(H)S$. Furthermore, every nilpotent element of H is in Rad(H).*

Proof. It is clearly sufficient to prove the assertions for the primary components of S separately. Therefore we assume that S is primary, i.e., C is a field. We can further consider S as a C-algebra rather than as a K-algebra. Thus it is sufficient to consider an algebra S where $\widetilde{Z} = Z(S/\mathrm{Rad}(S))$ is a purely inseparable field extension of K.

First we show that every nilpotent element of H is in the radical of S. To see this, let h be an arbitrary nilpotent element of H. Let $\widetilde{T}$ denote the image of T at the natural projection $\phi : S \to S/\mathrm{Rad}(S)$. By Lemma 4.4, (i), $\widetilde{T}$ is a maximal torus of $S/\mathrm{Rad}(S)$. Consider the centralizer $\widetilde{U}$ of the algebra $\widetilde{T}$ in $S/\mathrm{Rad}(S)$. Obviously $\phi(h)$ is a nilpotent element of $\widetilde{U}$. By Lemma 4.7, $\widetilde{U} = \widetilde{T}Z(S/\mathrm{Rad}(S))$ is a commutative semisimple algebra. Since in a commutative algebra every nilpotent element is in the radical, $\phi(h) \in \mathrm{Rad}(\widetilde{U}) = (0)$. This implies the last statement of the theorem together with the inclusion $\mathrm{Rad}(H) \subseteq \mathrm{Rad}(S)$. From this $S\mathrm{Rad}(H)S \subseteq \mathrm{Rad}(S)$ is immediate.

To prove the reverse inclusion, let K' be the separable algebraic closure of K, $S' = K' \otimes_K S$, $T' = K' \otimes_K T$, and $H' = K' \otimes_K H$. Then, by Lemma 4.6, T' is a maximal torus in the K'-algebra S' and H' is the centralizer of T' in S'. Let $I' = K' \otimes_K \mathrm{Rad}(A)$. Then $K' \otimes_K \mathrm{Rad}(H) = I' \cap H' = C_{I'}(T')$. Since K' is a separable extension of K, $\mathrm{Rad}(S') = K' \otimes_K \mathrm{Rad}(S)$ and $\mathrm{Rad}(H') = K' \otimes_K \mathrm{Rad}(H)$, therefore we are done if we prove that $\mathrm{Rad}(S') = S'\mathrm{Rad}(H)S'$. In order to simplify notation, we replace K with K', S with S', etc.

Let $\widetilde{T}$ denote the image of T at the natural projection $\phi : S \to S/\mathrm{Rad}(S)$. The algebra $S/\mathrm{Rad}(S)$ is a central simple $\widetilde{Z}$-algebra. By Theorem 13.5 of [80], there exists a finite separable field extension L of Z such that $L \otimes_Z S/\mathrm{Rad}(S) \cong M_d(L)$ for some integer d. Since $\widetilde{Z}$ is a purely inseparable extension of K which is closed under finite separable extensions, so is $\widetilde{Z}$ and hence $L = Z$. This implies $S/\mathrm{Rad}(S) \cong M_d(\widetilde{Z})$. Obviously $\widetilde{T}\widetilde{Z}$ is a torus of $S/\mathrm{Rad}(S)$. Since the minimal polynomial of every element of $\widetilde{T}\widetilde{Z}$ splits into linear factors over $\widetilde{Z}$, by Proposition 1.4.4 of [94], $\widetilde{T}\widetilde{Z}$ considered as a subalgebra of $M_d(\widetilde{Z})$ is conjugate in $M_d(\widetilde{Z})$ to a subalgebra of $Diag_d(\widetilde{Z})$. In other words, there exists a $\widetilde{Z}$-algebra isomorphism $\psi : S/\mathrm{Rad}(S) \cong M_d(\widetilde{Z})$ such that $\psi(\widetilde{T}\widetilde{Z}) \leq Diag_d(\widetilde{Z})$. Since $\psi(\widetilde{T})$ is the set of elements of $\psi(\widetilde{T}\widetilde{Z})$ which are separable over K, we have $\psi(\widetilde{T}) \leq Diag_d(K)$, the subalgebra of $M_d(\widetilde{Z})$ of diagonal matrices with entries from K. In particular, $\psi(\widetilde{T}) \leq M_d(K)$. On the right hand side of the inclusion stands a central simple K-subalgebra of $M_d(\widetilde{Z})$. Then $\widetilde{D} = \psi^{-1}M_d(K)$ is a central simple (and hence separable) K-subalgebra of $S/\mathrm{Rad}(S)$ containing $\widetilde{T}$ as a maximal torus. Corollary 2.1 implies the existence of a central simple K-subalgebra D of S containing T.

We are going to show the equality $\mathrm{Rad}(S) = D\mathrm{Rad}(H)D$. To this end we consider $\mathrm{Rad}(S)$ as a module over $D \otimes_K D^{op}$, where D^{op} is the algebra opposite to D. By Proposition 12.4b of [80], $D \otimes_K D^{op} \cong M_{d^2}(K)$ ($d = \dim_K T$), which is a simple algebra. In particular, every simple $D \otimes_K D^{op}$-module is isomorphic to the module D with multiplication law $(d_1 \otimes d_2)v = d_1 v d_2$ (cf. Corollary 12.3 of [80]). Obviously, this module is generated by the identity element 1_D of D which belongs to the subspace $\{v \in D | (1 \otimes a)v = (a \otimes 1)v \text{ for every } a \in D\} = Z(D)$. Now $\mathrm{Rad}(S)$, being a unital $D \otimes_K D^{op}$-module, can be decomposed into a direct sum of simple modules. The preceding observation, applied to the simple components, implies that $\mathrm{Rad}(S)$ is generated by the subspace $\{v \in \mathrm{Rad}(S) | (1 \otimes a)v = (a \otimes 1)v \text{ for every } a \in D\} = \{v \in \mathrm{Rad}(S) | va = av \text{ for every } a \in D\} = C_{\mathrm{Rad}(S)}(D) \leq C_{\mathrm{Rad}(S)}(T) = \mathrm{Rad}(H)$. This concludes the proof of the theorem. □

We shall make use of the following characterization of C which will enable us to compute C without calculating $\mathrm{Rad}(A)$ first.

Theorem 4.12. *Set $L = [A, A] \cap T$. Then L is a linear subspace of T and $C = \{x \in T | xL \subseteq L\}$.*

Proof. It is obvious that L is a linear subspace of T. Let $C_1 = \{x \in T | xL \subseteq L\}$. The inclusion $C \subseteq C_1$ follows easily from $C[A, A] = [CA, A] = [A, A]$. We have to show that $\dim_K C_1 \leq \dim_K C$.

We claim that $L = ([A, A] + \mathrm{Rad}(A)) \cap T$. The inclusion $L \subseteq ([A, A] + \mathrm{Rad}(A)) \cap T$ is obvious. To prove the reverse inclusion, let K' be the algebraic closure of K, $A' = K' \otimes_K A$, $T' = K' \otimes_K A$, and $L' = K' \otimes_K L$. Obviously $[A', A'] = K' \otimes_L [A, A]$ and hence $L' = [A', A'] \cap T'$. We have to show that $L' \supseteq [A', A'] + K' \otimes_K \mathrm{Rad}(A) \cap T'$. In view of $K' \otimes_K \mathrm{Rad}(A') \supseteq \mathrm{Rad}(A')$ it is sufficient to establish the inclusion $L' \supseteq ([A', A'] + \mathrm{Rad}(A')) \cap T'$. Since T' is a torus in A' and K' is perfect, by Corollary 2.1 there exists a subalgebra D' which contains T' and is isomorphic to $A'/\mathrm{Rad}(A')$. Obviously $[A', A'] + \mathrm{Rad}(A') = [D', D'] + \mathrm{Rad}(A')$. From $D \cap \mathrm{Rad}(A) = (0)$ and $[D', D'] \subseteq D'$ we infer that $D \cap [A', A'] + \mathrm{Rad}(A') = [D', D']$. As $T' \leq D'$ we have $T' \cap ([A', A'] + \mathrm{Rad}(A')) = [D', D'] \cap T' \subseteq [A', A'] \cap T' = L'$.

By the claim it is sufficient to verify the assertion modulo $\mathrm{Rad}(A)$. Furthermore, we can work separately in the simple components of $A/\mathrm{Rad}(A)$. Thus for the rest of the proof we may assume that A is a simple algebra. Then $Z = Z(A)$ is a purely inseparable extension of C. As $C_1 = ZT \cap T$ and $C = Z \cap T$ it is sufficient to establish the inequality $\dim_K ZC_1 \leq \dim_K Z$. Observe that $ZC_1 \subseteq \{x \in ZT | xZL \subseteq ZL\}$ and $ZL = [A, A] \cap ZT$. Consider A as a central simple algebra over Z. Then ZT is a maximal Z-torus in A and it is sufficient to show that $\dim_Z \{x \in ZT | xZL \subseteq ZL\} \leq 1$ In order to simplify notation, we write K in place of Z, T in place of ZT and ZL in place of L. Then it remains to prove $\dim_K C_1 \leq 1$ in the special case where A is a central simple algebra over K. It is also clear that we may assume that K is algebraically closed.

Then we can identify A with the full matrix algebra $M_d(K)$ where $d = \dim_K T$ and T can be identified with $Diag_d(K)$, the algebra of diagonal matrices. For an arbitrary element $x \in A$ let $\mathrm{Tr(x)}$ stand for the trace of x as a d by d matrix. It is well known that $[A, A] = \{x \in A | \mathrm{Tr(x)} = 0\}$ even if the characteristic is positive. (Both subspaces have codimension one.) From this fact we infer $L = \{x \in T | \mathrm{Tr}(x) = 0\}$. Observe that the bilinear form $< x, y >= \mathrm{Tr}(xy)$ is non-degenerate on T. The preceding characterization of L implies that C_1 is the orthocomplement of L in T with respect to the bilinear trace form, therefore $\dim C_1 = 1$, concluding the proof of the theorem. □

4.3 A reduction to the commutative case

This section is devoted to the proof of Theorem 4.1. Recall that the algebra $A \leq M_n(K)$ is assumed to be given by generators. That is, the input consists of matrices $g_1, \ldots, g_m \in M_n(K)$ and A is the algebra generated by $g_1, \ldots, g_m$ and the identity matrix. The output is expected to be an array $a_1, \ldots, a_t$ of matrices from $\mathrm{Rad}(A)$ such that the ideal of A generated by $a_1, \ldots, a_t$ is $\mathrm{Rad}(A)$.

Proof of Theorem 4.1. We can can calculate a basis $b_1, \ldots, b_s$ of A by a straightforward method with $n^{O(1)}$ operations in K. Then we find a K-basis $u_1, \ldots, u_d$ of a maximal torus T of A using the method of [45]. at s cost of $n^{O(1)}$ operations. We use the notation of

Section 4.2. Calculation of a K-basis of the centralizer $H = C_A(T)$ can be accomplished by solving the system of homogeneous linear equations $xu_i - u_ix = 0$ $(i = 1, \ldots, d)$ in A.

We find the subalgebra C using the characterization given in Theorem 4.12. We select a linear basis of the subspace $[A, A]$ from the commutators $[b_i, b_j]$ $(i, j = 1, \ldots, s)$ and calculate a basis of the intersection $L = T \cap [A, A]$ and then a basis $c_1, \ldots, c_k$ of the stabilizer $C = \{x \in T | xL \subseteq L\}$. Both tasks can be accomplished with $n^{O(1)}$ operations by solving systems of linear equations. We omit the details.

Now a basis of N can be selected from the commutators $[b_i, c_j]$ $(i = 1, \ldots, t,\ j = 1, \ldots, s)$. We can select a basis of the ideal $I = H[H, H]H$ in a similar way. Also, we can find a basis of the factor algebra $H_1 = H/I$ together with the multiplication table of H/I with respect to that basis. By Lemma 4.5, $H/\mathrm{Rad}(H)$ is commutative, therefore $I \leq \mathrm{Rad}(H)$ and hence H_1 is a commutative algebra. We pass the multiplication table of H_1 to the oracle for finding the radical of commutative algebras. Then $\mathrm{Rad}(H)$ is generated by a basis of $[H, H]$ and a system of representatives of the generators of H_1. These together with the basis of N generate $\mathrm{Rad}(A)$ as an ideal of A by Proposition 4.9 and Theorem 4.11. □

We remark that a nilpotent ideal J of B together with a presentation of B/J in terms of $s = O(\log_p \dim_K B)$ generators can be calculated in a rather straightforward way. (The technical details appear to be too complicated to include here.) Therefore computing $\mathrm{Rad}(B)$ can actually be reduced to calculating the radical of an ideal in the polynomial ring $K[x_1, \ldots, x_s]$.

4.4 Computing Fitting decomposition with respect to a semisimple matrix

Let $u \in M_m(K)$ be a semisimple matrix and T be the torus generated by u and the identity matrix. Let Φ_T stand for the element of $T \otimes_K T$ given in Subsection 4.1.1. Our aim is to calculate $\Phi_T a$ efficiently for an arbitrary matrix $a \in M_n(K)$. We know that $T \cong K[x]/(f(x))$ where $f(x)$ is the minimal polynomial of u. Let $V = K^n$, the vector space of column vectors of length n over K. We consider V as a T-module or, equivalently, as a $K[x]$-module. Then $C_{M_n(K)}(T) \cong End_T(V) \cong End_{K(x)}(V)$.

We presume that we have found a decomposition of V as a direct sum of cyclic T-submodules $V_1, \ldots, V_t$ such that for any pair V_i, V_j of components either $V_i \cong V_j$ as T-modules or $Hom_T(V_j, V_j) = (0)$. Then $End_K(V) = \bigoplus_{i,j} Hom_K(V_i, V_j)$ and for $a = \sum_{i,j} a_{ij}$ where $a_{ij} \in Hom_K(V_i, V_j)$ we have $\Phi_T a = \sum_{i,j} \Phi_T a_{ij}$. For non-isomorphic V_i, V_j we know that $Hom_K(V_i, V_j) = (0)$ therefore Φ_T is zero on $Hom_K(V_i, V_j)$. For isomorphic V_i and V_j we identify V_i and V_j using a T-module isomorphism.

The main task is computing Φ_T on a cyclic T-module W. Let $d = \dim_K W$. We may assume that T acts faithfully on W. Indeed, if I is an ideal of T such that $IW = (0)$ then $\Phi_{T/I}$ and Φ_T coincide on $End_K(W)$. We identify T with a subalgebra of $End_K(W)$. A faithful cyclic T-module is isomorphic to the regular module $T \cong K[x]/(g(x))$ where $g(x)$ is the minimal polynomial of the generator u on W. These isomorphisms (provided that we constructed it effectively) will allow us to perform a multiplication in T as well as multiplication of a vector and a matrix from T with $O(d\mathrm{polylog}\, d)$ operations in K using polynomial arithmetic modulo $g(x)$ (cf. Section 1.3 of [14]). The isomorphisms are assumed to be given as follows. Let w be a vector which generates W as a T-module. We work in the basis $w_i = u^i w$ $(i = 0, \dots, d-1)$ of W. Note that in this basis the coefficients of $g(x)$ can be read from the last column of the matrix of u. In T we use the basis $1, u, \dots, u^{d-1}$. If we have an element $a \in T$ represented as a matrix in terms of the basis $w_0, \dots, w_{d-1}$ then the coordinates of a with respect to the basis $1, \dots, u^{d-1}$ can be read from the vector aw_0 which is the first column of the matrix. Conversely, if a is given as $\sum_i \alpha_i u^i$ then the columns of the matrix of a are the vectors aw_i. Hence the matrix of a can be calculated at total cost $O(d^2 \text{polylog}\, d)$.

From $\dim_K W = \dim_K T$ we infer $End_T(W) = T$. (This can be seen by noting that over the algebraic closure of K the torus T is conjugate to the algebra of the diagonal matrices.) Hence we know that $\Phi_T a \in T$ for every element $a \in End_T(W)$. In view of the preceding discussion we have to show how to calculate $\Phi_T a w_0$ efficiently. In view of formula (4.1), we have $\Phi_T a = \sum_{i=0}^{d-1} u_i' a u^i$ where u_i' is the basis of T dual with respect to the trace form. Representation of u_i' in terms of $1, \dots, u^{d-1}$ can be obtained as the rows of the inverse of the matrix $(Tr(u^i u^j))_{i,j=0}^{d-1}$. The matrix $(Tr(u^i u^j))_{i,j=0}^{d-1}$ and its inverse can be calculated using $O(d^2 \text{polylog}\, d)$ operations using the method described as a part of Algorithm 2.6.1 in [14]. Observe that $au^i w_0 = aw_i$, which is the ith column of the matrix of a. Then for every $i \in \{0, \dots, d-1\}$ the vector $u_i' a u_i w_0$ can be calculated with $O(d \text{polylog}\, d)$ operations using polynomial arithmetic modulo $g(x)$. The total cost of computing $\Phi_T a$ (on a cyclic module with the presumed basis) is therefore $O(d^2 \text{polylog}\, d)$.

We return to determining $\Phi_T a$ on the whole V. Assume that we have a basis

$$v_{11}, \dots, v_{1d_1}, \dots, v_{t1}, \dots, v_{dt}$$

such that the subspaces V_i spanned by $v_{i1}, \dots, v_{id_j}$ are cyclic T-submodules such that V_i and V_j are either isomorphic T-modules or $Hom_T(V_i, V_j) = (0)$ $(i, j = 1, \dots, t)$ and the basis given on V_i is of the form $v_{ik} = u^{k-1} v_{i1}$ $(i = 1, \dots, t,\ k = 1, \dots, d_i)$. Then for every matrix $a \in M_n(K)$ writing a in terms of the new basis (i.e. conjugating a by the basis transition matrix) can be accomplished with $O(MM(n))$ operations. Then we calculate $\Phi_T a$ block-wise. The total cost of this amounts to $O(n^2 \text{polylog}\, n) = O(MM(n) \text{polylog}\, n)$ operations. Writing the result back in terms of the standard basis of V requires further $O(MM(n))$ operations.

It remains to show how to find a basis with the required properties. We follow the method of Giesbrecht for calculating the rational Jordan form, cf. [41]. However, here we are not allowed to factor the minimal polynomial. Recall that the the companion matrix $Comp(g(x))$ of a monic polynomial $g(x) \in K[x]$ of degree d is the matrix of the action of x on the $K[x]$-module $K[x]/(g(x))$. For every matrix $u \in M_n(K)$ there exists a unique block diagonal matrix $Frob(u)$ similar to u which is composed from the companion matrices of polynomials $f_1(x), \dots, f_s(x) \in K[x]$ satisfying $f_s(x) | f_{s-1}(x) | \cdots | f_1(x)$. The polynomials $f_1(x), \dots, f_s(x)$ are called the invariant factors of u and the matrix $Frob(u)$ is called the Frobenius form of u. Obviously $f_1(x)$ is the minimal polynomial of u and $f_1(x) \cdots f_s(x)$ is the characteristic polynomial of u. The Frobenius form $Frob(u)$ together with a matrix $b' \in GL_n(K)$ such that $b'^{-1} u b' = Frob(u)$ can be computed with $O(MM(n) \text{polylog}\, n)$ operations in K using the Las Vegas algorithm of Giesbrecht [41]. Let $f_1(x), \dots, f_s(x)$ be the invariant factors of U. Set $f_{s+1}(x) = 1$ and let $g_1(x), \dots, g_r(x)$ be the collection of non-constant quotients of the form $f_i(x)/f_{i+1}(x)$ $(i = 1, \dots, s)$. In our case where u is semisimple and therefore $f_1(x)$ is square-free we have $f_1(x) = g_1(x) \cdots g_r(x)$. Furthermore, it is easy to see that u is similar to the block diagonal matrix u' composed of s_1 companion matrices of $f_1(x)$, s_2 companion matrices of $f_2(x)$, and so on. The multiplicities s_i are determined by $\prod_{i=1}^{s} f_i(x) = \prod_{i=1}^{r} g_i(x)^{s_i}$. Since u' is similar to u, we have $Frob(u') = Frob(u)$ and, again by the method of Giesbrecht, we can calculate a matrix $b'' \in GL_n(K)$ such that $b''^{-1}u'b'' = Frob(u)$. With $b = b'b''^{-1}$ we have $u' = b^{-1}ub$. Now the columns of the matrix b form a basis with the required properties. The total cost amounts to $O(MM(n)\mathrm{polylog}\, n)$ operations. We have proved the following.

Proposition 4.13. *Let $u \in M_n(K)$ be a semisimple matrix and let T be the matrix algebra generated by u and the identity matrix. Let $a \in M_n(K)$ be an arbitrary matrix. Then $\Phi_T a$ can be calculated by a Las Vegas algorithm using $O(MM(n)\mathrm{polylog}\, n)$ operations in K.*

4.5 A Monte Carlo method for finding the radical

In this section we prove Theorem 4.3. Throughout the section we assume that K is a sufficiently large perfect field together with an efficient method for finding the square-free part of polynomials of degree n with $SF_K(n)$ operations. Also, K' stands for an algebraic closure of K and $A' = K' \otimes_K A$. We think of A as embedded into A'. The input is the same as described in Section 4.3. We assume that random elements of A are generated independently according to a distribution satisfying condition $AlgRand(A, n^2, \delta)$ defined in the introductory part of this chapter. The cost of selecting a single random element of A is denoted by $R(A)$. The algorithm follows the lines of the method described in Section 4.3. We describe the main ingredients using the notation of Section 4.2.

4.5.1 Jordan decomposition

Let $u \in M_n(K)$ be a matrix. Since K is perfect, there exists a semisimple matrix $u_s \in M_n(K)$ and a nilpotent matrix $u_n \in M_n(K)$ such that $[u_s, u_n] = 0$ and $u = u_s + u_n$ (cf. Propositions 1.4.6 and 1.4.10 of [94]). Furthermore, u_s and u_n are unique with these properties and both belong to the matrix algebra generated by u. The decomposition $u = u_s + u_n$ is referred to as the Jordan decomposition of u. The matrices u_s and u_n are called the semisimple respectively the nilpotent part of u. In this section it will be more convenient to denote u_s by $J_s(u)$ and u_n by $J_n(u)$. In [3] a method based on the Newton–Hensel lifting procedure is presented which calculates a polynomial $s(x) \in K[x]$ of degree less than n from the square-free part of the minimal polynomial of u such that $s(u) = J_s(u)$. Combining this with Giesbrecht's Las Vegas methods [41] for calculating the minimal polynomial and for evaluating $s(u)$ we can compute $J_s(u)$ with $O(MM(n)\mathrm{polylog}\, n + SF_K(n))$ operations.

4.5.2 Finding a maximal torus

We show that the semisimple part of a random element generates a maximal torus with a good chance. The argument used here is a simplified (and improved) version of a proof given by Eberly and Giesbrecht [30] for a special case.

Lemma 4.14. *Let d stand for the dimension of a maximal torus in A'. There exists a polynomial function $f : A' \to K'$ of degree $d^2 - d$ such that for $u \in A'$ the subalgebra T' generated by the semisimple part $J_s(u)$ of u and the identity matrix is a maximal torus of A' if and only if $f(u) \neq 0$.*

Proof. By Wedderburn's theorem $A'/\mathrm{Rad}(A) \cong \bigoplus_{i=1}^{s} M_{n_i}(K')$. A maximal torus in $M_{n_i}(K')$ is conjugated to the set of diagonal matrices. It follows that $d = \sum_{i=1}^{s} n_i$. We assume that $\bigoplus_{i=1}^{s} M_{n_i}(K')$ is embedded into $M_d(K')$ in the natural way. Let $\phi : A' \to M_d(K')$ be the composition of the natural projection $A' \to A'/\mathrm{Rad}(A')$ with this embedding. Observe that ϕ commutes with taking the semisimple part: $\phi(J_s(u)) = J_s(\phi(u))$ for every $u \in A'$. We claim that the torus T generated by the identity of A and semisimple part $J_s(u)$ of u has dimension d if and only if $\phi(u)$ has d distinct eigenvalues. Indeed, since $\ker \phi = \mathrm{Rad}(A')$ and $T \cap \mathrm{Rad}(A') = (0)$, T and $\phi(T)$ are isomorphic. On the other hand, $\phi(T)$ is generated by $\phi(J_s(u)) = J_s(\phi(u))$ and the identity, hence the dimension of $\phi(T)$ is the degree of the minimal polynomial of $J_s(\phi(u))$ which equals the number of distinct eigenvalues of $\phi(u)$.

Let $\chi_u(x)$ denote the characteristic polynomial of the adjoint action $\mathrm{ad}\phi(u) : w \mapsto \phi(u)w - w\phi(u)$ of $\phi(u)$ on $M_d(K')$. We claim that the nullity of $\mathrm{ad}\phi(u)$ is at least d and equality holds if and only if $\phi(u)$ has d distinct eigenvalues. Indeed, we may assume that $\phi(u)$ is of Jordan normal form. One easily verifies that $\mathrm{ad}\phi(u)$ acts nilpotently on the block diagonal matrices whose blocks correspond to the Jordan blocks of $\phi(u)$. This implies the inequality and the ,,only if" part of the claim concerning the equality. The ,,if" part is even easier.

It follows that $\phi(u)$ has d eigenvalues if and only if the coefficient c_u of the term x^d in $\chi_u(x)$ is zero. Let $f(u)$ stand for this coefficient. It is known that the coefficient of x^l in the characteristic polynomial of a linear transformation on a vector space W is a homogeneous polynomial function on $End(W)$ of degree $\dim W - l$. In our case $\dim W = d^2$ and $l = d$. Our function f being the composition of a homogeneous polynomial function of degree $d^2 - d$ and the linear maps ad and ϕ is either zero or homogeneous of degree $d^2 - d$. An element $u \in A'$ such that $\phi(u)$ is a diagonal matrix with distinct eigenvalues witnesses that this polynomial is not identically zero. □

Thus a semisimple matrix $u \in A$ such that the torus T generated by u is probably maximal (with error probability δ) can be found with $O(MM(n)\mathrm{polylog}\, n + SF_K(n) + R(A))$ operations. The error probability can be pushed under a prescribed bound ϵ by repeating this procedure $O(\log \frac{1}{\epsilon})$ times independently, and taking the element which has minimal polynomial of maximal degree, see Lemma 4.2.

In the steps described in the rest of this section we assume that we are provided with an element u which generates a maximal torus T. We keep the notation introduced in Section 4.2 (C, S, H, N). We denote $\dim_K T$ by d.

4.5.3 Calculating C

We follow the method suggested by Theorem 4.12. First we calculate the subspace $L = [A, A] \cap T$. The next two lemmas provide us with a tool for generating random elements of L.

Lemma 4.15. *The map $a \mapsto J_s(\Phi_T a)$ is a linear map of A onto T and the map $a \mapsto J_n(\Phi_T a)$ is a linear map from A onto Rad(H). Furthermore, $J_n(\Phi_T a) = a$, for every $a \in$ Rad(H), and $J_s(\Phi_T a) = a$, for every $a \in T$.*

Proof. We know that Φ_T is a linear projection of A onto H. Also, $\Phi_T \mathrm{Rad}(A) = \mathrm{Rad}(A)$ and J_s is zero on $\mathrm{Rad}(H)$. By Wedderburn–Malcev, $H = T + N$, a direct sum of vector spaces. Let $\pi : H \to T$ and $\mu : H \to \mathrm{Rad}(H)$ stand for projections corresponding to this decomposition. It remains to show that J_s and J_n (restricted to H) coincide with π

and μ, respectively. For every $a \in H$, $\pi(a)$ is semisimple and $\mu(a)$ is nilpotent. Since H centralizes T, $\pi(a)$ commutes with a and the same holds for $\mu(a) = a - \pi(a)$. From the uniqueness of the Jordan decomposition we infer that $\pi(a) = J_s(a)$ and $\mu(a) = J_n(a)$. □

Lemma 4.16. *$J_s(\Phi_T[A, A]) = L = [A, A] \cap T$.*

Proof. Since $J_s(\Phi_T a) = a$ for every $a \in T$ it suffices to show that $J_s(\Phi_T a) \in [A, A]$ for every $a \in [A, A]$. By Corollary 2.1 $A = B + \mathrm{Rad}(A)$ (direct sum as vector spaces) for some semisimple subalgebra $B \leq A$ containing T. Since $J_s(\Phi_T a) = 0$ for every $a \in \mathrm{Rad}(A)$ it is sufficient to prove the assertion for the semisimple algebra B in place of A. By Lemma 4.7 we have $T = C_B(T)$ hence $J_s(\Phi_T a) = \Phi_T a$ for every $a \in B$. We know that $\Phi_T a - a \in [T, B] \subseteq [B, B]$. Hence $\Phi_T a \in [B, B]$ if and only if $a \in [B, B]$. □

We calculate a basis of L by generating sufficiently many random elements of the form $J_s(\Phi_T[a, b])$.

Lemma 4.17. *Let $k \leq \dim_K L$, $0 < \epsilon, \delta < 1$, and let $h \geq k\lceil(\log k + \log \frac{1}{\epsilon})/\log \frac{1}{\delta}\rceil$. Assume that the elements $a_{11}, b_{11}, \ldots, a_{1,h}, b_{1,h}, \ldots a_{d1}, b_{d1}, \ldots, a_{d,h}b_{d,h}$ are chosen independently from A according to a probability distribution which satisfies the condition $AlgRand(A, \dim_K L, \delta)$. Then with probability at least $1 - \epsilon$, the set $\{J_s(\Phi_T[a_{ij}, b_{ij'}])|i = 1, \ldots, k,\ j, j' = 1, \ldots, h\}$ contains at least k linearly independent elements of L.*

Proof. Let $l = \dim_K L$. By fixing a K-basis $b_1, \ldots, b_l$ of L we identify L with K^l. For a tuple $(y_1, z_1, \ldots, y_k, z_k) \in A^{2k}$ let Y stand for the $l \times k$ matrix whose columns are $J_s(\Phi_T[y_{i,z_i}])$ $(i = 1, \ldots, k)$. Let Γ be the family of all k-element subsets of $\{1, \ldots, l\}$. For each $\gamma \in \Gamma$ let $f_\gamma(y_1, z_1, \ldots, y_k, z_k)$ be the determinant of the $k \times k$ minor of Y which consists of the rows indexed by the elements of γ. Obviously f_γ is a multilinear function. We observe that all the functions f_γ $(\gamma \in \Gamma)$ vanish on a particular tuple $(y_1, z_1, \ldots, y_k, z_k) \in A^{2k}$ if and only if the elements $J_s(\Phi_T[y_i, z_i])$ $(i = 1, \ldots, k)$ are linearly dependent over K. By Lemma 4.16 this cannot be the case for every $(y_1, z_1, \ldots, y_k, z_k) \in A^{2k}$ and hence there exists at least one $\gamma \in \Gamma$ such that f_γ is not identically zero. By Lemma 4.2 with probability at least $1-\epsilon$ there exist indices $j_1, \ldots, j_k, j_1', \ldots, j_k'$ such that $f_\gamma(a_1 j_1, b_{1j_1'}, \ldots, a_{kj_k}, b_{kj_k'}) \neq 0$. Then the elements $J_s(\Phi_T[a_{1j_1}, b_{1j_1'}]), \ldots, J_S(\Phi_T[a_{kj_k}, b_{kj_k'}])$ are linearly independent. □

Like in Section 4.4, it will be convenient to perform calculations in T in terms of the basis $1, u, \ldots, u^{d-1}$. If it has not been done before we calculate the Frobenius normal form $Frob(u)$ of u together with a transition matrix b such that $b^{-1}ub = Frob(u)$ using Giesbrecht's method with $O(MM(n)\mathrm{polylog}\, n)$ field operations. Then we can read the coordinates of an element $z \in T$ in terms of the basis $1, u, \ldots, u^{d-1}$ from the first column of the first block of $b^{-1}ub$.

We find a basis of L with $O(\log \frac{1}{\epsilon} \dim_K L(MM(n)+R(A)+SF_K(n))\mathrm{polylog}\, n)$ operations (even if $\dim_K L$ is not known a priori) as follows. Set $h = \lceil(\log d + \log \frac{d}{\epsilon})/\log \frac{1}{\delta}\rceil$. For $k = 1, 2, 4, \ldots, 2^{\lceil \log_2 d\rceil}$ select a maximal linearly independent system from $\{J_s(\Phi_T[a_{ij}, b_{ij'}])|i = 1, \ldots, k,\ \ j, j' = 1, \ldots, h\}$ where a_i, b_i are random elements of A chosen independently according to a distribution which satisfies $AlgRand(A, d, \delta)$. We stop if we obtained less than k elements, otherwise we proceed with $2k$ in place of k. By the lemma, the probability that we stop with a system which does not generate L is at most ϵ.

Note that $A/\mathrm{Rad}(A)$ is commutative iff $L = (0)$. Then $C = T$. Otherwise assume that we have a basis $b_1, \ldots, b_l$ of L. We choose linear function $f_1, \ldots, f_{d-l} : T \to K$ such that $L \bigcap_{i=1}^{d-l} \ker f_i$. Then $C = \{z \in T|zL \subseteq L\} = \{z \in T|f_i(zb_j) = 0\ (i = 1, \ldots, l,\ j =$

$1, \dots, d-l)\}$ whence we obtain a basis of C by solving a system of $l(d-l)$ linear equations in d variables. This costs $O(MM(d)l(d-l)/d) = O(dMM(d))$ operations. Finally we find an element $u' \in C$ which generates C as an algebra with identity by taking a random linear combination of these basis elements. (By Lemma 4.14, a random element of C will generate C. Note that we can verify whether u' generates C with $O(MM(d)\text{polylog}\, d)$ operations by testing linear independence of $1, u, \dots, u^{\dim C - 1}$.)

The total cost of the algorithm described in this subsection amounts to $O(\log \frac{1}{\epsilon} d(MM(n) + R(a) + SF_K(n))\text{polylog}\, n)$ operations in K. If $A/\text{Rad}(A)$ happens to be commutative then $O(\log \frac{1}{\epsilon}(MM(n) + R(a) + SF_K(n))\text{polylog}\, n)$ operations are sufficient.

4.5.4 Generating elements of N

Throughout this subsection we assume that we are provided with an element u' which generates C as an algebra with identity.

Lemma 4.18. *Assume that $a_1, \dots, a_m$ generate A as an algebra with identity. Then the elements $\{[u', a_1], \dots, [u', a_m]\}$ generate ANA as an ideal of A.*

Proof. Let J be the ideal generated by $[u', a_1], \dots, [u', a_m]$. Obviously $J \subseteq A[u', A]A \subseteq A[C, A]A = ANA$. Observe that $u' + J$ centralizes the generators $a_i + J$ of the factor algebra A/J. Hence $[u', A] \subseteq J$ and since C is generated by u' we have $[C, A] \subseteq J$. By definition $N = [C, A]$. □

Hence generators of ANA can be calculated with $O(mMM(n))$ operations by taking $[u', g_1], \dots, [u', g_m]$.

4.5.5 Generating elements of Rad(H)

We generate elements of $\text{Rad}(H)$ as follows. From a random element $a \in A$ we first calculate $\Phi_T a$ using the method described in Section 4.4. Then we compute the nilpotent part $J_n(\Phi_T a)$ of $\Phi_T a$. The cost is $O(MM(n)\text{polylog}\, n + SF_K(n))$ operations. Note that because of the linearity of the map $a \mapsto J_n(\Phi_T a)$ (cf. Lemma 4.15) the method can be considered as a way to generate ,,random" elements of $\text{Rad}(H)$. To be more specific, if we choose the element a according to a distribution satisfying $AlgRand(A, D, \delta)$ then the distribution of $J_n(\Phi_T a)$ satisfies condition $AlgRand(\text{Rad}(H), D, \delta)$.

We are going to give an upper bound for the number of elements from $\text{Rad}(H)$ which — in addition to the generators of ANA — are sufficient to generate $\text{Rad}(A)$ as an ideal. The following elementary lemma is well known. A proof can be obtained by combining Corollary 4.1b of [80] and Lemma 4.2 of loc. cit..

Lemma 4.19. *Let B be finite dimensional K-algebra and $M \subseteq Rad(B)$. Then $Rad(B) = BMB$ if and only if $Rad(B) = BMB + Rad(B)^2$. In other words, the ideal generated by M is $Rad(B)$ if and only if the same holds modulo $Rad(B)^2$.*

Lemma 4.20. *Assume that $A/Rad(A)$ is a central simple K-algebra of dimension d^2. Then $Rad(A)$ as an ideal of A can be generated by $\lceil \dim_K Rad(A)/d^3 \rceil$ elements from $Rad(H)$. Furthermore, A as an algebra with identity cannot be generated by less than $\lceil \dim_K Rad(A)/d^4 \rceil$ elements.*

Proof. Let ψ stand for the natural projection $A \to A/\mathrm{Rad}(A)^2$. Then $\psi(T)$ is a maximal torus in $A \to A/\mathrm{Rad}(A)^2$. We have $C_{\psi(A)}\psi(T) = \Phi_{\psi(T)}(\psi(A)) = \psi(\Phi_T A) = \psi(H)$. In view of this together with Lemma 4.19 it is sufficient to prove the assertion for $A/(\mathrm{Rad}(A))^2$ in place of A. In other words, we may assume that $\mathrm{Rad}(A)^2 = (0)$. By Wedderburn–Malcev, there exists a subalgebra $D \leq A$ such that $A = D + \mathrm{Rad}(A)$ (direct sum as vector spaces). Assume that A is generated by $a_1, \ldots, a_m$. Let $a_i = b_i + c_i$ where $b_i \in D$ and $c_i \in \mathrm{Rad}(A)$. One easily verifies that $c_1, \ldots, c_m$ generate $\mathrm{Rad}(A)$ as an ideal. On the other hand, since $\mathrm{Rad}(A)^2 = 0$ we have $Ac_iA = (D + \mathrm{Rad}(A))c_i(D + \mathrm{Rad}(A)) = Dc_iD$, whence $\dim_K Ac_iA \leq (\dim_K D)^2 = d^4$. This implies the inequality $m \geq \lceil \dim_K \mathrm{Rad}(A)/d^4 \rceil$.

To prove the first assertion we use a refinement of the argument of the proof of Theorem 4.11. We consider $\mathrm{Rad}(A)$ as a $D \otimes_K D$-module in the natural way. Then ideals of A contained in $\mathrm{Rad}(A)$ are exactly the $D \otimes_K D$-submodules and elements b of $\mathrm{Rad}(H) = \mathrm{Rad}(A) \cap C_A(T)$ are characterized as $(1 \otimes a)b = (a \otimes 1)b$ for every $a \in T$. We know that $D \otimes_K D \cong M_{d^2}(K)$ and $\mathrm{Rad}(A)$ as a $D \otimes_K D$-module is isomorphic to D^h, the direct sum of h copies of the simple $D \otimes_K D$-module D (with the natural module structure). Here $h = \dim_K \mathrm{Rad}(A)/d^2$. We claim that if $a_1, \ldots, a_r$ are linearly independent elements of D then $(a_1, \ldots, a_r)$ generates the $D \otimes_K D$-module D^r. This can be verified at once if we identify $D \otimes_K D$ with $M_{d^2}(K)$ and D with the standard $M_{d^2}(K)$-module K^{n^2}. Let $r \leq d$ and choose r linearly independent elements $a_1, \ldots, a_r$ from T. Then by the claim, $b = (a_1, \ldots, a_r)$ generates D^r as a $D \otimes_K D$-module and $(1 \otimes a)b = (a_1a, \ldots, a_ra) = (aa_1, \ldots, aa_r) = (a \otimes 1)b$. Hence $\lceil h/d \rceil$ generators of $\mathrm{Rad}(A)$ with the required property can be constructed by distributing the irreducible summands of $\mathrm{Rad}(A)$ into appropriate blocks and taking a single generator in each block. □

Corollary 4.21. *Assume that A as an algebra with identity is generated by m elements. Suppose that the simple components of $A/Rad(A)$ are $\widetilde{A}_1, \ldots, \widetilde{A}_r$ with dimensions $\dim_K \widetilde{A}_i / \dim_K Z(\widetilde{A}_i) = d_i^2$. Then there exists a subset $M \subseteq Rad(H)$ of size at most*

$$\mathrm{Max}\{\mathrm{Min}(md_i, \lceil \dim_K A/d_i^3 \rceil) | i = 1, \ldots, r\} \leq \lceil (\dim_K A)^{\frac{1}{4}} m^{\frac{3}{4}} \rceil \leq \lceil n^{\frac{1}{2}} m^{\frac{3}{4}} \rceil$$

such that $A(M + N)A = Rad(A)$.

Proof. As in the proof of Lemma 4.20 we can assume that $\mathrm{Rad}(A)^2 = (0)$. Then $N^2 = (0)$ as well and by Proposition 4.8, N is an ideal of A. Hence $S \cong A/N$ and S is also generated by m elements. This means that for the rest of the proof we may further assume that $N = (0)$, or, equivalently, $A = S$. By Proposition 4.10, A is a direct sum of subalgebras $A_1, \ldots, A_r$, where $A_i/\mathrm{Rad}(A_i) \cong \widetilde{A}_i$. Assume that $M_i \subseteq \mathrm{Rad}(H_i) = \mathrm{Rad}(H) \cap H_i$ such that $A_iM_iA_i = \mathrm{Rad}(A_i)$ $(i = 1, \ldots, r)$. It is easy to construct a set $M \subseteq \mathrm{Rad}(H)$ of cardinality $\mathrm{Max}\,|M_i|$ such that for every $i \in \{1, \ldots, r\}$ $\pi_i(M) = M_i$ where $\pi_1, \ldots, \pi_r$ are the projections corresponding to the direct decomposition of A. It is immediate that such an M generates $\mathrm{Rad}(A)$ as an ideal. Hence it is sufficient to prove the assertion in the special case where $\widetilde{A} = A/\mathrm{Rad}(A)$ is a simple K-algebra. Then C is a field in $Z(A)$ and we can consider A as a C-algebra. The statement now follows from Lemma 4.20, applied to A as a C-algebra. (The bound independent of the d_is is obtained by taking an appropriate weighted geometric mean of md_i and $\dim_K A/d_i^3$.) □

The next lemma gives a bound on the random elements of $\mathrm{Rad}(H)$ which probably generate $\mathrm{Rad}(A)$ modulo the ideal AMA. We omit the proof which is rather technical and can be carried out in a fashion similar to the proof of Lemma 4.17.

Lemma 4.22. *Assume that there exists a subset $M \subseteq Rad(H)$ of size k such that $A(M+N)A = Rad(A)$. Let $0 < \epsilon, \delta < 1$ and $h \geq k\lceil(\log k + \log\frac{1}{\epsilon})/\log\frac{1}{\delta}\rceil$. Assume that the elements $a_1, \ldots, a_h, \in A$ are chosen independently according to a probability distribution satisfying $AlgRand(A, \dim_K Rad(A), \delta)$. Then with probability at least $1-\epsilon$ the subspace $N \cup \{J_n(\Phi_T a_i) | i = 1, \ldots, h\}$ generate $Rad(A)$ as an ideal of A.*

4.5.6 Computing Rad(A)

Here we summarize the algorithm for computing Rad(A) and conclude the proof of Theorem 4.3. The input consists of matrices $g_1, \ldots, g_m$ such that A is the matrix algebra generated by the identity matrix and $g_1, \ldots, g_m$. We assume that random elements of a are generated independently according to a probability distribution satisfying condition $AlgRand(A, n^2, \delta)$ for a constant $0 < \delta < 1$, say 1/2. An error probability bound $0 < \epsilon < 1$ is also given as a part of the input. We require that each of the three big steps which make use of randomization of the algorithm works correctly with probability at least $1 - \frac{\epsilon}{3}$.

First we find a semisimple matrix u which generates a maximal torus T by the method of Subsection 4.5.2. Then we calculate the subalgebra $C \leq T$ (and a generator u' of C) using the method described in Subsection 4.5.3. If $C = T$ then we set $k = m$ otherwise $k = \lceil n^{\frac{1}{2}} m^{\frac{3}{4}} \rceil$. Then we calculate the commutators $[u', g_i]$ ($i = 1, \ldots, m$) as well as $J_n(\Phi_T a)$ for $O(\log \frac{1}{3\epsilon} k \log k)$ random elements $a \in A$. (The exact constant is given in Lemma 4.22.) These elements generate Rad(A) with probability at least $1 - \epsilon$. This finishes the proof of Theorem 4.3.

4.6 Remarks

Assume that $\operatorname{char} K = 0$, ϵ is a constant and m is small, say $m = O(\log n)$. Then the cost of the algorithm is $O(nMM(n)\text{polylog}\, n) = O(n^4)$ operations provided that we can draw random elements very efficiently (e.g., $R(A) = O(MM(n)\text{polylog}\, n)$. If a basis for A is also available then we can produce random elements of A using $O(n^4)$ field operations. Then the cost of the algorithm is roughly $O(n^5)$. Even this is definitely better than the cost of the previously known methods based on Dickson's characterization (see [38]) which appears to be around $O(n^6)$.

Note that for applications it seems to be important to exhibit a single nonzero element of Rad(A) (provided that Rad(A) $\neq$ (0)). For this task an algorithm of complexity around $O(MM(n))$ could be considered optimal. By a version of the algorithm presented here we can almost achieve this bound in the special cases where Rad(H) $\neq$ (0) or A/Rad(A) is (nearly) commutative. In the general case computation of the subalgebra C with its complexity roughly $O(nMM(n))$ appears to be the weakest point of the present algorithm. The results of the next two chapters rely on alternative methods for computing subalgebras analogous to C or at least certain replacements in the finite ground field case.

Chapter 5

Treating the exceptional cases of the MeatAxe

In this chapter, based on the paper [58] (joint work with Klaus Lux), we show that the Holt-Rees extension of the standard MeatAxe procedure finds submodules of modules over finite algebras with positive probability in slightly greater generality than originally claimed. For the case when the Holt–Rees method fails we propose a further, but still simple and efficient extension.

The problem of finding the irreducible composition factors of a finite module M for a finite dimensional associative algebra A over a finite field F is one of the fundamental tasks in computational modular representation theory. The most commonly used practical approach to this problem is the MeatAxe algorithm described in R. Parker's paper [79]. It solves the problem of proving constructively that M is irreducible but originally, the method did not perform satisfactorily when the ground field F is large. D. F. Holt and S. Rees in [52] propose an extension to Parker's method which is based on factoring the characteristic polynomial of random elements from A. They provide an accurate analysis and show that their approach proves efficiently that a given module is irreducible independently of the size of the ground field. Furthermore, in most cases of reducible modules, they also have a good chance of finding a nontrivial submodule. In this chapter we prove that the extension works for a wider class of inputs than claimed in [52]. However, there is still one type of modules where the algorithm definitely fails. We propose a method for this case. We remark that the implementation of M. Ringe as part of the C-MeatAxe shows that our algorithm is also practically feasible. See the original paper [58] for a report on experimental results with this implementation.

In this section by F we denote the finite field $GF(q)$ consisting of q elements. We assume we have an associative algebra A (with identity) over F and we work with the module is $M = F^d$, the space of column vectors of length d where the action of A on M is unital, faithful and is given in terms of matrices for generators of A. (Note that, in contrast to the GAP implementation of the MeatAxe, which is based on row vectors and right action, the discussion of this chapter is presented in terms of column vectors and left action.) Furthermore, we identify A with its image in $\mathcal{M}_d(F)$. Finally, we assume that we are provided with an auxiliary procedure which generates random elements of A (independently and uniformly). This differs from the "algebraic random" model used in Chapter 4. Most notably, this assumption does not imply that that the ground field is large. The procedure either concludes that M is an irreducible A-module or returns a nontrivial submodule of M.

The rest of this chapter is structured as follows. In Section 5.1 we briefly comment on the original algorithm proposed in [52] in order to extend the probability analysis to slightly more cases and to describe a class of algebras A which contains all the situations where the method fails. In terms of the structure theory of Chapter 4, it will turn out that these algebras are very special cases of those where the radical has only "commutator part". Therefore what we need is finding efficiently the subalgebra C of a maximal torus – or at least a good replacement for C. This replacement will be $C \cdot \iota$, where ι is a primitive idempotent. The algorithm for this class of algebras is outlined in Section 5.2. The probability of success will be estimated in Section 5.3.

It will be convenient to introduce some additional notation. By the Wedderburn–Malcev principal theorem (see the Section 2.2.4), A can be written as

$$A = S + \mathrm{Rad}(A), \text{ where } S \cong A/\mathrm{Rad}(A).$$

Since the complementary subalgebra S is unique up to a conjugation by an inner automorphism of A, we can speak about the structural properties of A in terms of S even if S is not specified explicitly.

5.1 The exceptional algebras

In [52], the extension of MeatAxe is proved to succeed in constructing a nontrivial submodule with probability at least 0.144 in many cases. In particular, it recognizes irreducible modules, finds a nontrivial submodule if $M/\mathrm{Rad}(M)$ is decomposable or M contains non-isomorphic composition factors. The submodule is generated from the kernel of $p(x)$, where x is a random element and $p(t)$ is an appropriate irreducible factor of the characteristic polynomial of x on M (see Lemma 5.1 below). The probability analysis of success is based on the following observation, which will be useful in the analysis of the present chapter as well.

Lemma 5.1. *Let W be an irreducible A-module and $E = End_A(W)$, the algebra consisting of the A-endomorphisms of the module W. Then for at least 21.4% of the elements $x \in A$ the characteristic polynomial over F of x on the module W has an unrepeated irreducible factor of degree $\dim_F E$.*

Proof. By Schur's lemma and Wedderburn's theorem on finite division algebras, E is a finite extension field of F. Note also that if W as an E-module is isomorphic to E^n and $I = \{x \in A | xW = (0)\}$ is the annihilator ideal of W, then $A/I \cong \mathcal{M}_n(E)$. Since uniform selection of elements in A corresponds to uniform selection in the factor A/I, we may assume throughout the proof that $I = (0)$ and identify A with $\mathcal{M}_n(E)$. The statement for the case $E = F$ is proved in [52] (with a somewhat bigger constant), therefore we may restrict ourselves to the case $e = \dim_F E > 1$. The argument given in [52] for this case appears to contain a minor mistake, therefore we give a corrected proof below.

The condition is equivalent to that x, considered as a matrix over E, has an unrepeated eigenvalue λ such that λ is not contained in any proper subfield E' with $F \leq E' < E$ and for every automorphism $\sigma \in Gal(E|F)$ such that $\lambda^\sigma \neq \lambda$, λ^σ is not an eigenvalue of x. (This follows from the fact that the characteristic polynomial of x over F is $\prod_{\sigma \in Gal(E|F)} c(t)^\sigma$, where $c(t) \in E[t]$ is the characteristic polynomial of x, regarded as a matrix over E. See for example Theorem 9.10 and Exercise 9.4 in [82].)

Note that at most half of the elements in E can be contained in a proper subfield of E. This establishes the case $n = 1$. For the rest of the proof we assume $n > 1$.

Let $F = GF(q)$ and $E = GF(q^e)$. Following the arguments given in [52], let H denote the number of matrices $x \in \mathcal{M}_n(E)$ such that a specific $\lambda \in E$ is an unrepeated eigenvalue of x. Also, let H' stand for the number of matrices with two distinct specific unrepeated eigenvalues $\lambda, \mu \in E$. In [52] it is shown that H and H' are independent of the particular choice of λ and μ, and

$$H = \frac{1}{q^e - 1} \prod_{i=0}^{n-1} (q^{en} - q^{ei}) \quad \text{and} \quad H' \leq \frac{H}{q^e - 1}.$$

Let R denote the set of elements $\lambda \in E$ such that λ has exactly e conjugates over $Gal(E|F)$ and let $r = |R|$. By inclusion-exclusion, at least $rH - \binom{r}{2}H' \geq (r - \frac{r(r-1)}{2(q^e-1)})H$ matrices have some unrepeated eigenvalue from R. For the number of matrices having at least two eigenvalues from some orbit of $Gal(E|F)$ on R we have the crude upper bound $\frac{r}{e}\binom{e}{2}H' \leq \frac{r}{e}\binom{e}{2}\frac{H}{q^e-1}$. Hence the number of matrices with the required property is at least

$$\begin{aligned}
\left(r - \frac{r(r-1) + r(e-1)}{2(q^e-1)}\right) H &= \left(1 - \frac{r+e-2}{2(q^e-1)}\right) rH \\
&\geq \left(1 - \frac{q^e + e - 4}{2(q^e-1)}\right) \frac{q^e}{2} H \\
&= \left(\frac{1}{2} - \frac{e-3}{2(q^e-1)}\right) \frac{q^e}{2} H \geq \frac{7}{30} q^e H.
\end{aligned}$$

The first inequality follows from $q^e/2 \leq r \leq q^e - q \leq q^e - 2$, while the second one from that the maximal value of $(e-3)/(2q^e - 2)$ for the integers $q, e \geq 2$ is $\frac{1}{30}$ (taken at $q = 2, e = 4$). Hence the proportion of such matrices is at least

$$\frac{7}{30} q^e H / q^{en^2} = \frac{7}{30} \prod_{i=2}^{n} (1 - q^{-ei}) \geq \frac{7}{30} \prod_{i=2}^{\infty} (1 - 4^{-i}) \geq 0.214.$$

□

Remark. The mere assumption that x, regarded as a matrix over E contains an unrepeated eigenvalue λ which is not contained in any proper subfield (cf. [52]) appears to be insufficient even for the purposes of the MeatAxe. Indeed, if an algebraic conjugate λ' of λ, different from λ, is also an eigenvalue of x, then the characteristic polynomial of x over F contains the minimal polynomial $p(t)$ of λ at least twice and therefore the dimension of the kernel of $p(x)$ over E is at least 2.

The only possible situations when the Holt-Rees extension of MeatAxe may fail are modules M such that $\text{Rad}(M) \neq (0)$, $M/\text{Rad}(M)$ is irreducible and all the composition factors are isomorphic to $M/\text{Rad}(M)$. Since M is faithful, this implies that every irreducible A-module is isomorphic to $M/\text{Rad}(M)$. Let $E = \text{End}_A(M/\text{Rad}(M))$, as in Lemma 5.1. Then E is a finite extension field of F and $M/\text{Rad}(M)$ is isomorphic to E^n as an S-module for some integer n, where S is a subalgebra of A isomorphic to $A/\text{Rad}(A)$. Note that the multiplicity of E^n in M is d/en, where $e = \dim_F E$ and $S \cong \mathcal{M}_n(E)$. The center of S is therefore isomorphic to E. We may and shall identify E with $\mathrm{Z}(S)$. In summary, we have

$$\mathrm{Rad}(A) \neq (0), \quad S \cong \mathcal{M}_n(E), \quad E = \mathrm{Z}(S) \text{ is an extension field of } F. \tag{5.1}$$

The Holt-Rees extension of the MeatAxe is shown to succeed even in this case provided that $E = F$. We extend the proof given in [52] to the more general case where $E \leq \mathrm{Z}(A)$.

Proposition 5.2. *Assume that* (5.1) *holds, $M/Rad(M)$ is irreducible and $E \leq \mathrm{Z}(A)$. Then, for at least* 14.4% *of the elements x in A, there exists a factor $p(t) \in F[t]$ of the characteristic polynomial of x on M such that the kernel of $p(x)$ is a nonzero subspace of $Rad(M)$.*

Proof. The statement is proved for $E = F$ in [52]. Assume that $E > F$. Note that every element $x \in A$ can be uniquely written in the form $x = x_0 + x_1$ where $x_0 \in S$ and $x_1 \in \mathrm{Rad}(A)$. Assume that the characteristic polynomial (over F) of $x_0 \in S$ on the irreducible S-module $M/\mathrm{Rad}(M) \cong E^n$ has an unrepeated irreducible factor $p(t) \in F[t]$ of degree $e = \dim_F E$. By Lemma 5.1, this is the case for at least 21.4% of the possible choices for x_0. Let $\lambda_1, \ldots, \lambda_e$ be the roots of $p(t)$ in E. Then there exists an element $\lambda \in \{\lambda_1, \ldots, \lambda_e\}$, say $\lambda = \lambda_1$ such that the kernel of $x_0 - \lambda$ is an E-submodule of $M/\mathrm{Rad}(M)$ of rank 1 (i.e., a one dimensional E-linear subspace). Furthermore, $x_0 - \lambda_i$ is a unit in S for $i = 2, \ldots, e$.

Obviously, for every $x_1 \in \mathrm{Rad}(A)$, the kernel of $x_0 + x_1 - \lambda$ in M is nonzero, since the quotient map on $M/\mathrm{Rad}(M)$ is $x_0 - \lambda$. Let L stand for the set consisting of $x_1 \in \mathrm{Rad}(A)$ for which this kernel is not contained in $\mathrm{Rad}(M)$. We claim that L is contained in a proper E-submodule of $\mathrm{Rad}(A)$. To this end consider M as an S-module. Since S is a simple algebra there exists an S-submodule M_0 complementary to $\mathrm{Rad}(M)$. Then M_0, as an S-module, is isomorphic to $M/\mathrm{Rad}(M)$. In particular, there exists a nonzero element $v \in M_0$ such that $(x_0 - \lambda)v = 0$. Then for every element $x_1 \in \mathrm{Rad}(A)$, the kernel of $x_0 + x_1 - \lambda$ is contained in the E-submodule $Ev + \mathrm{Rad}(M)$. Assume now that $x_1 \in L$, i.e., this kernel contains an element $u \in M \setminus \mathrm{Rad}(M)$. Then $u = \beta v + w$ for some unit $\beta \in E$ and some element $w \in \mathrm{Rad}(M)$. Multiplying by β^{-1}, we may assume that $u = v + w$ with $w \in \mathrm{Rad}(M)$. Now

$$0 = (x_0 + x_1 - \lambda)(v + w) = x_1 v + (x_0 - \lambda)w + x_1 w,$$

and hence

$$x_1 v = -(x_0 - \lambda)w - x_1 w$$

is in

$$(x_0 - \lambda)\mathrm{Rad}(M) + \mathrm{Rad}(A)\mathrm{Rad}(M) = (x_0 - \lambda)\mathrm{Rad}(M) + \mathrm{Rad}^2(M).$$

Thus

$$L \subseteq L' = \{x_1 \in \mathrm{Rad}(A) | x_1 v \in (x_0 - \lambda)\mathrm{Rad}(M) + \mathrm{Rad}^2(M)\}.$$

Obviously L' is an E-submodule of $\mathrm{Rad}(A)$. Assume that $L' = \mathrm{Rad}(A)$. Then

$$\mathrm{Rad}(M) = \mathrm{Rad}(Av) = \mathrm{Rad}(A)v = L'v \subseteq (x_0 - \lambda)\mathrm{Rad}(M) + \mathrm{Rad}^2(M).$$

(Here, the first equality holds because of $M = Av + \mathrm{Rad}(M)$ and Nakayama's lemma.) From this we infer that $x-\lambda$ acts surjectively on the factor module $\mathrm{Rad}(M)/\mathrm{Rad}^2(M)$, and hence on its composition factors as well. Since all these composition factors are isomorphic to $M/\mathrm{Rad}(M)$, this is a contradiction to the fact that $x-\lambda$ is singular on $M/\mathrm{Rad}(M)$. Thus L is included in the proper E-submodule L' of $\mathrm{Rad}(A)$, as claimed.

By the claim, for at least $1-\frac{1}{|E|}$ of the possible choices for x_1, the kernel of $x-\lambda = x_0 + x_1 - \lambda$ is a subspace of $\mathrm{Rad}(M)$. Let $\rho = \prod_{i=2}^{e}(x-\lambda_i)$. Then ρ is a unit modulo $\mathrm{Rad}(A)$ and hence ρ itself is a unit in A. Therefore the kernel of $(x-\lambda)\rho = p(x)$ is equal to the kernel of $x-\lambda$. Thus, the kernel of $p(x)$ is a nonzero subspace of $\mathrm{Rad}(M)$ provided that the kernel of $x-\lambda$ is. As the components x_0 and x_1 of x are chosen independently, this gives $0.214(1-\frac{1}{|E|}) \geq 0.214 \cdot 3/4 > 0.16$, so that at least $16\% > 14.4\%$ of the elements $x \in A$ satisfy the desired property. □

This means that the Holt-Rees extension of MeatAxe succeeds with probability at least 0.144 in this case. Hence we can restrict our attention to the case where E is not central, i.e., algebras A satisfying (5.1) and the additional hypothesis

$$[A, E] > (0) \tag{5.2}$$

5.2 The algorithm

We propose the method described below for treating algebras with properties (5.1) and (5.2).

In the following, we assume that a random element $x \in A$ is selected and the irreducible factors of the characteristic polynomial $c(t)$ of x over F are computed. Note that these computations are carried out as a part of the original algorithm described in [52]. We select a factor $p(t)$ of minimum degree among the factors of $c(t)$ of minimum multiplicity and do the following.

(i) Determine the polynomial $i(t)$, a representative of the primitive idempotent of the algebra $F[t]/(c(t))$ corresponding to the factor $p(t)$. More precisely, by the Chinese Remainder Theorem,

$$F[t]/(c(t)) \cong F[t]/(p^l(t)) \oplus F[t]/(q(t)),$$

where l is the multiplicity of $p(t)$ in $c(t)$ and $q(t) = c(t)/p^l(t)$ and we want the identity element of the component isomorphic to $F[t]/(p^l(t))$. To be explicit, 1 can be expressed using the extended Euclidean algorithm in the form

$$1 = a(t)p^l(t) + b(t)q(t)$$

with polynomials $a(t)$ and $b(t)$. Then $i(t) \equiv b(t)q(t) \pmod{c(t)}$.

(ii) Choose another random element $y \in A$ as well as a random vector $v \in M$ and calculate the submodule N generated by $[x, i(x)yi(x)]v$. If this is a proper nonzero submodule then return N, otherwise report failure.

We make comments only on the costs of steps which are additional to the Holt-Rees extension of the MeatAxe procedure. The polynomial $i(t)$ can be determined with $O(d^2)$ operations in F (Note that l is less than d). The cost of computing the vector $[x, i(x)yi(x)]v$

is $O(d^3)$ arithmetical operations assuming that we use a method based on performing $O(d)$ matrix-by-vector multiplications. We remark that using a method based on fast calculation of Krylov sequences (see [14]) the cost can be reduced to $O(\mathrm{MM}(d)\log d)$ operations, where $\mathrm{MM}(d)$ stands for the number of arithmetic steps required to multiply two d by d matrices. We remark that in [30], Lemma 3.1, an efficient algorithm is described which computes all the primitive idempotents of the subalgebra generated by x simultaneously in explicit matrix form. The method is based on computing the rational canonical form of x (cf. [41]), and the running time is essentially $O(\mathrm{MM}(d)\log d)$.

Thus the total number of arithmetical steps required by the algorithm is dominated by the cost of computing the submodule N in step (ii), which is $O(d^3)$, provided that the number of generators of A is fixed.

5.3 Probability of success

Below we give an estimate for the probability of finding a proper submodule in the situation where the algebra A satisfies conditions (5.1) and (5.2). Actually we show that the commutator $[x, i(x)yi(x)]$ has a positive chance for being a nonzero element of $\mathrm{Rad}(A)$.

Lemma 5.3. *Assume that the finite dimensional F-algebra A with identity satisfies conditions* (5.1) *and* (5.2)*. Let ι be an idempotent of S. Then*

(a) $[\iota E\iota, \iota A\iota] = \iota[E, A]\iota$,

(b) $S(\iota[E, A]\iota)S = [E, A]$, *and*

(c) $(0) \subset [\iota E\iota, \iota A\iota] \subseteq \mathit{Rad}(A)$.

Proof. First we note that since ι commutes with E, $\iota b\iota = \iota b = b\iota$ for every $b \in E$ and hence $\iota E\iota = \iota E = E\iota$. Part (a) is immediate from the following equalities which hold for every $b \in E$ and $a \in A$.

$$\iota b \cdot \iota a\iota - \iota a\iota \cdot \iota b = \iota b\iota \cdot a\iota - \iota a \cdot \iota\iota b = \iota b \cdot a\iota - \iota a \cdot b\iota = \iota(ba - ab)\iota.$$

To prove part (b), let $s, s' \in S$, $b \in E$, $a \in A$. Then

$$s\iota[b, a]\iota s' = s\iota ba\iota s' - s\iota ab\iota s' = bs\iota a\iota s' - s\iota a\iota s' b = [b, s\iota a\iota s'],$$

where the second equality holds because b commutes with the elements $\iota, s, s' \in S$. From this we infer that $S\iota[E, A]\iota S = [E, S\iota A\iota S]$. It remains to establish the equality $S\iota A\iota S = A$. To this end observe that $S\iota S$ is a nonzero ideal in the simple algebra S, therefore $S\iota S = S$. Hence $S\iota A\iota S = S\iota SAS\iota S = SAS = A$. (The first and the last equalities are obvious because S contains 1_A.)

Part (c) follows from (a) and (b) and the fact that E is central modulo $\mathrm{Rad}(A)$. □

After these preparations we are ready to give a lower bound on the probability of success of the algorithm.

Proposition 5.4. *Assume that the matrix algebra $A \leq \mathcal{M}_d(F)$ satisfies conditions* (5.1) *and* (5.2)*. Then the proportion of the triples $(x, y, v) \in A \times A \times F^d$ for which the algorithm described in the preceding section finds a proper submodule is at least* 0.08.

Proof. Assume that $p(t)$ is an unrepeated irreducible factor of the characteristic polynomial of $x + \mathrm{Rad}(A)$ on E^n. Then the degree of $p(t)$ is the dimension (over F) of the kernel of $p(x + \mathrm{Rad}(A))$. This subspace is obviously a $\mathrm{Z}(A/\mathrm{Rad}(A))$-submodule of E^n, and hence the degree of $p(t)$ is at least $e = \dim_F \mathrm{Z}(A/\mathrm{Rad}(A)) = \dim_F E$. Assume that the degree of $p(t)$ is exactly e. By Lemma 5.1, such a factor does exist for at least 21.4% of the elements $x \in A$. Furthermore, all the factors of this kind are characterized as the minimum degree factors amongst the factors of minimal multiplicity of the characteristic polynomial of x on the whole module M.

Referring to the homomorphism $F[t]/(c(t)) \to A$ induced by $x \mapsto x$, it is immediate that $\iota = i(x)$ is an idempotent. Let $\overline{x} = x + \mathrm{Rad}(A)$ and $\overline{\iota} = \iota + \mathrm{Rad}(A)$. Furthermore, the characteristic polynomial of $\overline{\iota x}$ on E^n is $p(t)x^{(n-1)e}$. It follows that $\overline{\iota x}$ and $\overline{\iota}$ have rank e, therefore $\overline{\iota}$ is a primitive idempotent of $A/\mathrm{Rad}(A)$. Hence $\overline{\iota}(A/\mathrm{Rad}(A))\overline{\iota} = \overline{\iota}\mathrm{Z}(A/\mathrm{Rad}(A))\overline{\iota}$. In particular, $\overline{\iota x} \in \overline{\iota}\mathrm{Z}(A/\mathrm{Rad}(A))$. On the other hand, the minimum polynomial of $\overline{\iota x}$ on $\overline{\iota}Z(A/\mathrm{Rad}(A))$ is of degree e, therefore $\overline{\iota x}$ generates the whole $\overline{\iota}\mathrm{Z}(A/\mathrm{Rad}(A))$.

Now $[x, \iota y \iota]$ is a nonzero element of $\mathrm{Rad}(A)$ for at least $1 - \frac{1}{|E|} \geq \frac{3}{4}$ of the elements $y \in A$, see Lemma 5.5 below, and let us assume in the following that this is the case. Then $[x, \iota y \iota]$ is a nontrivial F-linear transformation and hence the kernel has codimension at least 1. Therefore for at least $1 - \frac{1}{|F|} \geq \frac{1}{2}$ of the elements $v \in M$ the vector $[x, \iota y \iota]v$ is a nonzero element of the proper submodule $\mathrm{Rad}(A)M = \mathrm{Rad}(M)$. Putting the bounds together, the algorithm finds a proper submodule with probability at least $0.214 \cdot \frac{3}{4} \cdot \frac{1}{2} > 0.08$. □

The proposed method, complemented with the Holt-Rees approach gives an algorithm of Las Vegas type for every case.

We now give the promised proof of the statement used above.

Lemma 5.5. *Assume that the finite dimensional F-algebra A with identity satisfies conditions* (5.1) *and* (5.2)*. Assume further that x is an element of A and ι is an idempotent of the subalgebra of A generated by x and 1_A such that the subalgebra of $A/RadA$ generated by $\iota x + Rad(A)$ is $(\iota + Rad(A))\mathrm{Z}(A/Rad(A))$. Then $[x, \iota A \iota] \subseteq Rad(A)$ and $[x, \iota y \iota] \neq 0$ for at least $1 - \frac{1}{|E|}$ of the elements $y \in A$.*

Proof. Let A_x denote the subalgebra of A generated by ιx. We first note that ι is the identity element of A_x. Indeed, $\iota a = a \iota = a$ holds for every element $a \in A_x$. On the other hand, it is straightforward to see that $A'_x = A_x + F\iota$ is a subalgebra and A_x is an ideal of A'_x. By the assumption

$$(A'_x + \mathrm{Rad}(A))/\mathrm{Rad}(A) = (A_x + \mathrm{Rad}(A))/\mathrm{Rad}(A) \cong Z(A/\mathrm{Rad}(A)),$$

thus A'_x is a local algebra and A_x is not a nilpotent ideal. But since in a local algebra every proper ideal is contained in the radical, $A_x = A'_x$, establishing the containment $\iota \in A_x$.

We are now going to replace S and E with appropriate conjugates in order to achieve the situation where $\iota \in S$ and ιE is a subalgebra of A_x. By the Wedderburn–Malcev principal theorem, $A_x = S_x + \mathrm{Rad}(A_x)$, where S_x is a semisimple subalgebra of A_x. Since every maximal semisimple subalgebra of A is a conjugate of S by an inner automorphism (cf. [76]), there exists a unit $a \in A$ such that $S^a = a^{-1}Sa \geq S_x$. Because conditions (5.1) and (5.2) are invariant under automorphisms, we may replace S with S^a and E with E^a, or, equivalently, assume that $S_x \leq S$. Note that ι is just the identity element of S_x.

By the assumption

$$(A_x + \mathrm{Rad}(A))/\mathrm{Rad}(A) = (\iota + \mathrm{Rad}(A))\mathrm{Z}(A/\mathrm{Rad}(A)) \cong E,$$

is a simple algebra, therefore $\mathrm{Rad}(A_x+\mathrm{Rad}(A)) = \mathrm{Rad}(A)$. On the other hand, $\mathrm{Rad}(A_x)+\mathrm{Rad}(A)$ is obviously a nilpotent ideal of $A_x+\mathrm{Rad}(A)$. It follows that $\mathrm{Rad}(A_x) \leq \mathrm{Rad}(A)$, $A_x+\mathrm{Rad}(A) = S_x+\mathrm{Rad}(A)$ and $S_x = \iota E$.

Observe that, since the idempotent ι commutes with x, for every $y \in A$ we have

$$[x, \iota y\iota] = x\iota y\iota - \iota y\iota x = x\iota\iota y\iota - \iota y\iota\iota x = \iota x\iota y\iota - \iota y\iota\iota x = [\iota x, \iota y\iota].$$

The equality $\iota E = S_x$ and the preceding lemma give $[S_x, \iota A\iota] \subseteq \mathrm{Rad}(A)$. Since $S_x \leq A_x \leq S_x + \mathrm{Rad}(A)$, we have

$$[A_x, \iota A\iota] \subseteq [S_x, \iota A\iota] + \mathrm{Rad}(A) \subseteq \mathrm{Rad}(A).$$

The first inclusion of the formula above holds because $\mathrm{Rad}(A)$ is a two-sided ideal and hence $[\iota A\iota, \mathrm{Rad}(A)] \subseteq \mathrm{Rad}(A)$. In particular, $[x, \iota A\iota] = [\iota x, \iota A\iota] \subseteq \mathrm{Rad}(A)$. So we have proved the first part of the statement.

In order to see the second part, notice that, since $S_x \leq A_x$,

$$\mathrm{C}_{\iota A\iota}(x) = \mathrm{C}_{\iota A\iota}(A_x) \leq \mathrm{C}_{\iota A\iota}(S_x) < \iota A\iota.$$

The latter inclusion is strict because not the whole $\iota A\iota$ commutes with $S_x = \iota E$ by Lemma 5.3. Obviously, $\mathrm{C}_{\iota A\iota}(S_x)$ is an S_x-submodule of $\iota A\iota$ (multiplication by elements from S_x from the left hand side). The set of elements y such that $[x, \iota y\iota] = 0$ is the F-linear subspace $(1_A - \iota)A + A(1_A - \iota) + \mathrm{C}_{\iota A\iota}(x)$. By the preceding argument the codimension of this subspace is at least $\dim_F S = \dim_F E$, whence the second part of the assertion follows. □

Chapter 6

Computing the radical of matrix algebras over finite fields

In this chapter, based on the paper [54], we discuss randomized algorithms which compute algebra generators of a Wedderburn complement as well as ideal generators of the radical of a matrix algebra over a finite field given by algebra generators. The cost of the algorithms is comparable to that of a polylogarithmic number of matrix multiplications.

Concerning fast randomized computations in matrix algebras over finite fields given by generators, the first results are due to Eberly and Giesbrecht [30, 31]. They presented randomized algorithms for determining the structure of semisimple matrix algebras over finite fields using a few (i.e. $(\log n)^{O(1)}$) matrix multiplications. These algorithms are nearly optimal, one cannot expect substantially faster methods. The most important result of [30, 31] is valid for arbitrary matrix algebras over finite fields: they can efficiently find a complete system of pairwise orthogonal primitive idempotents.

In Chapter 4 we described a randomized algorithm for computing the radical of matrix algebras over a wide range of ground fields. The number of matrix multiplications performed by the method presented therein is roughly $O(n^4)$ if we ignore the cost of producing random algebra elements and that of computing squarefree part of polynomials. In this chapter, based on the algorithm of Eberly and Giesbrecht, we give an improvement on this result in the special case where the ground field is finite. Furthermore, we also compute a subalgebra which is isomorphic to the radical-free part.

We fix some notation used throughout this chapter. The ground field (which is a finite field) is denoted by F. Following the notation introduced in Section 2.1, we assume that $O(MM(n))$ operations are sufficient to multiply two $n \times n$ matrices over F. We also make the following (very reasonable) assumptions on the function $MM(n)$: $MM(n_1) + MM(n_2) = O(MM(n_1 + n_2))$ and $n_1^2 MM(n_2) = O(MM(n_1 n_2))$. Notice that these assumptions hold for the examples $O(n^3)$ or $O(n^{2.376})$ given in Section 2.1.

Let $A \leq M_n(F)$ be a matrix algebra containing the n by n identity matrix I_n. We denote by V the vector space of the length n column vectors. We assume that A is given by matrices $a_1, \ldots, a_m$, which, together with the identity matrix I_n generate A as an F-algebra.

Like the MeatAxe procedure discussed in the preceding chapter, the algorithm of Eberly and Giesbrecht [30, 31] for finding a complete system of pairwise orthogonal primitive idempotents presumes the presence of a method for selecting random elements of A independently according to a (nearly) uniform distribution. The number of arithmetical operations in F performed by the algorithm is $O((MM(n) + n^2 \log |F| + R(A))\mathrm{polylog}\, n)$, where $R(A)$ stands for the cost of selecting a single random element of A. Note that, since there can be as many as n pairwise orthogonal idempotents, writing the output as a list of matrices would not fit within the desired complexity bound. Instead, the output is given with the aid of a matrix of a basis transformation such that the idempotents written in the new basis are diagonal.

The algorithm is of Monte Carlo type, i.e., it may return a false output, although with an error probability which can be made arbitrarily small. (Actually, the only possible error is that not all of the idempotents in the system are primitive.) In the semisimple case Eberly and Giesbrecht also showed how to supplement the algorithm with a randomized correctness test of cost within the same complexity bound. This upgrades the algorithm for a semisimple algebra A to a Las Vegas method, i.e., a randomized algorithm which may report failure (with a small error probability) but never returns a false output.

In this chapter we give positive answers to a part of the questions posed in [31]. By the Wedderburn–Malcev principal theorem there exists a subalgebra $S \leq A$ such that $S \cap \mathrm{Rad}(A) = (0)$ and $A = S + \mathrm{Rad}(A)$. Furthermore, any pair of such subalgebras are conjugated by an element of the form $1+r$ where $r \in \mathrm{Rad}(A)$. We refer to such subalgebras S as *Wedderburn complements* of A. Obviously, a subalgebra $S \leq A$ is a Wedderburn complement iff $S \cong A/\mathrm{Rad}(A)$. Also, it is an easy consequence of conjugacy part of the principal theorem that Wedderburn complements are just the maximal semisimple subalgebras of A. Note that to construct a Wedderburn complement of A we do not need to work with the whole algebra A. Indeed, if A' is a subalgebra such that $A' + \mathrm{Rad}(A) = A$ then every Wedderburn complement of A' is a Wedderburn complement of the whole A.

It will be convenient to introduce the following concept. Let S be a Wedderburn complement of A. Then the map $\sigma_S : A \to A$ which is the identity on S and zero on $\mathrm{Rad}(A)$ is an algebra epimorphism from A to S. We refer to σ_S as a *Wedderburn projection* of A. Assume that $S \leq M_n(F)$ is a semisimple matrix algebra containing the identity matrix. By an *absolute Wedderburn projection* to S we mean a map $\sigma : M_n(F) \to M_n(F)$ such that σ restricted to S is the identity map and for any matrix algebra $A' \leq M_n(F)$ having S as a Wedderburn complement (i.e., $A' = S + \mathrm{Rad}(A')$), σ restricted to $\mathrm{Rad}(A')$ is the zero map. For an arbitrary subalgebra $A \leq M_n(F)$, an absolute Wedderburn projection of A is just an absolute Wedderburn projection σ to some Wedderburn complement of A. We require that σ is given by a procedure which computes $\sigma(a)$ for an arbitrary matrix a. By the *complexity* of σ we mean the maximum number of arithmetical operations sufficient to compute $\sigma(a)$ for $a \in M_n(F)$. Note that we do not require σ to be linear on the whole $M_n(F)$. The advantage of the concept is that an absolute Wedderburn projection of a sufficiently large subalgebra A' (such that $A' + \mathrm{Rad}(A) = A$) is automatically a Wedderburn projection of the whole A.

Our main result is an efficient method for finding an absolute Wedderburn projection of complexity roughly $O(MM(n)\mathrm{polylog}\, n)$.

Theorem 6.1. *Assume that matrices $e_1, \ldots, e_s$ are given such that $e_1, \ldots, e_s$ form a complete system of pairwise orthogonal primitive idempotents of A. Then an absolute Wedderburn projection σ of A can be constructed by a Las Vegas algorithm performing $O(m(MM(n) + n^2 \log |F|)\mathrm{polylog}\, n)$ operations. The complexity of σ is also $O((MM(n) + n^2 \log |F|)\mathrm{polylog}\, n)$.*

Applying σ to the generators $a_1, \ldots, a_m$, it is obvious that $\sigma(a_1), \ldots, \sigma(a_m)$ (together with the identity matrix) generate the Wedderburn complement $\sigma(A)$ as an F-algebra.

Also, $a_1 - \sigma(a_1), \ldots, a_m - \sigma(a_m)$ generate $\mathrm{Rad}(A)$ as an ideal of A.

Corollary 6.2. *Keeping the assumptions of the theorem, m matrices which generate Rad(A) as an ideal of A as well as m matrices which generate a Wedderburn complement of A as an algebra with identity can be calculated by a Las Vegas algorithm performing $O(m(MM(n) + n^2 \log |F|)\mathrm{polylog}\, n)$ arithmetical operations in F.*

Combining with the result of [31], we obtain

Corollary 6.3. *Assume that we have an auxiliary method which produces random elements of A independently and uniformly at the cost of $O(R(A))$ operations per each element. Then m matrices which generate Rad(A) as an ideal of A as well as m matrices which generate a Wedderburn complement of A as an algebra with identity can be calculated by a Monte Carlo algorithm performing $O((m(MM(n) + n^2 \log |F|) + R(A))\mathrm{polylog}\, n)$ arithmetical operations in F.*

It turns out that the map σ returned by the algorithm of Theorem 6.1 (if it succeeds) is always a Wedderburn projection to a semisimple subalgebra of A. Thus, to upgrade the algorithm to a Las Vegas method it is sufficient to test nilpotency of the (right) ideal generated by $a_1 - \sigma(a_1), \ldots, a_m - \sigma(a_m)$. In this direction we have the following result.

Theorem 6.4. *Let $A \leq M_n(F)$ be a matrix algebra given by generators $a_1, \ldots, a_m$. Let $b_1, \ldots, b_{m'}$ be further elements of A. Assume that S is a subalgebra of A containing the identity matrix such that S and $b_1, \ldots, b_{m'}$ generate A. Suppose further that we have an auxiliary procedure for generating random elements of S uniformly and independently at the cost of $O(R(S))$ operations per each random element. Then there exists a Las Vegas algorithm performing $O(((m+m')n^3 + nR(S)) \log_{|F|} n)$ operations which either detects that not all of the matrices b_i are in Rad(A), or constructs a chain of A-submodules $(0) = V_0 \leq V_1 \leq \ldots \leq V_n = V$ such that $b_i V_j \leq V_{j+1}$ for every $i \in \{1, \ldots, m'\}$ and for every $j \in \{1, \ldots, n\}$.*

Note that a divide and conquer method based on iterative application of the MeatAxe procedure gives a composition series of V. No accurate complexity analysis of such an algorithm can be found in the literature. We think that the implementation in the C-MeatAxe package requires $\Omega(n^4)$ operations for certain nilpotent algebras. A chain of submodules with the properties stated in the theorem together with a complete system of primitive idempotents seem to be applicable to construct a composition chain with $O(n^3 \mathrm{polylog}\, n)$ operations.

Such a chain of submodules witnesses that all the elements b_i are in the radical. Indeed, b_i must be upper triangular in terms of a basis of V compatible with the chain. We shall also see in Section 6.3 that for a Wedderburn-complement S constructed by the method of Theorem 6.1, after a preprocessing of cost $O(mn^3)$ operations, random elements can be drawn using $R(S) = O(n^2)$ operations per each. Thus, by combining Theorem 6.4 with Corollary 6.3 we obtain the following.

Corollary 6.5. *Under the assumptions of Corollary 6.3, m matrices which generate Rad(A) as an ideal of A as well as m matrices which generate a Wedderburn complement of A as an algebra with identity can be calculated by a Las Vegas algorithm performing $O((m(n^3 + n^2 \log |F| + R(A)))\mathrm{polylog}\, n)$ arithmetical operations in F.*

In Section 6.1, we give a proof of Theorem 6.1 for local algebras. In Section 6.2, using techniques similar to those given in [31], we describe a reduction to the local case. The proof of Theorem 6.4 is given in Section 6.3.

We remark that one could propose variants of the algorithms given in this chapter which use random elements of A instead of the generators. The complexity bounds would involve terms $R(A)$ in place of the multiplicative factor m. We have chosen a presentation in terms of generators because – although these variants might be more efficient in practice – it is not known how to generate random elements of A efficiently in a mathematically rigorous way. Also, we hope that the algorithms presented here can contribute to eliminating the random choice of elements of the whole algebra from the algorithm of [30, 31] for finding a complete system of primitive idempotents.

6.1 Wedderburn complements in local algebras

Recall that A is a local algebra if $A/\mathrm{Rad}(A)$ is a field. Throughout this section we assume $A \leq M_n(F)$ is a local algebra. Then a Wedderburn complement S of A is a subfield of $M_n(F)$ (containing the identity matrix).

Let $C = C_A(S)$. Then S is a Wedderburn complement of C and $S \leq Z(C)$. By the conjugacy part of the principal theorem, S is the unique Wedderburn complement of C. We call a matrix $b \in M_n(F)$ semisimple if the matrix algebra generated by b is semisimple. Since S is the unique maximal semisimple subalgebra of C, it is just the set of semisimple elements. It follows that every element $b \in C$ can be uniquely written in the form $b = b_s + b_n$ where $b_s \in S$ and $b_n \in \mathrm{Rad}(C)$, or, equivalently, b_s is a semisimple matrix and b_n is a nilpotent matrix from C. Of course, this construction also works for the subalgebra generated by b in place of C. The decomposition $b = b_s + b_n$ is referred as the Jordan decomposition of b. The matrices b_n and b_s are called the nilpotent part and the semisimple part of b, respectively. The Jordan decomposition can be computed by with $O((MM(n) + n \log |F|)\mathrm{polylog}\, n)$ operations using the Las Vegas rational Jordan normal form algorithm of Giesbrecht [41]. Thus the map $b \mapsto b_s$ is a good Wedderburn projection of C.

We use a simple version of the Fitting decomposition technique of Chapter 4. This gives a well defined subspace of A complementary to C as follows. Since S is a finite field there exists an element $a \in S$ which generate S as an F-algebra. Of course, a is a semisimple matrix. The linear map $\mathrm{ad}_a : b \mapsto ab - ba$ is a semisimple linear transformation on $M_n(F)$. Therefore $M_n(F) = \ker \mathrm{ad}_a + \mathrm{im}\, \mathrm{ad}_a$, a direct sum as vector spaces. Note that this decomposition is inherited by any subspace of $M_n(F)$ invariant under ad_a, such as A and $\mathrm{Rad}(A)$. Also note that $\ker \mathrm{ad}_a = C_{M_n(F)}(a)$. Hence $A = C + \mathrm{ad}_a(A)$, a direct sum as vector spaces. Since $A/\mathrm{Rad}(A)$ is commutative, $\mathrm{ad}_a(A)$ is in fact a subspace of $\mathrm{Rad}(A)$.

In this chapter Φ_a denotes the map which is the identity on $C_{M_n(F)}(a)$ and zero on $\mathrm{im}\, \mathrm{ad}_a$. A Las Vegas algorithm for computing $\Phi_a(b)$ with $O(MM(n)\mathrm{polylog}\, n)$ operations is described in Section 4.4.

It is straightforward to check that the composition map $\sigma_a : b \mapsto (\Phi_a(b))_s)$ is zero on $\mathrm{Rad}(A)$ and maps S identically onto itself. We obtained the following.

Proposition 6.6. *Assume that we have a matrix a such that the matrix algebra S generated by a and the identity matrix is a field. Then the map $\sigma_a : M_n(F) \to M_n(F)$ given above is an absolute Wedderburn projection of S. The complexity of σ_a is $O((MM(n) + n \log |F|)\mathrm{polylog}\, n)$.*

Assume that the matrices $a_1, \ldots, a_m$ generate the local algebra A. It is not difficult to show that if the ground field is sufficiently large then the semisimple part of a random linear combination of $a_1, \ldots, a_m$ will generate a Wedderburn complement with high probability. In the rest of this section we describe an iterative algorithm which works over small fields as well. The method is also based on projecting onto the centralizer and then taking the semisimple part.

We calculate a sequence $c_1, \ldots, c_m$ of semisimple elements of A such that the subalgebra S_i of A generated by c_i satisfies $a_1, \ldots, a_i \in S_i + \mathrm{Rad}(A)$. Then $a = c_m$ will be a semisimple matrix such that for the subalgebra S generated by a we have $S + \mathrm{Rad}(A) = A$. Since S is semisimple $S \cap \mathrm{Rad}(A) = (0)$ and hence S is in fact a Wedderburn complement.

We start with $c_0 = I_n$, the n by n identity matrix. Assume that $0 \leq i < m$ and we have already calculated a matrix c_i with the desired property. Then the subalgebra S_i generated by c_i is semisimple and hence $S_i \cap \mathrm{Rad}(A) = (0)$. Therefore the natural projection $A \to A/\mathrm{Rad}(A)$ embeds S_i into $A/\mathrm{Rad}(A)$, which is a field. We first calculate $b_{i+1} = \Phi_{c_i}(a_{i+1})$ and take the semisimple part d_{i+1} of b_{i+1}. Then d_{i+1} is a semisimple element of A commuting with S_i. Since commutative algebras generated by semisimple matrices are semisimple, the algebra S_{i+1} generated by S_i and d_{i+1} is semisimple.

We claim that for every $j \in \{1, \ldots, i+1\}$, $a_j \in S_{i+1} + \mathrm{Rad}(A)$. For $j \leq i$ it is obvious from the inductive hypothesis as $S_i \leq S_{i+1}$. It remains to establish the containment $a_{i+1} \in S_{i+1} + \mathrm{Rad}(A)$. Since $A/\mathrm{Rad}(A)$ is commutative, $\mathrm{ad}_{c_i} A \subseteq \mathrm{Rad}(A)$ and hence $a_{i+1} - \Phi_{c_i}(a_{i+1}) \in \mathrm{Rad}(A)$. Thus it is sufficient to show that $b_{i+1} = \Phi_{c_i}(a_{i+1}) \in S_{i+1} + \mathrm{Rad}(A)$. But $b_{i+1} - d_{i+1}$ is nilpotent. In particular $b_{i+1} - d_{i+1} + \mathrm{Rad}(A)$ is a nilpotent element of $A/\mathrm{Rad}(A)$. Since $A/\mathrm{Rad}(A)$ is a field, this implies $b_{i+1} - d_{i+1} \in \mathrm{Rad}(A)$ and hence $b_{i+1} \in S_{i+1} + \mathrm{Rad}(A)$. We have proved the claim.

To finish, we need an element $c_{i+1} \in S_{i+1}$ which generate S_{i+1} as an F-algebra. As by induction S_i is generated by c_i, therefore S_{i+1} is a field generated by c_i and d_{i+1}, this can be done as described below.

Lemma 6.7. *Given matrices $c, d \in M_n(F)$ such that the matrix algebra S generated by c and d is a field, a single matrix which generates S can be calculated by a Las Vegas algorithm performing $O(MM(n)\log n + n^2 \log n \log\log n)$ operations.*

Proof. We adopt a version of the method of Giesbrecht [41] proposed for evaluating univariate matrix polynomials. Let S_c and S_d be the subfields of S generated by c and d, respectively. Set $r = \dim_F S$, $r_c = \dim_F S_c$, and $r_d = \dim_F S_d$. Then r is the least common multiple of r_c and r_d and the vectors $d^j c^i$ ($i = 0, \ldots, r_c - 1$, $j = 0, \ldots, r/r_c - 1$) form a basis of S over F. Choose random coefficients α_{ij} from F uniformly and independently. Then the linear transformation $h = \sum_{i,j} \alpha_{ij} d^j c^i$ generates S with probability at least $1/4$ (cf. [42] , Theorem 5.2). We compute the matrix of h in terms of a special basis of $V = F^n$. When it is done we write h in terms of the natural basis at the cost of $O(MM(n))$ operations.

Assume we already know r_c, r_d and r. (These dimensions can be computed using the Frobenius normal form of c and d, which will be needed for other purposes, too.) $V = F^n$ can be considered as a vector space of dimension n/r over S. The probability of that n/r random vectors $v_1, \ldots, v_{n/r}$ (chosen uniformly and independently from V) are linearly independent over S is

$$\begin{aligned}(1 - |S|^{-n/r})(1 - |S|^{1-n/r}) \cdots (1 - |S|^{-1}) \\ > \prod_{i=1}^{\infty}(1 - |S|^{-i}) \geq \prod_{i=1}^{\infty}(1 - 2^{-i}) \quad > \quad 0.28.\end{aligned}$$

Suppose that the vectors $v_1, \ldots, v_{n/r}$ are linearly independent over S. Then the vectors $d^j c^j v_l$ ($l = 1, \ldots, n/r$, $i = 0, \ldots, r_c - 1$, $j = 1, \ldots, r/r_c - 1$) form a basis of V. In this basis the (l, i, j)th column of h will be the vector $\sum_{i',j'} \alpha_{i'j'} d^{j'} c^{i'} d^j c^i v_l = d^j c^i w_l$, where $w_l = \sum_{i,j} \alpha_{ij} d^j c^i v_l$. Note that in the basis $d^j v^i v_l$ the vector w_l is just the column vector having α_{ij} in the (l, i, j)th position and zero elsewhere. Thus computing h is equivalent to computing the vectors $d^j c^i w_l$, for $l = 1, \ldots, n/r$, $i = 0, \ldots, r_c - 1$, $j = 1, \ldots, r/r_c - 1$. This can be done as follows.

By (a slightly simplified version of) the algorithm of [41] for Frobenius normal forms, we can find matrices $u_c, u_d \in \mathrm{GL}_n(F)$ such that both $u_c^{-1} c u_c$ and $u_d^{-1} d u_d$ have at most $2n$ nonzero entries. The algorithm requires $O(MM(n) \log n + n^2 \log n \log \log n)$ operations. The simplification is that, in spite of the general case of the algorithm, we do not need to go over an extension field if F is small. Indeed, if we choose random vectors $w_1, \ldots, w_n$, the first n/r_c (resp. n/r_d) of them will form a basis of V over the field S_c (resp. S_d) with probability at least $1/2$.

Having u_c and u_d at hand, we start with computing $u_c^{-1} v_l$. This requires $O(MM(n))$ operations. Then we compute $u_c^{-1} c^i v_l = (u_c^{-1} c u_c)^i (u_c^{-1} v_l)$ for $i = 0, \ldots, r_c - 1$ iteratively. This can be done with $O(n r_c n/r)$ operations because the cost of multiplying a vector by $u_c^{-1} c u_c$ is $O(n)$ as $u_c^{-1} c u_c$ has only at most $2n$ nonzero entries. Next we multiply these vectors simultaneously by u_c using $O(MM(n))$ operations in order to obtain the vectors $c^i w_l$. From these vectors, computing $d^j c^i v_l$ for all i, j and l can be accomplished in a similar fashion with $O(MM(n) + n(r/r_c) r_c (n/r)) = O(MM(n) + n^2) = O(MM(n))$ operations. We see that the cost of the whole algorithm is dominated by the $O(MM(n) \log n + n^2 \log n \log \log n)$ operations used to compute the normal forms of c and d.

We can test the correctness of the output as follows. Let $v_1, \ldots, v_{n/r}$ be as above. Then the vectors $h^i v_l$ ($i = 0, \ldots, r - 1$, $l = 1, \ldots, n/r$) form a basis of V. This basis can be computed with $O(MM(n) \log n)$ operations using the algorithm of Keller-Gehrig [66]. In this basis h is in Frobenius normal form. Writing c in terms of this basis requires $O(MM(n))$ operations. If $c = \sum_{i=0}^{r-1} \gamma_i h^i$ then $c v_1 = \sum \gamma_i (h^i v_1)$, thus the coefficients γ_i are just the first r entries of the first row of c in the new basis (all the other entries of this raw must be zero). This means that we can read the (possible) coefficients γ_i from the matrix of c. Computing the polynomial $\sum_{i=0}^{r-1} \gamma_i h^i$ (in the basis $h^i v_l$) requires further $O(n^2)$ operations (cf. [41], Lemma 6.1). Finally, testing whether the result equals c requires n^2 comparisons. Testing whether d is a polynomial of h can be accomplished in the same way.

Note that at the cost of further $O(n \log |F|)$ operations we can verify irreducibility of the minimal polynomial of h, which is equivalent to that h generates the field. □

The overall cost of computing the generator $a = c_m$ of a Wedderburn complement of A amounts to $O(m(MM(n) + n \log |F|) \mathrm{polylog}\, n)$ operations. Together with Proposition 6.6, this gives the following.

Proposition 6.8. *Assume that A is local and matrices $a_1, \ldots, a_m$ are given such that $a_1 + Rad(A), \ldots, a_m + Rad(A)$ generate $A/Rad(A)$ as a F-algebra. Then an absolute Wedderburn projection σ of A of complexity $O((MM(n) + n \log |F|) \mathrm{polylog}\, n)$ can be constructed by a Las Vegas algorithm performing $O(m(MM(n) + n \log |F|) \mathrm{polylog}\, n)$ operations. Furthermore, the algorithm calculates a matrix a which generates the Wedderburn complement $\sigma(A)$.*

Remark. Assume that we run this algorithm on an input when the algebra A is not local. If the algorithm succeeds and a has an irreducible minimal polynomial (passes the additional test mentioned at the end of the proof of Lemma 6.7), then $\sigma(A)$ is subfield of A and σ is an absolute Wedderburn projection of $\sigma(A)$. However, σ is in general not linear on A and has very little to do with the structure of A.

6.2 Using a complete system of primitive idempotents

In this section we show how to use a complete orthogonal system of primitive idempotents to reduce the construction of a Wedderburn projection of an arbitrary algebra to the local case. The approach follows the lines of the primary decomposition of A (see Subsection 2.2.3). Assume that $e_1, \ldots, e_s$ form a complete system of pairwise orthogonal primitive idempotents of A. Two primitive idempotents e and f are called equivalent if $e + \mathrm{Rad}(A)$ and $f + \mathrm{Rad}(A)$ lie in the same simple component of $A/\mathrm{Rad}(A)$. In other words, e_i and e_j are equivalent if $eAf \not\subseteq \mathrm{Rad}(A)$ (or, equivalently, $fAe \not\subseteq \mathrm{Rad}(A)$). Let $f_1, \ldots, f_t$ denote the sums of the equivalence classes of $e_1, \ldots, e_s$. Then the primitive central idempotents of $A/\mathrm{Rad}(A)$ are $f_1 + \mathrm{Rad}(A), \ldots, f_t + \mathrm{Rad}(A)$ and hence $A = f_1Af_1 + \ldots + f_tAf_t + N'$, a direct sum as vector spaces, where $N' = \sum_{i \neq j} f_iAf_j$. Recall that the components f_iAf_i are pairwise orthogonal primary subalgebras of A and N' is a linear subspace of $\mathrm{Rad}(A)$. The following observation is obvious.

Lemma 6.9. *Using the notation introduced above, assume that for every $l \in \{1, \ldots, t\}$, the map $\sigma_l : f_lAf_l \to f_lAf_l$ is a Wedderburn projection of f_lAf_l. Then the map $\sigma : A \mapsto A$ which is zero on N' and coincides with σ_l on f_lAf_l is a Wedderburn projection of A.*

For primary algebras we have the following. Assume that A is primary. Then, by Theorems 49.4 and 26.8 of [67], $A \cong M_s(A_1)$ for some local algebra A_1 and $\mathrm{Rad}(M_s(A_1)) = M_s(\mathrm{Rad}(A_1))$ (the subalgebra consisting of matrices with entries from $\mathrm{Rad}(A)$). As a consequence, if S_1 is a Wedderburn complement in A_1 then $M_s(S_1)$ (the subalgebra of matrices with all entries from S_1) is a Wedderburn complement of $M_s(A_1)$. Hence the following is straightforward.

Lemma 6.10. *Assume that A_1 is a local algebra and σ_1 is a Wedderburn projection of A_1. Then the map $\sigma : M_s(A_1) \to M_s(A_1)$ given as $\sigma((u_{ij})_{i,j=1}^s) = (\sigma_1(u_{ij}))_{i,j=1}^s$ is a Wedderburn projection of $M_s(A_1)$.*

In order to accomplish a construction suggested by the two lemmas above, we need to determine the equivalence relation between the given idempotents, and then, for each primary component, an explicit isomorphism with a full matrix algebra over a local algebra. In an interpretation in the spirit of Chapter 4, the complete set of idempotents corresponds to a special torus T (a maximal *split* torus rather than a maximal torus) of A and determining the equivalence relation corresponds to computing the semi-central part of T.

Throughout this section we assume that a complete system $e_1, \ldots, e_s$ of pairwise orthogonal primitive idempotents of A is given as a part of the input. We assume that $e_1, \ldots, e_s$ are given with the aid of a new basis of V such that the e_is written in that basis are diagonal matrices of the form $e_i = \mathrm{diag}(0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)$. The reader is referred to the papers [30, 31] for the details of a Monte Carlo algorithm for finding such a system with $O((MM(n) + n^2 \log |F| + R(A)) \mathrm{polylog}\, n))$ operations.

It will be convenient to supplement the system $a_1, \ldots, a_s$ of generators with $a_0 = I_n$. We shall represent linear transformations in terms of the basis described above. Writing

$a_1, \ldots, a_m$ in terms of that basis can be accomplished with $O(mMM(n))$ operations by conjugating by the appropriate basis transformation matrix. Similarly, if the result of any computation consists of m' n by n matrices, we can write it back in terms of the natural basis of V at the cost of $O(m'MM(n))$ operations.

The following observation is obvious.

Lemma 6.11. *The set $\{e_i a_l e_j | i, j = 1, \ldots, s,\ l = 0, \ldots, m\}$ generate A as an F-algebra.*

Note that in terms of the basis we are working with, for an arbitrary matrix a, the product $e_i a e_j$ is zero except in its i,jth block, which is equal to that of a. With some abuse of notation, sometimes we will also denote the i,jth block of a by $e_i a e_j$.

Our first aim is to determine the equivalence classes of $e_1, \ldots, e_s$.

Lemma 6.12. *Let e and f be two orthogonal primitive idempotents of A and $a \in A$. Then either $eaf \in Rad(A)$ or the restriction of eaf to fV is an F-linear isomorphism $fV \cong eV$. To be more specific, in the latter case there exists an element $b \in A$ such that the map $fbe : eV \to fV$ is an inverse of the map $eaf : fV \to eV$.*

Proof. Assume that $eaf \notin \mathrm{Rad}(A)$. Then there exists an element $d \in A$ such that $deaf$ is not nilpotent. An easy induction shows that $(deaf)^k = (fdeaf)^k + (1-f)deaf(fdeaf)^{k-1}$ for every positive integer k. Hence $fdeaf$ is not nilpotent either. Since fAf is a local algebra, this means that $fdeaf$ is a unit in fAf: there exists an element $c \in A$ such that $f = fcffdeaf = fcfdeeaf$. Put $b = cfd$. Then $f = fbeeaf$ and, since f acts on fV as the identity, fbe is indeed an inverse of eaf. □

Note that the "or" in the first sentence of the lemma above is not exclusive: it can happen that eaf is in $\mathrm{Rad}(A)$ but eaf is still a regular map from fV to eV. However, in this case there is no inverse of eaf from fAe. In particular we have the following useful characterization of equivalence. It is a generalization of Lemma 3.4 of [30].

Lemma 6.13. *Let e and f two orthogonal primitive idempotents of A. Then e and f are equivalent if and only if there exist elements $a, b \in A$ such that both of the linear maps $eaf : fV \to eV$ and $fbe : eV \to fV$ are F-linear isomorphisms between the vector spaces eV and fV.*

Proof. The "only if" part of the statement follows from the preceding lemma. To establish the reverse implication, assume that e and f are not equivalent. Then both eAf and fAe are subspaces of $\mathrm{Rad}(A)$ and hence $fAeeAf \subseteq \mathrm{Rad}(A)$. In particular, the product $fbeeaf$ is nilpotent for $a, b \in A$. This implies that either fbe or eaf is singular. □

The lemmas provide us with a tool for determining the equivalence classes of idempotents $e_1, \ldots, e_s$ efficiently. We define a directed graph G on vertices $e_1, \ldots, e_s$ as follows. Let $e_j \to e_i$ be an edge if there exists a generator $a \in \{a_0, a_1 \ldots, a_m\}$ such that $e_i a e_j$ is a linear isomorphism $e_j V \to e_i V$.

Lemma 6.14. *The strongly connected components of the graph G defined above are the equivalence classes of the idempotents $e_1, \ldots, e_s$.*

Proof. By taking compositions, the existence of a path in G from e_j to e_i clearly implies the existence of an element $a \in A$ such that $e_i a e_j$ is a linear isomorphism $e_j V \cong e_j V$. Thus, by Lemma 6.13, if e_i and e_j are in the same strongly connected component then e_i and e_j are equivalent idempotents.

To establish the reverse implication, observe that e_iAe_j is the linear span of all the products of the form $e_{j_1}a_{l_1}e_{j_2}a_{l_2}e_{j_2}\cdots e_{j_u}a_{l_u}e_{j_{u+1}}$, where u is a positive integer, $j_1 = i$, $j_{u+1} = j$, each j_h is from $\{1,\ldots,s\}$, and each l_h is from $\{0,\ldots,m\}$. Therefore if e_i and e_j are equivalent there exists such a product which is not in Rad(A). Since Rad(A) is a two sided ideal, none of the components $e_{j_h}a_{l_h}e_{j_{h+1}}$ of that product is in Rad(A). By Lemma 6.12, the map $e_{j_h}a_{l_h}e_{j_{h+1}}$ is a linear isomorphism $e_{j_{h+1}}V \cong e_{j_h}V$ for every $h \in \{1,\ldots t\}$. This means that there is a path from j to i. The existence of a path in the reverse direction can be proved in the same way. □

To determine whether the block $e_ia_he_j$ is nonsingular (of course, only in the case if this block is a square, i.e., e_i and e_j are of the same rank) requires $O(MM(\mathrm{rk}\,(e_i)))$ operations. Assume that for every positive integer r the number of idempotents e_i of rank r is u_r. Then the overall cost of determining the edges of G is

$$\sum_{r=1}^{\infty} u_r^2 O(MM(r)) = \sum_{r=1}^{\infty} O(MM(ru_r)) = O(MM(\sum_{r=1}^{\infty} ru_r)) = O(MM(n))$$

by the assumptions on the function $MM(n)$. The strongly connected components can be determined by the usual algorithms based on depth first search at uniform cost $O(n^2)$.

Assume now that the components are $E_1,\ldots,E_t$. For every $l \in \{1,\ldots,t\}$, let f_l stand for the sum $\sum_{e_i\in E_l} e_i$. Then the primary components of A are $f_1Af_1,\ldots,f_tAf_t$. Effective faithful matrix representations of f_lAf_l are given by the actions on f_lV. These just correspond to embeddings into the blocks $f_lM_n(F)f_l \cong M_{r_l}(F)$ (where r_l stand for the rank of f_l). Assume that for every index $1 \le l \le t$ the map $\sigma_l : f_lM_n(f)f_l$ is an absolute Wedderburn projection of f_lAf_l. Then, by Lemma 6.9, the direct sum $\sigma : M_n(F) \mapsto M_n(F)$, which is zero on the non-diagonal blocks $f_lM_n(F)f_{l'}$ ($l \neq l'$) and σ_l on the diagonal blocks $f_lM_n(F)f_l$ is an absolute Wedderburn projection of A. Furthermore, if the complexity of σ_l is $C_l = O((MM(r_l) + r_l^2 \log|F|)\mathrm{polylog}\, r_l)$ ($l = 1,\ldots,t$), then the complexity of σ amounts to $\sum_{l=1}^{t} C_l = O((MM(n)+n^2\log|F|)\mathrm{polylog}\, n)$ by the assumptions on the function $MM(n)$. A similar statement holds for the cost of computing the data structure representing σ.

Thus it is sufficient to construct Wedderburn projections (and prove the complexity bounds stated in Theorem 6.1) for the primary components f_lAf_l separately. In general, we cannot give generators for the whole f_lAf_l efficiently. However, it will be sufficient to work with somewhat smaller subalgebras which are equal to f_lAf_l modulo the radical. Let A_l stand for the subalgebra generated by the set $\{e_ia_he_j | e_i, e_j \in E_l,\ h = 0,\ldots,m\}$.

Lemma 6.15. *For every $l \in \{1,\ldots,t\}$, A_l is a subalgebra of f_lAf_l and $A_l + Rad(f_lAf_l) = f_lAf_l + Rad(f_lAf_l)$.*

Proof. The inclusion $A_l \le f_lAf_l$ is obvious. To prove the second part, observe that f_lAf_l is the linear span of products of elements of the form $e_ia_he_j$, where $h \in \{0,\ldots,m\}$ and $i,j \in \{1,\ldots,s\}$, and the first and last idempotents in the product are from E_l. However, such products containing also idempotents from different $E_{l'}$s are in Rad(A) and hence $A_l + \mathrm{Rad}(A) = f_lAf_l + \mathrm{Rad}(A)$. To finish the proof, observe that $\mathrm{Rad}(f_lAf_l) = f_lAf_l \cap \mathrm{Rad}(A)$. □

Assume that for every $l \in \{1,\ldots,t\}$, $E_l = \{e_1^l,\ldots,e_{s_l}^l\}$. We fix a basis B_1^l of e_1^lV and along a spanning tree in G rooted in e_1^l we propagate this basis to the other subspaces e_i^lV as follows. Assume that $e_j^l \to e_i^l$ is an edge of the spanning tree and B_i^l is already defined. Choose an element e^l_{ij} of the form $e^l_i a_h e^l_j$ ($h \in \{0, \dots, m\}$) which is an isomorphism $e^l_j V \cong e^l_i V$ and set $B^l_i := e^l_{ij} B^l_j$. For a fixed l this procedure requires $O(s_l)$ multiplications of pairs of $n_l \times n_l$ matrices where $n_l = \operatorname{rk} e^l_1 = \dim_K e^l_1 V$. The total cost of calculating the bases for all the components requires $\sum_{l+1}^t O(s_l MM(n_l)) = O(MM(n))$ operations. The union of these bases is a basis of V. From now on we work with this new basis. Again, writing the generators in terms of the new basis can be accomplished be performing $O(mMM(n))$ operations and a similar bound can be given for the operations required to writing the possible results back in terms of the original basis.

The new basis has the following remarkable property. For every $l \in \{1, \dots, t\}$ let e^l_{ij} stand for the matrix which is the n_l by n_l identity matrix in the block corresponding to $e^l_i V$ and $e^l_j V$ and zero otherwise ($i, j = 1, \dots, s_l$).

Lemma 6.16. *For every $l = 1, \dots, t$ and for every $i, j = 1, \dots, s_l$, the matrix e^l_{ij} is an element of A_l.*

Proof. First we note that the matrix e^l_{ij} already defined for a pair i, j of indices such that $e^l_j \to e^l_i$ is an edge of the spanning forest is of the desired form. We claim that for such an edge the matrix e^l_{ji} is in A_l as well. Indeed, e^l_{ji} is the matrix $x \in e^l_j M_n(F) e^l_i$ uniquely determined by the equation $xe^l_{ij} = e^l_j$. We know that there exists an element $a \in e^l_j A_l e^l_i$ such that a is a regular linear map $e^l_i V \to e^l_j V$. It follows that ae^l_{ij} is a unit in $e^l_j A_l e^l_j$: there is an element $b \in e^l_j A_l e^l_j$ such that $bae^l_{ij} = e^l_j$. Then $x = ba \in e^l_j A_l e^l_i$ and $xe^l_{ij} = e_j$. Thus the claim holds. The assertion for the remaining e^l_{ij}s follows as these elements are products of e^l_{ij}s for edges and reverse edges of the spanning forest. □

We remark that the proof of the lemma also suggests an algorithm to give effective presentations of e^l_{ij} in terms of the idempotents e^l_i and the generators a_h.

The following is an immediate consequence of Lemma 6.15.

Lemma 6.17. *The subalgebra $e^l_1 A_l e^l_1$ is generated by the set*

$$\{e^l_{1i} a_h e^l_{j1} | h = 0, \dots, m,\ i, j = 1, \dots, s_l\}.$$

Again, the action of $e^l_1 M_n(F) e^l_1$ on $e^l_1 V$ gives a faithful representation of $e^l_1 M_n(F) e^l_1$. Also, for every matrix $a \in M_n(F)$, the matrix of $e^l_{1i} a e^l_{j1}$ in this representation is just the $((l, i), (l, j))$th block of a (i.e, the block $e^l_i a e^l_j$).

Proof of Theorem 6.1. We have $s_l^2 m$ n_l by n_l matrices which generate the local subalgebra $e^l_1 A_l e^l_1$. By Proposition 6.8, an absolute Wedderburn projection σ^l_1 of $e^l_1 A e^l_1$ can be calculated with $O(ms_l^2(MM(n_l) + n_l \log |F|) \operatorname{polylog} n_l)$ operations. The complexity of σ^l_1 is $O((MM(n_l) + n_l \log |F|) \operatorname{polylog} n_l)$. Using Lemma 6.10, one can see that the map σ_l built from σ^l_1 as an application of σ^l_1 block-wise is an absolute Wedderburn projection of A_l. The complexity of σ_l is $O(s_l^2(MM(n_l) + n_l \log |F|) \operatorname{polylog} n_l)$. By the assumptions on the function $MM(n)$, this is $O((MM(s_l n_l) + s_l^2 n_l \log |F|) \operatorname{polylog} n_l) = O((MM(r_l) + r_l^2 \log |F|) \operatorname{polylog} r_l)$, where $r_l = s_l n_l = \operatorname{rk} f_l$. Similarly, the cost of constructing the data structure of σ_l (which is essentially the block structure of A_l together with the data structure for σ^l_1) is $O(mMM(r_l) + r_l^2 \log |F|) \operatorname{polylog} r_l)$. Combined with the direct sum construction described earlier, this establishes Theorem 6.1. □

Remark If not all of the idempotents $e_1, \dots, e_s$ are primitive in A, the algorithm presented here – if imprimitivity of some idempotent is not discovered – calculates a map

$\sigma : M_n(F) \to M_n(F)$ of the given complexity. Actually, it is an absolute Wedderburn projection of the semisimple subalgebra $\sigma(A)$ of A and maps A into A. Also, we can verify that for every $l = 1, \ldots, t$, the subalgebra $S_l = \sigma(e_1^l A_l e_1^l)$ is a field. Indeed, the algorithm of Section 6.1 gives an explicit generator for S_l and we can test whether the minimal polynomial of this generator is irreducible. We also have an explicit representation of the subalgebra $\sigma(f_l A f_l)$ as a full matrix algebra over S_l.

We conclude this section with a related result.

Lemma 6.18. *The subalgebra $\sum_{l,l'=1}^t e_1^l A e_1^{l'}$ is generated by the set $\{e_{1i}^l a_h e_{j1}^k | h = 0, \ldots, m,\ k, l = 1, \ldots, t,\ i = 1, \ldots, s_l,\ j = 1, \ldots, s_k\}$.*

Again, the only possible nonzero block of $e_{1i}^l a e_{j1}^k$ equals the appropriate block of a. The subalgebra $B(A) = \sum_{l,l'=1}^t e_1^l A e_1^{l'}$ is called the basic algebra of A. It is up to isomorphism independent of the choice of the system of primitive idempotents.

Corollary 6.19. *Matrices which generate $B(A)$ can be calculated with $O(mMM(n))$ operations.*

6.3 Verifying correctness

We keep the notation of Section 6.2. As already mentioned, if the algorithm (supplemented with irreducibility tests) succeeds we can be sure that σ is a Wedderburn projection of the semisimple subalgebra $\sigma(A)$ of A. As $\sigma(A)$ together with the matrices $b_l = a_l - \sigma(a_l)$ $(l = 1, \ldots, m)$ generate A, to prove that σ is a Wedderburn projection of the whole A it is sufficient to verify that $b_1, \ldots, b_m \in \mathrm{Rad}(A)$. A witness of that can be a chain $(0) = V_0 \leq V_1 \leq \ldots \leq V_n = V$ of A-submodules of $V = F^n$ such that $b_i V_j \leq V_{j-1}$ $(i = 1, \ldots, m,\ j = 1, \ldots, n$.

We can work with a special basis of V such that random elements of $\sigma(A)$ can be drawn efficiently. We assume that generators b_l of $S_l = \sigma(e_1^l A e_1^l)$ are given for $l = 1, \ldots, t$, and a bases B_1^l of $e_1^l V$ are computed such that for every l the matrix of b_l in term of B_1^l is in Frobenius normal form. Furthermore, the basis of the whole V is the prolongation of B_1^ls described in the preceding section. Finding such a basis and writing the generators of that basis requires $O(mMM(n)\mathrm{polylog}\, n) = O(mn^3)$ operations. In the new basis random elements of $\sigma(A)$ can be generated uniformly as follows. We fill every "big block" corresponding to the subspace $f_l V$ with n_l^2 independent random polynomials in b_l of degree at most $\dim_F S_l$. Since b_l is in Frobenius normal form, by the results of [41] we can evaluate all of these polynomials at total cost of $O(n^2)$ operations. An application of Theorem 6.4 gives that verifying correctness of the Wedderburn decomposition as well as verifying that all the idempotents are primitive can be done by a Las Vegas algorithm performing $O(mn^3 \log n)$ operations. The rest of this section is devoted to the proof of the theorem.

Proof of Theorem 6.4. Let B stand for the linear span of the matrices $b_1, \ldots, b_{m'}$. We can produce random elements of B by taking linear combinations of $b_1, \ldots, b_{m'}$ with random coefficients at the cost of $O(m'n^2)$ operations. Let $R = BA$ be the right ideal of A generated by B. Note that the products bc $(b \in B,\ c \in S)$ generate R as an F-algebra (without identity). Since $\mathrm{Rad}(A)$ is the largest nilpotent right ideal of A, the subspace B is contained in $\mathrm{Rad}(A)$ if and only if R is nilpotent. The algorithm consists of two parts. It is analogous to the two-phase depth first search method to determine the strongly connected components of a directed graph. In the first part we attempt to calculate a basis $v_1, \ldots, v_n$ of V such that for every index i, the subspace Rv_i is contained in the subspace generated by the first $i-1$ basis vectors $v_1, \ldots, v_{i-1}$. Note that R is nilpotent if and only if such a sequence exists. In the second part we determine the A-submodules V_i generated by the the first i basis vector $v_1, \ldots, v_i$ ($V_0 = (0)$). If R is nilpotent and the first part works correctly then $BU_i = B(Av_i + V_{i-1}) \subseteq BAv_i + V_{i-1} = Rv_i + V_{i-1} \subseteq V_{i-1}$, i.e., we obtain a chain with the desired property.

Both parts make use of a data structure representing a dynamically increasing subspace U of V supporting efficient test for containment of vectors in U as well as adding single vectors to U. (By adding v to U we mean replacing U with the subspace generated by U and v.) The data structure is standard and based on the idea of Gaussian elimination: a semi-echelonized basis of U. The cost of a test whether $v \in U$ $O(n^2)$ as well as the cost of pivoting a new vector when it is added to U. Actually, this is used in the C-MeatAxe program as a part of a procedure for computing bases of submodules given by generators (cf. Algorithm 2.1.18 of [75]). We begin with the description of the second part because it can be implemented by a procedure which is deterministic and essentially the same as the C-MeatAxe subroutine mentioned above. We initialize U to (0) and iteratively for every index $i = 1, \ldots, n$ (in this order) do the following. Test if $v = v_i$ is already in U. If not, add v to U and do the same for $a_1v, \ldots, a_mv$ (recursively, in this order). Since there are exactly n additions to U and for each addition there are at most m vector-by-matrix multiplications and containment tests, the total cost of this part is $O(mn^3)$ operations.

The first part is randomized. During the course of this part, if no error occur due to randomization, U is always an R-submodule of V. Initially $U = (0)$. For every element v of a basis of V (in any fixed order) we test whether v is already in U, and if not, we invoke the recursive procedure described below to compute (in the variable U) the R-submodule $U + Rv$.

The input of the procedure is always a vector $v \in V \setminus U$. Actually, if the depth of the recursive calls is j then this vector is from $R^{j-1}V$. Thus if we check the depth and find that it is bigger than n then we can stop and conclude that R is not nilpotent. Otherwise we do the following. Take a random element b from B and a random element c from S. If the vector bcv is not already in U we invoke the procedure itself with input bcv. After this call we test whether v is in U. If not, we continue with selecting new random elements b and c. Otherwise we stop and test deterministically whether the subspace U_0, the value of U before the invocation, is an R-submodule. This can can be done by computing first the A-submodule AU_0 and then the subspace BAU_0 at total cost of $O((m'+m)n^3)$ arithmetical operations. If U_0 turns out to be a submodule then we conclude that R is not nilpotent. (Indeed, if R is nilpotent then $Rv + U_0$ must be a proper submodule of $v + Rv + U_0$.) If U_0 is not a submodule we report failure because this is a consequence of an error due to randomization.

We repeat this with subsequent random choices for b and C until we observe that $bcv \in U$. If $bcv \in U$ then U contains Rv with sufficiently large probability. We finish the procedure by adding the vector v to U and then return.

The result of this part will be the sequence of the basis vectors of U (in the order of addition to U.) To analyze complexity, observe that each successful return of the recursive procedure can be bound to an addition to a vector to U. Thus the total number of such returns is at most n. (Exactly n, if the whole of this part succeeds.) As we have a control of depth of recursive calls, to total number of the calls is less than $2n$. Also note that, apart of the at most n containment tests outside from the recursive procedure, all matrix by vector multiplications and containment tests can be bound to calls of the recursive procedure. To each call, $O(1)$ such operations are associated. The same holds for producing random elements of S. Thus the total cost of this part if $O(n(n^2 + (R(S)) + m'n^2) + MM(n)) = O((m + m')n^3 + nR(S))$ arithmetical operations in F.

The only possible error can occur at the point where $BSv \not\subseteq U$ but we happen to choose b, c such that $bcv \in U$. The probability of such an event is at most $1/|F|^2$. Thus the probability of that error never occurs is $(1 - \frac{1}{|F|^2})^{O(n)}$ which can improved to a positive constant by $O(\log_{|F|} n)$ repetitions. (Or, by testing containment $bcv \in U$ for $O(\log_{|F|} n)$ random b, c before giving up.)

We see that if the algorithm concludes that R is not nilpotent, the answer is always correct. If a chain $V_0 \leq V_1 \leq \ldots \leq V_n$ is constructed, the result can be checked by writing the generators $a_1, \ldots, a_m$ and $b_1, \ldots, b_{m'}$ in terms of a basis consistent with the chain. (The matrices of the a_is should be block upper triangular while the matrices of the b_is should be strictly block triangular.) This test requires further $O((m + m')MM(n)) = O(m + m')n^3)$ operations. This concludes the proof of the theorem. □

Chapter 7

Constructing module isomorphisms

This chapter, based on a part of the paper [20], joint work with Alexander Chistov and Marek Karpinski. Here we present a deterministic algorithm for testing and constructing isomorphisms between modules.

Module isomorphism corresponds to equivalence of matrix representations. By taking images of a set of algebra generators, it is reduced to the following conjugacy problem. Assume that we are given two collections $a_1, \dots, a_m$ and $a'_1, \dots, a'_m$ of $n \times n$ matrices with entries from the field K. Our task is to find (if exists) a nonsingular matrix $x \in GL_n(K)$ such that

$$xa_ix^{-1} = a'_i$$

for every $i \in \{1, \dots, m\}$. A natural approach to this problem is to consider the set V of $n \times n$ matrices x satisfying

$$xa_i = a'_ix.$$

This condition is equivalent to a system of homogeneous linear equations in the entries of the matrix x, whence V is a linear subspace of $M_n(K)$. Obviously, the conjugacy problem is equivalent to finding a nonsingular matrix in the subspace V. Thus, the conjugacy problem can be considered as a special case of finding matrices of maximum possible rank in linear subspaces of $M_n(K)$, which was formulated by J. Edmonds [32]. Note that if our ground field K is sufficiently large then this problem admits an efficient randomized solution: If there exists a nonsingular matrix in the linear subspace V of $M_n(K)$ then a random matrix from V (i.e., a random linear combination of a basis) is nonsingular with high probability by the Schwartz-Zippel lemma (see Section 2.4). However, no deterministic polynomial time method is known to this general problem. Here we present a deterministic polynomial time algorithm for our particular case, i.e., the conjugacy problem.

We remark that special instances of Edmonds' problem include the case where the subspace V of $M_n(K)$ is spanned by rank one matrices and the case where V is spanned by skew-symmetric matrices of rank two. A mathematical result describing the maximal rank in the former case can be derived from the Matroid Intersection Theorem, while in the latter case the Parity Theorem for linear matroids gives the information on the rank. See the paper [71] of L. Lovász for a discussion of these and further examples. The paper [40] by J. F. Geelen gives deterministic polynomial time algorithms for finding maximal rank matrices in these special instances. M. Domokos, in a sequence of papers starting with [27], presents results part of which are related to invariant theoretic aspects of the problem.

Below we state the main results of this chapter. We need to assume that there is a deterministic polynomial algorithm for computing the Jacobson radical of finite dimensional algebras over K. Important examples for such fields are fields of zero characteristic, finite fields and extensions of constant transcendence degree thereof. We assume A is a finite dimensional algebra with identity over the field K. By an A-module we mean a finite dimensional unital A-module given by matrices of the action of a generating set for A. The complexity of the algorithm will be measured by the number of field operations measured in terms of the dimension of the module.

Our main result is actually a cyclicity test. Recall that an A-module V is *cyclic* if V is generated by a single element, i.e, there exists an element $v \in V$ such that $Av = V$. Note that V is a cyclic module if and only if there exists an epimorphism ${}_AA \to V$. (By ${}_AA$ we denote the regular A-module. This is the space A where the elements of A act by multiplication from the left. The submodules of ${}_AA$ are left ideals of A.)

Theorem 7.1. *Let K be field and let A be a K-algebra and V be a A-module. Given A, Rad(A) and V, one can decide by a deterministic polynomial time algorithm, whether the module V is cyclic. If this is the case the algorithm returns a generator of V.*

As an application, we have the following result on a generalized conjugacy problem. The result can also be used in effective versions of the Skolem-Noether theorem.

Theorem 7.2. *Let A be a K-algebra where K is a field which admits a deterministic polynomial time procedure for computing the radical of finite dimensional K-algebras. Assume that we are given two collections $a_1, \dots, a_m$ and $a'_1, \dots, a'_m$ of elements from A. We can decide in deterministic polynomial time whether there exists an element $x \in A^*$ such that $xa_ix^{-1} = a'_i$ for every $i = 1, \dots, m$, and exhibit such an element if one exists.*

Since the module isomorphism problem is equivalent to the conjugacy problem in the full matrix algebra $M_n(K)$, we obtain the following.

Corollary 7.3. *Let A be a K-algebra, K is a field which admits a deterministic polynomial time procedure for computing the radical of finite dimensional K-algebras. Assume that we are given two A-modules V and W. Then one can decide in deterministic polynomial time whether V and W are isomorphic, and if it is the case then construct an isomorphism between these two modules.*

Note that in [21], a deterministic polynomial time method for testing isomorphism of semisimple modules is given. The present method is constructive and works for arbitrary modules.

We prove Theorem 7.1 first for semisimple modules (Sec. 7.1). Then, in Section 7.2, we "lift" the result from the factor by the radical. Finally we show how Theorem 7.1 applies to the generalized conjugacy problem (Sec. 7.3).

7.1 Finding free submodules over semisimple algebras

In this section A is a semisimple algebra over the field K and V is an A-module. For every $v \in V$ we consider the module homomorphism $\phi_v : {}_AA \to V$ given as $\phi_v(x) = xv$. We define the rank $\operatorname{rk} v$ of $v \in V$ as the rank of the linear transformation $\phi_v : A \to V$.

Recall that the annihilator $\operatorname{Ann}_A(v)$ of an element $v \in V$ in A is a left ideal of A given as $\{x \in A | xv = 0\}$. This is the kernel of the A-module homomorphism $\phi_v : x \mapsto xv$, whence $Av \cong {}_AA/\operatorname{Ann}_A(v)$. We have $\operatorname{rk} v = \dim \operatorname{im} \phi_v = \dim Av = \dim A - \dim \operatorname{Ann}_A(v)$.

An element $v \in V$ is of maximal rank if $\operatorname{rk} w \leq \operatorname{rk} v$ for every $w \in V$. The following lemma suggests a method for testing whether v is of maximal rank.

Lemma 7.4. *Let V be a module over the semisimple K-algebra A. The element $v \in V$ is of maximal rank if and only if $Ann_A(v)V \subseteq Av$.*

Proof. Let $V_1, \dots, V_s$ be representatives of the isomorphism classes of the simple A-modules. Assume that ${}_AA \cong V_1^{\mu_1} \oplus \dots \oplus V_s^{\mu_s}$ and $V \cong V_1^{\nu_1} \oplus \dots \oplus V_s^{\nu_s}$. It is easy to see that the A-module U is cyclic if and only if $U \cong V_1^{\kappa_1} \oplus \dots \oplus V_s^{\kappa_s}$, where $\kappa_1 \leq \mu_1, \dots, \kappa_s \leq \mu_s$. It follows that v is of maximal rank in V if and only if the submodule Av is isomorphic to $V_1^{\mathrm{Min}\{\mu_1,\nu_1\}} \oplus \dots \oplus V_s^{\mathrm{Min}\{\mu_1,\nu_1\}}$.

Assume that v is not of maximal rank. Then there exists a simple A-module, say V_1, such that the multiplicity κ_1 of V_1 in AV is less than $\mathrm{Min}\{\mu_1, \nu_1\}$. Assume further that $\mathrm{Ann}_A(v)V \subseteq Av$, in other words, $\mathrm{Ann}_A(v)$ annihilates the factor module V/Av. The multiplicity of V_1 in that factor module is $\nu_1 - \kappa_1 > 0$, therefore $\mathrm{Ann}_A(v)$ annihilates the module V_1 as well. But $\mathrm{Ann}_A(V_1)$ is the ideal of A complementary to the ideal generated by the minimal left ideals isomorphic to V_1. This is a contradiction since the multiplicity of V_1 in $\mathrm{Ann}_A(v)$ is $\mu_1 - \kappa_1 > 0$. The "if" part is proved.

We give a proof of the "only if" part that will be useful in algorithms. Since A is semisimple there exists a left ideal L in A complementary to $\mathrm{Ann}_A(v)$: ${}_AA = L \oplus \mathrm{Ann}_A(v)$. Similarly, there exists a submodule V' of V complementary to Av. The map ϕ_v induces an isomorphism $L \cong Av$. Assume that we have bases of A and V, respectively, that reflect the decompositions described above. By this we mean that the basis of A is a union of bases of L and $\mathrm{Ann}_A(v)$, while the basis of V is a union of bases of Av and V'. For every $w \in V$ we consider the block structure of the matrix of ϕ_w. We see that the matrix of ϕ_v is a regular matrix on the block corresponding to $L \times Av$, and zero outside that block. Assume that there exists element $w \in V$ such that $\mathrm{Ann}_A(v)w$ is not a subset of Av. Decompose w as $w = cv + w'$, where $c \in A$ and $w' \in V'$. Since $\mathrm{Ann}_A(v)cv \subseteq Av$, $\mathrm{Ann}_A(v)w' \not\subseteq Av$. Observe that both blocks of the matrix of $\phi_{w'}$ corresponding to Av are zeros. It follows that the matrix of $\phi_{v+w'}$ is a block triangular matrix (both ϕ_v and $\phi_{w'}$ are zeros in the block corresponding to $\mathrm{Ann}_A(v) \times Av$), whence the sum of the ranks of the diagonal blocks is a lower bound for $\operatorname{rk}(v + w')$. In particular, since the lower right corner of $\phi_{v+w'}$ is nonzero, $\operatorname{rk}(v + w') > \operatorname{rk}(v)$. We have proved the lemma. □

The argument above also suggests a test of rank maximality as well as a method for incrementing the rank if it is possible. Indeed, let $v \in V$ and let $v_1, \dots, v_r$ be a basis of V. Obviously, $\mathrm{Ann}_A(v)V \leq Av$ if and only $\mathrm{Ann}_A(v)w \subseteq Av$ for every $w \in \{v_1, \dots, v_r\}$. We can compute the annihilator $\mathrm{Ann}_A(v)$ and test whether $\mathrm{Ann}_A(v)w \in Av$ for every $w \in \{v_1, \dots, v_r\}$ via solving systems of linear equations. This procedure terminates either with the conclusion that v is of maximal rank or with the first element $w \in \{v_1, \dots, v_r\}$ such that $\mathrm{Ann}_A(v)w \not\subseteq Av$. We can compute a projection $\pi \in \mathrm{End}_A(V)$ such that $\operatorname{im} \pi = Av$ and $\pi(v) = v$ via solving a system of linear equations. We take $w' = w - \pi(w)$. The argument of the proof of the lemma shows that $\operatorname{rk}(v + w') > \operatorname{rk}(v)$.

This method could serve as a basic step of iteration in a procedure for finding an element $v \in V$ of maximal rank. In fact, the procedure performs polynomially many field operations. Unfortunately, over infinite ground fields, we solve systems of linear equations that depend on the previous intermediate vector v, therefore we do not have any good control over the sizes of the vectors that occur during the iteration. Over sufficiently large fields we have the following generalization of [6], Lemma 5.2.

Lemma 7.5. *Let V be an r-dimensional module over the semisimple K-algebra A and $v_1, \ldots, v_r$ be a K-basis of V. Assume that $v \in V$ is an element of non-maximal rank. Let Ω be a subset of K^* consisting of at least $rk v + 1$ elements. Then there exists a scalar $\omega \in \Omega$ and a basis element $u \in \{v_1, \ldots, v_r\}$ such that $rk(v + \omega u) > rk Av$, i.e., $\dim_K A(v + \omega u) > \dim_K Av$.*

Proof. We use an argument similar to the proof of Lemma 5.2. in [6]. Let $w \in \{v_1, \ldots, v_r\}$ such that $\mathrm{Ann}_A(v)w \not\subseteq Av$. As in the proof of the preceding lemma, we consider decompositions ${}_AA = L \oplus \mathrm{Ann}_A(v)$ and $V = Av \oplus V'$ as well as the related block structure of matrices of ϕ_v and ϕ_w. Let $l = \mathrm{rk}\,(v)$. By choosing bases appropriately, we can achieve the situation where the matrix of ϕ_v is zero except the $l \times l$ principal minor, and the entry in position $(l+1, l+1)$ of the matrix of ϕ_w is nonzero. We also know that the $l \times l$ principal minor of ϕ_v has rank l. Let x be an indeterminate and $d(x)$ be the determinant of the $(l+1)\times(l+1)$ minor of the matrix of $\phi_{v+xw} = \phi_v + x\phi_w$. Obviously, $d(x) \in K[x]$ is of degree at most $l+1$. Expanding the determinant at the last row one sees that the coefficient of the linear term in $d(x)$ is the determinant of the $l \times l$ principal minor of ϕ_v. In particular, $d(x)$ is a nonzero polynomial of degree at most $l+1$. Since $\Omega \cup \{0\} > l+1$, there exists $\omega \in \Omega$ such that $d(\omega) \neq 0$. This implies that for such a scalar ω $\mathrm{rk}\,(v + \omega w) \leq l+1$. □

This lemma suggests another iterative method for finding an element $v \in V$ of maximal rank, provided that our ground field K is sufficiently large. Let $v_1, \ldots, v_r$ be a basis of V and Ω be a subset of K^* of cardinality r. Initially we take $v = 0$. In each round, we compute the ranks $\mathrm{rk}\,(v + \omega w)$, $(w \in \{v_1, \ldots, v_r\}, \omega \in \Omega)$. We replace v with the first element $v + \omega w$ such that $\mathrm{rk}\,(v + \omega w) > \mathrm{rk}\,(v)$. We stop if no such element exists. The procedure terminates in at most r iterations and the intermediate element v after t rounds is in the form $\omega_1 w_1 + \ldots + \omega_t w_t$, where $\omega_i \in \Omega$ and $w_i \in \{v_1, \ldots, v_n\}$. If K is an algebraic number field, we take $\Omega = \{1, \ldots, r\}$. This gives a polynomial bound on the size of the vectors we compute with. We have proved the following.

Theorem 7.6. *Let V be a module over the semisimple K-algebra A, where K is a finite field or an algebraic number field. There is a deterministic polynomial time algorithm that finds an element $v \in V$ of maximal rank.* □

We also have a straightforward generalization of the procedure *findfree* of the paper [6].

Theorem 7.7. *Let V be a module over the semisimple K-algebra A, where K is a finite field or an algebraic number field. There is a deterministic polynomial time algorithm that finds (free generators of) a maximal free submodule of V.* □

7.2 Finding a single generator

In this section we turn to the general case where V is a module over the (not necessarily semisimple) K-algebra A and prove Theorem 7.1.

Note that we assumed an efficient method for computing the the radical $\mathrm{Rad}(A)$. Using $\mathrm{Rad}(A)$, we can compute $\mathrm{Rad}(V) = \mathrm{Rad}(A)V$. We consider the action of the factor-algebra $\overline{A} = A/\mathrm{Rad}(A)$ on $\overline{V} = V/\mathrm{Rad}(V)$. Let $v \in V$ be an arbitrary vector and $\overline{v} = v + \mathrm{Rad}(V)$. It is obvious that $Av = V$ implies $\overline{A}\overline{v} = \overline{V}$. We claim that converse also holds. The proof relies on the well known fact that elements of $\mathrm{Rad}(V)$ can be omitted from any system of A-module generators. Assume that $\overline{v}$ is a generator of the $\overline{A}$-module $\overline{V}$. This means that

$Av + \mathrm{Rad}(V) = V$. Assume that Av is a proper submodule of V. Let M be a maximal proper submodule of V containing Av. Since $\mathrm{Rad}(V)$ is the intersection of the maximal proper submodules, we have $AV + \mathrm{Rad}(V) \leq M < V$, a contradiction. We have proved the claim.

$\overline{V}$ is a unital module over the semisimple algebra $\overline{A}$. Using the method of Theorem 7.6 we compute an element $\overline{v} \in \overline{V}$ of maximal rank. If $\mathrm{rk}\,\overline{v} < \dim_K \overline{V}$, i.e, $\overline{v}$ is not a generator then neither V nor $\overline{V}$ is cyclic. On the other hand, if $\overline{v}$ is a generator of $\overline{V}$ then we can return any element $v \in \overline{v}$ as a generator of V. This finishes the proof of Theorem 7.1. □

7.3 The general conjugacy problem

This section is devoted to the proof of Theorem 7.2.

We consider the linear subspace V of A given as

$$V = \{v \in A \,|\, va_i = a'_i v \ (i = 1, \dots, m)\}.$$

The task is equivalent to finding a unit in V. Let A' be the centralizer of the elements $a'_1, \dots, a'_m$:

$$A' = \{x \in A \,|\, xa'_i = a'_i x \ (i = 1, \dots, m)\}.$$

A' is a subalgebra of A containing 1_A and V is closed under multiplication by elements from A' from the left, i.e., V is a left A'-module. Let v be an arbitrary element from V. We use the linear map $\phi_v : A' \to V$ mapping x to xv. We claim that if $A^* \cap V \neq \emptyset$ then V is a cyclic A'-module and every generator v of V is a unit in A. Indeed, let $y \in A^* \cap V$. Then the map ϕ_y is a A' module isomorphism between A' and V: the inverse of ϕ_y is the map $w \mapsto wy^{-1}$. In particular, V is cyclic. Let x be an arbitrary generator. Then xy^{-1} is a generator of ${}_{A'}A'$, therefore xy^{-1} is a unit in A', whence $xy^{-1} \in A^*$, and $x \in A^*$. The claim is proved.

We compute V and A' as the solution spaces of systems of linear equations. We attempt to find a generator of V by the method of Theorem 7.1. If V is not cyclic then the conjugacy problem admits no solution. If V is cyclic then the method of Theorem 7.1 also returns a generator x of V. Again, if x is not a unit then there exist no units in V at all. Otherwise we can return x. This finishes the proof of Theorem 7.2.

7.4 Remarks

The weakness of the algorithms presented in this chapter is that they depend on the ability of computing the radical of algebras over the ground field in polynomial time. Recently P. Brooksbank and E. M. Luks developed a deterministic polynomial method for testing and finding isomorphisms of modules unconditionally [16].

Chapter 8

Deciding universality of quantum gates

In this chapter, based on the paper [56], we show that universality of quantum gate sets is decidable. We say that collection of n-qudit gates is universal if there exists $N_0 \geq n$ such that for every $N \geq N_0$ every N-qudit unitary operation can be approximated with arbitrary precision by a circuit built from gates of the collection. Our main result is an upper bound on the smallest N_0 with the above property. The bound is roughly $d^8 n$, where d is the number of levels of the base system (the 'd' in the term qu*d*it.) The proof is based on a recent result of R. Guralnick and P. H. Tiep on invariants of (finite) linear groups.

A qu*d*it is a vector of norm 1 from the Hilbert space $\mathbb{C}^d$, an n-qu*d*it state is an element of norm 1 of $(\mathbb{C}^d)^{\otimes n} \cong \mathbb{C}^{d^n}$. The space $(\mathbb{C}^d)^{\otimes n}$ is called an n-qu*d*it quantum system and the the factors of the n-fold tensor product $(\mathbb{C}^d)^{\otimes n}$ are referred as the qu*d*its of the system. An n-qu*d*it quantum operation (or gate) is a unitary transformation acting on the n-qu*d*it states, i.e., an element of the unitary group U_{d^n}. As in quantum computation, states which are scalar multiples of each other are considered equivalent, quantum operations are also understood projectively. In particular, for every $u \in U_{d^n}$, the normalized operation $(\det u)^{-1/d^n} \cdot u$ represents the same gate as u (here α^{1/d^n} stands for any d^nth root of a complex number α).

Let $\Gamma \subset U_{d^n}$ be a (finite) collection of n-qu*d*it quantum gates. We say that Γ is a *complete* set of n-qu*d*it gates if a scalar multiple of every n-qu*d*it operation from U_{d^n}, can be approximated with an arbitrary precision by a product of operations from Γ. In other words, Γ is complete if the semigroup of U_{d^n} generated by Γ and the unitary scalar matrices is dense in U_{d^n}. The latter condition, because of compactness, is equivalent to saying that the *group* generated by Γ and the unitary scalar matrices is dense in U_{d^n}.

Note that in the quantum computation literature complete sets of gates are frequently called universal. In this Chapter, partly following the terminology of [64], we reserve the term *universal* for a weaker notion discussed below.

For $N \geq n$ we can view $(\mathbb{C}^d)^{\otimes N}$ as a bipartite system $(\mathbb{C}^d)^{\otimes n} \otimes (\mathbb{C}^d)^{\otimes N-n}$ and let an n-qu*d*it gate u act on the first part only. Formally, the N-qu*d*it extension u_N of u is the operation $u \otimes I$ where I stands for the identity of $(\mathbb{C}^d)^{\otimes N-n}$. For an n-qu*d*it gate set Γ the gate set Γ_N is the collection of the extensions of gates from Γ obtained this way: $\Gamma_N = \{u_N | u \in \Gamma\}$.

More generally, we can extend an n-qu*d*it gate u to N qu*d*its by selecting an embedding μ of $\{1, \dots, n\}$ into $\{1, \dots, N\}$ and let act u on the components indexed by $\mu(1), \dots, \mu(n)$

(in this order) and leave the rest "unchanged". It will be convenient to formalize this in terms of permutations of the qu*d*its of the larger system as follows. Each permutation from the symmetric group S_N acts on $(\mathbb{C}^d)^{\otimes N}$ by permuting the tensor components. For an N-qu*d*it gate v and $\sigma \in S_N$ the operation $v^\sigma = \sigma v \sigma^{-1}$ is also a quantum gate which can be considered as the gate v with "fans" permuted by σ. We denote by Γ^N the collection of gates obtained from gates in Γ_N this way: $\Gamma^N = \{u_N^\sigma | u \in \Gamma, \sigma \in S_N\}$.

We say that for $N \geq n$ the n-qu*d*it gate set Γ is *N-universal* if Γ^N is complete. The collection Γ is called ∞-universal or just universal, for short, if there exists $N_0 \geq n$ such that Γ is N-universal for every $N \geq N_0$. It turns out that for $n \geq 2$, every complete n-qu*d*it gate is N-universal for every $N \geq n$. This claim follows from the fact that the Lie algebra su_{d^N} is generated by $su_{d^2}^N = \{(u \otimes I)^\sigma | u \in su_{d^2}, \sigma \in S_N\}$. This is shown in [26] for $d = 2$ but essentially the same proof works for $d > 2$ as well. By the claim, N-universality of a fixed gate set for $N \geq 2$ is a monotone property in N: for $n \geq 12$, an n-qu*d*it gate set Γ is universal if and only if there exists an integer $N \geq n$ such that Γ is N-universal. On the other hand, no 1-qu*d*it gate set can be universal as the resulting group preserves the natural tensor decomposition.

Completeness of a gate set can be decided by computing the (real) Zariski closure of the group generated by the gates using the method of H. Derksen, E. Jeandel and P. Koiran [23]. A polynomial time algorithm for gates defined over a number field is given in [64, 65]. Reducing the problem of universality to completeness requires a bound for the smallest N such that a universal set of gates is N-universal. In [64, 65] Jeandel gives a 6-qubit gate set which is 9-universal but not 6-universal and it is explained how to extend this example to a gate set over $2^k + 2$ qubits which is $2^{k+1} + 1$-universal but not $2^{k+1} - 2$-universal where k is an integer greater than 1. (A qubit is a qu*d*it with $d = 2$.) Our main result is the following.

Theorem 8.1. *Let Γ be an n-qudit gate set where $n, d \geq 2$. Then Γ is universal if and only if it is N-universal for some integer $N \leq d^8(n-1) + 1$.*

Our main technical tool, a criterion for completeness based on invariants of groups, is given in Section 8.1. It can be considered as a "more algebraic" variant of Jeandel's criterion given in [64, 65]. Correctness is a consequence of a recent result of R. Guralnick and P. H. Tiep stating that certain low degree invariants distinguish the special linear group from its closed (in particular finite) subgroups. Needless to say, the proof of the applied result heavily uses the classification of finite simple groups and their representations.

We prove Theorem 8.1 in Section 8.2. The outline of the proof is the following. We relate polynomial ideals to gate sets. The completeness criterion gives that the Hilbert polynomial of the ideal corresponding to a universal gate set must be the constant polynomial 24. Our result is then a consequence of D. Lazard's bound on the regularity of Hilbert functions of zero dimensional ideals.

8.1 Completeness

In Jeandel's work [64, 65], testing gate sets for completeness is based on the following observation.

Fact 8.2. *Let $d \geq 2$ and let G be subgroup of SU_{d^N}. Assume further the real vector space su_{d^N} (the Lie algebra of SU_{d^N}) consisting of the traceless skew Hermitian $d^N \times d^N$ matrices* *is an irreducible $\mathbb{R}G$-module under the conjugation action by elements of G. Then G is either finite or dense in SU_{d^N}.*

Therefore if Γ is a finite collection of normalized gates then testing Γ for completeness amounts to testing irreducibility of su_{d^N} under conjugation of elements of Γ and to testing if the linear group generated by Γ is finite. The first test can be accomplished by solving a system of linear equations (see below) while for the other – in the case where the gates are defined over an algebraic number field – the method [4] of L. Babai, R. Beals and D. Rockmore is available. Here, informally, we are going to replace the second test with a test similar to the first one.

Set $V = \mathbb{C}^{d^N}$, the complex column vectors of length d^N. The vector space V is a left $\mathbb{C}G$-module for every linear group $G \leq GL_{d^N}(\mathbb{C})$. The dual space $V^* = \mathrm{Hom}_{\mathbb{C}}(V, \mathbb{C})$ is a right $\mathbb{C}G$-module. It can be made a left $\mathbb{C}G$ module by letting u^{-1} act in place of u. This module (denoted also by V^*) is called the module contragradient to V. In terms of matrices, the contragradient matrix representation can be obtained by taking the inverse of the transpose of the original matrix representation. Note that for $u \in U_{d^N}$ the matrix of u in the contragradient representation will be simply the complex conjugate of the matrix of u.

For every positive integer k and $G \leq GL_{d^N}(\mathbb{C})$ we define the quantity $\mathrm{M}_{2k}(G)$ as

$$\mathrm{M}_{2k}(G) = \dim_{\mathbb{C}} \mathrm{Hom}_{\mathbb{C}G}((V \otimes V^*)^{\otimes k}, \mathbb{C}).$$

Recall that for a left $\mathbb{C}G$-module W

$$\mathrm{Hom}_{\mathbb{C}G}(W, \mathbb{C}) = \{f \in W^* | f(gw) = f(w) \text{ for every } g \in G, w \in W\}.$$

Note that if a finite set Γ generates a dense subgroup of G and B is a basis of W then

$$\mathrm{Hom}_{\mathbb{C}G}(W, \mathbb{C}) = \{f \in W^* | f(gw) = f(w) \text{ for every } g \in \Gamma, w \in B\}, \tag{8.1}$$

and hence (a basis of) the space $\mathrm{Hom}_G(W, \mathbb{C})$ can be computed by solving a system of linear equations.

Also note that $V \otimes V^* \cong \mathrm{End}_{\mathbb{C}}(V)$ and $\mathrm{M}_2(G)$ is the dimension of the centralizer of G (in $\mathrm{End}_{\mathbb{C}}(V)$). In particular, $\mathrm{M}_2(G) = 1$ if and only if V is an irreducible $\mathbb{C}G$-module. Similarly, $\mathrm{M}_4(G)$ is the dimension of the centralizer of the conjugation action of G on $d^N \times d^N$ complex matrices.

M. Larsen observed that if $\mathcal{G}$ is the entire complex linear group $GL_{d^N}(\mathbb{C})$, or the complex orthogonal group or the complex symplectic group and G is a Zariski closed subgroup of $\mathcal{G}$ such that the connected component of the identity in G is reductive (including the case when this component is trivial) and $\mathrm{M}_4(G) = \mathrm{M}_4(\mathcal{G})$ then either G is finite or $G \geq [\mathcal{G}, \mathcal{G}]$. (Notice that Fact 8.2 can be viewed as the unitary analogue of Larsen's alternative.) Larsen also conjectured that for a finite subgroup $G < \mathcal{G}$ we have $\mathrm{M}_{2k}(G) > \mathrm{M}_{2k}(\mathcal{G})$ with some $k \leq 4$. Recently R. M. Guralnick and P. H. Tiep [48], using the classification of finite simple groups and their irreducible representations, settled Larsen's conjecture. The conjecture holds basically true, there are only two exceptions. In any case, $\mathrm{M}_{2k}(G) > \mathrm{M}_{2k}(\mathcal{G})$ with some $k \leq 6$. The following statement is an easy consequence of the results from [48]. In order to shorten notation, for a collection $\Gamma \subseteq U_{d^N}$ we define $\mathrm{M}_{2k}(\Gamma)$ as $\mathrm{M}_{2k}(G)$ where G is the smallest closed subgroup of U_{d^N} containing Γ (in the norm topology). Also, in view (8.1) and the comment following it, computing $\mathrm{M}_{2k}(\Gamma)$ can be accomplished by computing the rank of a $d^{N^{2k}}$ by $|\Gamma| d^{N^{2k}}$ matrix if Γ is finite.

Proposition 8.3. *Assume that $d^N > 2$ and let $\Gamma \subset U_{d^N}$. Then Γ is complete if and only if $M_8(\Gamma) = M_8(GL_{d^N}(\mathbb{C}))$. If $d^N = 2$ then the necessary and sufficient condition for completeness is $M_{12}(\Gamma) = M_{12}(GL_{d^N}(\mathbb{C}))$.*

Proof. We only prove the first statement, the second assertion can be verified with a slight modification of the arguments. Let G be the smallest closed subgroup of U_{d^N} containing Γ (in the norm topology). We replace each $u \in G$ with its normalized version $\det^{-1} u \cdot u$. In this way we achieve that G is a closed subgroup of SU_{d^N}. As the action of $\det^{-1} u \cdot u$ is the same as that of u on $V^{\otimes k} \otimes V^{*\otimes k}$, this change does not affect the quantities $\mathrm{M}_{2k}(G)$. If Γ is complete then $G = SU_{d^N}$. Therefore the Zariski closure of G in $GL_{d^N}(\mathbb{C})$ (over the complex numbers) is $SL_{d^N}(\mathbb{C})$ and hence $\mathrm{M}_{2k}(G) = \mathrm{M}_{2k}(SL_{d^N}(\mathbb{C})) = \mathrm{M}_{2k}(GL_{d^N}(\mathbb{C}))$ for every k. This shows the "only if" part.

To prove the reverse implication, assume that $\mathrm{M}_8(G) = \mathrm{M}_8(GL_{d^N}(\mathbb{C}))$. By Lemma 3.1 of [48], $\mathrm{M}_{2k}(G) = \mathrm{M}_{2k}(GL_{d^N}(\mathbb{C}))$ for $k = 1, 2, 3$ as well. In particular, $\mathrm{M}_4(G) = \mathrm{M}_4(GL_{d^N}(\mathbb{C})) = 2$. Notice that G is a compact Lie group therefore every finite dimensional representation of G is completely reducible. Hence the the conjugation action of G on $d^N \times d^N$ matrices has two irreducible components: one consists of the scalar matrices the other one is the Lie algebra $sl_{d^N}(\mathbb{C})$ of traceless matrices. As a real vector space, $sl_{d^N}(\mathbb{C})$ is the direct sum of su_{d^N} and $i \cdot su_{d^N}$ (here $i = \sqrt{-1}$). Both subspaces are invariant under the action of U_{d^N}, therefore they are $\mathbb{R}G$-submodules and multiplication by i gives an $\mathbb{R}G$-module isomorphism between them. It follows that su_{d^N} must be an irreducible $\mathbb{R}G$-module. Hence by Fact 8.2, either $G = SU_{d^N}$ or G is finite. In the first case Γ is complete. In the second case we can apply the results of [48]. By Theorems 1.4 and 2.12 therein, G must be $SL_2(5)$ and $d^N = 2$. This contradicts the assumption $d^N > 2$. □

8.2 Universality

We begin with a lemma which establishes a condition for N-universality which suits better our purposes than the original definition.

Lemma 8.4. *Let $d > 1$ and Γ be an n-qudit gate set, let $N \geq n$ and let Σ be an arbitrary generating set for S_N. Then Γ is N-universal if and only if $\Gamma_N \cup \Sigma$ is complete.*

Proof. Let H resp. G denote the closure of the subgroup of SU_{d^N} generated by the normalized gates from Γ^N and $\Gamma_N \cup \Sigma$, respectively. As Γ^N is in the subgroup generated by $\Gamma \cup \Sigma$, the group H is a subgroup of G and hence the "only if" part of the statement is obvious. On the other hand, H is closed under conjugation by the elements of $\Gamma \cup S_N$, therefore H is a closed normal subgroup of G. Assume that $\Gamma_N \cup \Sigma$ is complete, i.e., $G = U_{d^N}$. Then Γ_N must contain at least one non-scalar matrix since otherwise G would be finite (every matrix in G would be a permutation, multiplied by a d^Nth root of unity). Therefore H is a normal subgroup of SU_{d^N} containing a non-scalar matrix. Because of simplicity of PSU_{d^N} this implies $H = SU_{d^N}$, that is, Γ^N is complete. □

By Lemma 8.4, we can consider gate sets on N qu*d*its which consist of two parts. The gates in the first part act on the first n qu*d*its while the rest consists of permutations. We exploit this property in Subsection 8.2.1, where we relate polynomial ideals to such a sequence of gate sets where N varies. We finish the proof of Theorem 8.1 in Subsection 8.2.2 by observing that the sequence M_8 for letting an n-qu*d*it gate set together with the symmetric group S_N act on $(C^d)^{\otimes N}$ $(N = n, n+1, \ldots)$ take the same values as the Hilbert function of the corresponding ideal.

8.2.1 The ideal of a gate set

In this subsection $W = \mathbb{C}^m$ for some integer $m > 0$ and G is a subgroup of $GL(W^{\otimes n})$. For every $N \geq n$ we establish a relation between $\mathrm{Hom}_{\langle G,S_n\rangle}(W^{\otimes n},\mathbb{C})$ and $\mathrm{Hom}_{\langle G\otimes I,S_N\rangle}(W^{\otimes N},\mathbb{C})$. Here S_N denotes the subgroup of $GL(W^{\otimes N})$ consisting of the permutations of tensor components and I stands for the identity on $W^{\otimes(N-n)}$.

We work with the tensor algebra $T = \oplus_{j=0}^{\infty} W^{\otimes j}$ of W. We use some elementary properties of T and its substructures. Most of the proofs can be found in Section 9 of [47]. We say that an element w of T is homogeneous of degree j if $w \in W^{\otimes j}$. If we fix a basis $w_1,\dots,w_m$ of W, then a basis of T consists of the non-commutative monomials of the form $w_{i_1}\otimes\cdots\otimes w_{i_j}$ and T can be interpreted as the ring of non-commutative polynomials in $w_1,\dots,w_m$ over $\mathbb{C}$. In this interpretation, for every $j \geq 0$ the elements of $W^{\otimes j}$ are identified with the homogeneous non-commutative polynomials of degree j. A right (or two sided) ideal J of T is called graded if J equals the sum $\oplus_{j=0}^{\infty} J^j$ where $J^j = W^{\otimes j}\cap J$. The component J^j is called the degree j part of J. It turns out that a right (resp. two-sided) ideal J of T is graded if and only if there is a set of homogeneous elements of J which generate J as a right (resp. two-sided) ideal.

Let M be the two-sided ideal of T generated by $w_i\otimes w_j - w_j\otimes w_i$ ($i,j \in \{1,\dots m\}$), and let $\phi : T \to R = T/M$ be the natural map. Then M is a graded ideal with degree j parts M^j which are spanned by $w_{i_1}\otimes\cdots\otimes w_{i_j} - w_{i_{\sigma(1)}}\otimes\cdots\otimes w_{i_{\sigma(j)}}$ where $(i_1,\dots i_j) \in \{1\dots,m\}^j$ and $\sigma \in S_j$. The factor algebra R is called the symmetric algebra of W. Set $x_i = \phi(w_i)$ for $i = 1,\dots,m$. Then R is identified with the (commutative) polynomial ring $\mathbb{C}[x_1,\dots,x_m]$. The image of R^j of $W^{\otimes j}$ under ϕ is the jth symmetric power of W. In interpretation of R as polynomial ring, R^j consists of the homogeneous polynomials of degree j.

For a subspace L of $(W^{\otimes N})^*$ we denote by $L^\perp$ the subspace of $W^{\otimes N}$ annihilated by L: $L^\perp = \{w \in W^{\otimes N} | l(w) = 0 \text{ for every } l \in L\}$. Because of duality, $\dim L = \dim(W^{\otimes N}/L^\perp)$ and $(L_1\cap L_2)^\perp = L_1^\perp + L_2^\perp$. In particular, $\mathrm{Hom}_{\langle G\otimes I\cup S_N\rangle}(W^{\otimes N},\mathbb{C})^\perp = \mathrm{Hom}_{G\otimes I}(W^{\otimes N},\mathbb{C})^\perp + \mathrm{Hom}_{S_N}(W^{\otimes N},\mathbb{C})^\perp$.

As $\mathrm{Hom}_{G\otimes I}(W^{\otimes N},\mathbb{C}) = \mathrm{Hom}_G(W^{\otimes n},\mathbb{C})\otimes(W^{\otimes(N-n)})^*$, we obtain that

$$\mathrm{Hom}_{G\otimes I}(W^{\otimes N},\mathbb{C})^\perp = \mathrm{Hom}_G(W^{\otimes n},\mathbb{C})^\perp \otimes W^{\otimes(N-n)},$$

in other words, the space $\mathrm{Hom}_{G\otimes I}(W^{\otimes N},\mathbb{C})^\perp$ is the degree N part of the right ideal $H(G)$ in T generated by $\mathrm{Hom}_G(W^{\otimes n},\mathbb{C})^\perp$.

The space $\mathrm{Hom}_{S_N}(W^{\otimes N},\mathbb{C})$ corresponds the symmetric N-linear functions, i.e., it consists of the linear functions $W^{\otimes N} \to \mathbb{C}$ which take identical values on $w_{i_1}\otimes\cdots\otimes w_{i_N}$ and $w_{i_{\sigma(1)}}\otimes\cdots\otimes w_{i_{\sigma(N)}}$ for every permutation $\sigma \in S_N$. Therefore $\mathrm{Hom}_{S_N}(W^{\otimes N},\mathbb{C})^\perp$ coincides with the degree N part M^N of the ideal M.

We obtain that $\mathrm{Hom}_{\langle G\otimes I\cup S_N\rangle}(W^{\otimes N},\mathbb{C})^\perp$ is the degree N part of $H(G)+M$. As $H(G)$ is a right ideal and M is an ideal in T with $R = T/M$ commutative, $H(G)+M$ is an ideal in T containing M. Setting $J(G) = \phi(H(G)+M)$ we conclude that for every $N \geq n$, $J^N(G) = \phi(\mathrm{Hom}_{\langle G\otimes I\cup S_N\rangle}(W^{\otimes N},\mathbb{C})^\perp)$ is the degree N part of $J(G)$. Furthermore, $J(G)$ is the ideal of the commutative polynomial ring R generated by $J^n(G)$ and

$$\dim \mathrm{Hom}_{\langle G\otimes I\cup S_N\rangle}(W^{\otimes N},\mathbb{C}) = \dim(R^N/J^N(G)).$$

8.2.2 The proof of Theorem 8.1

Let $n,d \geq 2$, let $\Gamma \subseteq GL((\mathbb{C}^d)^{\otimes n})$ and let G be the subgroup of $GL(\mathbb{C}^d)$ generated by Γ. For every integer $N \geq n$, we consider the G-module $V = (\mathbb{C}^d)^{\otimes N}$ where the action of G is given by $G \otimes I$ (here I is the identity on $V^{\otimes(N-n)}$). We set $W = (\mathbb{C}^d)^{\otimes 4} \otimes ((\mathbb{C}^d)^*)^{\otimes 4}$ and, with some abuse of notation, consider G as a subgroup of $GL(W^{\otimes n})$. For every $N \geq n$ we have the G-module isomorphism $V^{\otimes 4} \otimes (V^*)^{\otimes 4} \cong W^{\otimes N}$ where the action of G on the right hand side is $G \otimes I$ (this time I is the identity on $W^{\otimes(N-n)}$). Applying the notation and observations of the preceding subsection in this context, we obtain that

$$\mathrm{M}_8(\langle G \otimes I \cup S_N \rangle) = \dim(R^N/J^N(G))$$

for every $N \geq n$.

First we consider the full linear group $GL_{d^n}(\mathbb{C})$. The n-universality of U_{d^n} for $n \geq 2$ gives $\dim(R^N/J^N(GL_{d^n}(\mathbb{C})) = \mathrm{M}_8(GL_{d^N}(\mathbb{C}))$. From invariant theory it is known that $\mathrm{M}_8(GL_{d^N}(\mathbb{C})) = 4! = 24$, see [93].

Now consider an arbitrary gate set $\Gamma \subseteq U_{d^n}$ and let $G \leq GL_{d^n}(\mathbb{C})$ the group generated by Γ. The preceding discussion and Proposition 8.3 give that Γ is universal if and only if $\dim(R^N/J^N(G)) = 24$ for sufficiently large degree N.

The ideal $J(G)$ is an ideal of $R = \mathbb{C}[x_1, \ldots, x_m]$ generated by homogeneous polynomials of degree n. In the context of polynomial rings, graded ideals are called homogeneous. That is, an ideal J of the polynomial ring R is called homogeneous if J is the direct sum its homogeneous components $J^j = R^j \cap J$; and an ideal generated by homogeneous polynomials is homogeneous. The *Hilbert function* of the homogeneous ideal J is given as $j \mapsto h_J(j) = \dim R^j/J^j$. It turns out that the Hilbert function is ultimately a polynomial: there is a polynomial p_J (in one variable) and an integer N such that $h_J(j) = p_J(j)$ for $j \geq N$. The smallest N with this property is called the regularity of the Hilbert function of J. The degree of the Hilbert polynomial is the *dimension* of J. (Actually, it is the dimension of the projective variety consisting of the common projective roots of the polynomials in J.)

The discussion above shows that the Hilbert polynomial of the ideal $J(G)$ corresponding to a universal gate set is the constant 24. In particular, the dimension of $J(G)$ is zero. In [73], D. Lazard proved that the regularity of the Hilbert function of a zero dimensional ideal in $\mathbb{C}[x_1, \ldots, x_m]$ generated by homogeneous polynomials of degree n is bounded by $mn - m + 1$. From this, the proof of Theorem 8.1 is finished by observing that the smallest N for which Γ is N-universal coincides with the regularity of the Hilbert function of $J(G)$.

8.3 Remarks

Very likely the bound proved in Theorem 8.1 is not tight. However, for fixed d it is linear in n and Jeandel's construction discussed at the beginning of this chapter shows that in fact the smallest N such that a universal n-qubit gate set is N-universal can be at least $2n - 6$. Proving better upper bounds would require deeper knowledge of subspaces of $V^{*\otimes 4} \otimes V^{\otimes 4}$ which occur as $\mathrm{Hom}_G(V^{\otimes 4} \otimes V^{*\otimes 4}, \mathbb{C})$ for $G \leq GL(V)$. Using the isomorphism $\mathrm{Hom}_G(V^{\otimes 4} \otimes V^{*\otimes 4}, \mathbb{C}) \cong \mathrm{End}_G(V^{\otimes 4})$, a natural restriction is that these subspaces must be subalgebras of $\mathrm{End}_{\mathbb{C}}(V^{\otimes 4})$. However, it is not clear to us how to exploit this fact.

Effectiveness and complexity of algorithms for testing completeness and universality based on Proposition 8.3, Theorem 8.1 and Lemma 8.4 depend on the computational model and on the way the input gate set is represented. In particular, in the Blum–Shub–Smale model for the real numbers (this is based on black boxes performing field operations and inequality tests), if the input gates are given as arrays of $n \times n$ complex numbers, the completeness test can be accomplished in polynomial time. With the same assumption on the input, for constant d (e.g., for qubits or qutrits) even universality can be tested in polynomial time. Similar results can be stated for Boolean complexity if the entries of the matrices representing the input gates are from an algebraic number field. Even the problem whether there is a non-universal gate set which is ϵ-close to a given collection of gates in the Hadamard norm of matrices is decidable. Indeed, existence is equivalent to solvability of a (huge) system of polynomial equations and inequalities over the real numbers. Of course, this straightforward method is far from practical.

Chapter 9

A quantum algorithm for finding hidden subgroups in a class of solvable groups

This chapter is based on parts of the paper [36], joint work with Katalin Friedl, Frédéric Magniez, Miklos Santha, and Pranab Sen. Here we give a quantum algorithm for solving the hidden subgroup problem (HSP) in polynomial time for a class of solvable groups. Our approach is based on considering permutation problems closely related to the hidden subgroup problem.

Efficient solutions to some cases of the hidden subgroup problem (see Subsection 2.5.6) constitute probably the most notable success of quantum algorithms. To be efficient, an algorithm has to be polynomial in the length of strings encoding group elements, which is usually logarithmic in the order of G. While classically not even query efficient algorithms exist for the HSP, it can be solved efficiently in abelian groups by a quantum algorithm. A detailed description of the so called standard algorithm can be found for example in [78] or in [10]. The main quantum tool of this algorithm is Fourier sampling, based on the efficiently implementable Fourier transform in abelian groups, see Subsection 2.5.7. Factorization and discrete logarithm [90] are special cases of this solution.

After settling the case of finite abelian groups, substantial research was devoted to the hidden subgroup problem in finite noncommutative groups. The interest in this problem is enhanced by the fact, that the graph isomorphism is a special case. The standard algorithm has been extended to some special cases in non-abelian groups, including finding hidden normal subgroups in groups admitting efficient quantum procedure for the so-called noncommutative Fourier transform, see [86, 49, 46, 77]. In this chapter we present a method for finding non-normal hidden subgroups in a class of solvable groups.

We assume that G is a finite solvable group of constant derived length given by a refined polycyclic presentation, see Section 2.3. We use normal words for encoding elements of G and suppose that this encoding requires $\ell = O(\log |G|)$ bits.

The main advantage of using such a presentation is that it allows fast computation of a unique encoding of subgroups of G given by generators. This unique encoding allows us to work with ”clean” subroutines in the sense of Subsection 2.5.2.

By a *quantum permutation action* of G we mean a permutation action of G on a set Ψ, where Ψ consists of pairwise orthogonal unit vectors (states) from $\mathbb{C}^{2^t}$ for some natural number t. We use the left multiplicative notation $x\underline{\psi}$ for permutation actions. That is, a permutation action of G on Ψ is a map $(x, \underline{\psi}) \mapsto x\underline{\psi}$ from $G \times \underline{\psi}$ onto Ψ satisfying $(xy)\underline{\psi} = x(y\underline{\psi})$. The adjective "quantum" expresses that Ψ does not need to be a subset of the computational basis. If Ψ is a subset of the computational basis then we refer to the action of G on Ψ as a classical permutation action. We assume that the action is given by an efficient quantum procedure, more precisely, a by quantum circuit of size polynomial in ℓt mapping $|x\rangle \otimes \underline{\psi}$ to $|x\rangle \otimes x\underline{\psi}$ where $x \in G$ and $\underline{\psi} \in \Psi$. (In the spirit of Subsection 2.5.2, we actually allow that the procedure uses some workspace which contains both initially and finally zero qubits.) With some abuse of notation, we will denote merely by Ψ the permutation action of G on Ψ. This will not cause any confusion as in this chapter we do not consider different permutation actions on identical sets.

We define two computational problems related to the hidden subgroup problem. The input for the *quantum stabilizer* problem $\mathrm{STABILIZER}_K(G, \Psi, \underline{\psi})$ is the tensor power $\underline{\psi} \otimes \cdots \otimes \underline{\psi}$ of length K, that is, it consists of K identical copies (so called clones) of the vector $\underline{\psi} \in \Psi$. Taking multiple copies of the input vector $\underline{\psi}$ is necessary because there is no general method for producing clones of arbitrary quantum states. The task is to compute (generators for) the stabilizer of $\underline{\psi}$. Our main result is the following.

Theorem 9.1. *Assume that the finite solvable G has constant derived length and the derived series of G' is such that the factors of the consecutive members are of exponent bounded by a constant. Suppose that G is given by a polycyclic presentation where normal words for group elements are encoded by bit strings of length ℓ. Suppose further that we have a quantum permutation action of G on the orthonormal set $\Psi \subseteq \mathbb{C}^{2^t}$. Then, with $K = (\log|G|)^{\theta(1)} \log \frac{1}{\epsilon}$, the problem $\mathrm{STABILIZER}_K(G, \Psi, \underline{\psi})$ for $\underline{\psi} \in \Psi$ can be solved by a quantum algorithm in time $(\log|G|\ell t)^{O(1)} \log \frac{1}{\epsilon}$ with error at most ϵ.*

The input for the *constructive orbit membership* problem (or just orbit membership for short) $\mathrm{ORBIT\text{-}MEMBER}_K(G, \Psi, \underline{\psi}_0, \underline{\psi}_1)$ is the tensor product $\underline{\psi}_0 \otimes \cdots \otimes \underline{\psi}_0 \otimes \underline{\psi}_1 \otimes \cdots \otimes \underline{\psi}_1$ of length $2K$ for $\underline{\psi}_0, \underline{\psi}_1 \in \Psi$. Intuitively, the input consists of K copies of a pair of vectors from Ψ. The task is to decide if $\underline{\psi}_1$ is in the orbit of $\underline{\psi}_0$ and to find the set of element $x \in G$ carrying $\underline{\psi}_0$ to $\underline{\psi}_1$ if they are in the same orbit. Thus the output is either "none" or a coset of the stabilizer of ω_0. In this chapter we only need to solve instances of $\mathrm{ORBIT\text{-}MEMBER}_K(G, \omega_0, \cdots)$ where G is an abelian group and the stabilizer is trivial. Then the solution is "none" or a single element of G.

Note that, while the inputs for the stabilizer and constructive orbit membership problems are allowed to be "quantum", the outputs are assumed to be "classical", i.e., the computational basis elements corresponding to the strings describing the output. Therefore we can apply the cleanup method of Subsection 2.5.2 after performing algorithms for these tasks: we make a separate copy of the output and then undo (perform the inverse of) the algorithm.

Also note that if the input is "classical", that is, an element of the computational basis for the input part then actually from one copy of the input we can produce arbitrary many copies. In view of this, the quantum stabilizer problem (over the symmetric group rather than a solvable one) includes computing automorphism of graphs and the constructive orbit membership problem includes the constructive version of the graph isomorphism problem.

The hidden subgroup problem can be translated to the quantum stabilizer problem as follows. Let $f : G \to \{0, 1\}^s$ be a function which hides the subgroup H. The *right shift* of the function f by $y \in G$ is the function ${}^{y}f$ given by ${}^{y}f(x) = f(xy)$. Obviously the group G acts as a permutation group on the right shifts of f:

$${}^{yz}f(x) = f(xyz) =^{z} (f)(xy) =^{y} ({}^{z}f(x)).$$

With respect to this action, the stabilizer of f is the hidden subgroup H. To turn the action on the shifts of f into a quantum permutation action we consider the *graphs* of the shifted functions ${}^yf$. The graph of the function $f : G \to \{0,1\}^t$ is the unit vector $\frac{1}{\sqrt{|G|}}\sum_{x\in G}|x\rangle \otimes |f(x)\rangle$ in $\mathbb{C}^{2^{\ell+t}}$. If the elements of G are encoded by normal words in a polycyclic presentation, the Fourier transform of an abelian group having the same order as G (see Subsection 2.5.7) can be used to compute the vector $\frac{1}{\sqrt{|G|}}\sum_{x\in G}|x\rangle \otimes |\underline{0}\rangle$ in polynomial time. From this vector the graph of f is obtained by an application of the oracle for f. Finally, observe that given $y \in G$, the graph of the shift ${}^yf$ of of an arbitrary function f can be obtained from the graph of f just by multiplying the the first register by y^{-1} from the right:

$$\frac{1}{\sqrt{|G|}}\sum_{x\in G}|x\rangle \otimes |f(xy)\rangle = \frac{1}{\sqrt{|G|}}\sum_{x'\in G}|x'y^{-1}\rangle \otimes |f(x')\rangle.$$

Thus, if the quantum stabilizer problem over a G can be solved in time polynomial in $\ell + t$ then the same holds for the hidden subgroup problem over G. (Note that in this reduction the number of copies of the input vector used in the stabilizer computation is just the number of queries to the function oracle.) In particular, from Theorem 9.1 we immediately obtain.

Corollary 9.2. *Let G be as in Theorem 9.1. Then the hidden subgroup problem over G can be solved in time polynomial in ℓt, where ℓ is the encoding length of G and the values of the subgroup hiding function are t-bit strings.*

We remark that the orbit membership problem over an abelian group G is related to the hidden subgroup problem over the semidirect product $G \rtimes \mathbb{Z}_2$ where the nontrivial element of $\mathbb{Z}_2$ act on G by taking inverses (”flipping signs” if we use the additive notation for the group operation in G). By solving the hidden subgroup problem for the restriction to G, and then going over the factor by the subgroup obtained this way, one can see that the really interesting case of the hidden subgroup problem in such semidirect product groups is the case where the hidden subgroup intersects G trivially. This means that the hidden subgroup is either trivial or of the form $1 \cup (y,1)$ where y is an element of G. Let us define two functions f_0 and f_1 on G by $f_0(x) = f(x,0)$ and $f_1(x) = f(x,1)$ where f is the function for $G \rtimes \mathbb{Z}_2$. It turns out that if the hidden subgroup is $1 \cup (y,1)$ then f_1 is the shift of f_0 by y and hence – going over the graphs – y can be found by solving the constructive orbit membership problem over G. On the other hand, if the hidden subgroup is trivial then the ranges of f_0 and f_1 are distinct therefore the corresponding graphs are in different orbits.

One of our ”low-level” tools will be a straightforward adaptation of the standard abelian hidden subgroup algorithm for a solution of the quantum stabilizer problem. The other tool will be an algorithm that solves the effective orbit membership problem over elementary abelian groups of constant exponent. The ”high-level” structure of our method is based on the following.

Observation 9.3. *Let G be a finite group acting on a finite set Ψ where the stabilizer of $\underline{\psi} \in \Psi$ is the subgroup H. Let N be a normal subgroup of G and let $x_1,\ldots,x_r$ be elements of G such that the cosets $x_1N,\ldots,x_rN$ generate the subgroup HN/N of G/N. For every index $i = 1,\ldots,r$ let y_i be an element of N such that $y_i\underline{\psi} = x_i\underline{\psi}$. (Existence of such an element y_i is granted because $x_i \in HN$.) Then the subgroup H is generated by the set*

$$(H \cap N) \cup \{y_i^{-1}x_i | i \in I\}.$$

Proof. Notice that for every $i \in \{1, \ldots, r\}$, we have $x_i N \cap H = y_i^{-1} x_i (H \cap N)$. Let H_0 be the subgroup of G generated by the set in the statement. Obviously $H_0 \subseteq H$. To see that equality holds, it is sufficient to show equality of orders of H and H_0. To this end notice that $H_0 N = HN$ because $x_1, \ldots, x_r$ generate H modulo N. From this, using the isomorphism theorem, we infer $H/(H \cap N) \cong HN/N = H_0N/N \cong H_0/(H \cap N)$, whence the desired equality. □

Assume that N is a normal subgroup of G. Observation 9.3 suggests a reduction of the quantum stabilizer problem to determining the stabilizer modulo N, to computing the intersection of the stabilizer with N, and to certain instances of the constructive orbit membership problem.

Computing the stabilizer modulo N is based on the following. Assume that we have a quantum permutation action of G on $\Psi \subset \mathbb{C}^{2^t}$. Then the factor group G/N acts on the set Ψ' consisting of the vectors (orbit superpositions)

$$\underline{\psi}' = \frac{1}{\sqrt{|N|}} \sum_{x \in N} x\underline{\psi}.$$

The set Ψ' is an orthonormal set in $\mathbb{C}^{2^t}$ because Ψ is orthonormal. Every orbit of N on Ψ is collapsed to a single point. If H is the stabilizer of $\underline{\psi}$ then the stabilizer of $\underline{\psi}'$ under the action of G/N will be HN/N. Also, for $y \in G$ such that $y\underline{\psi}_0 = \underline{\psi}_1$ for $\underline{\psi}_0, \underline{\psi}_1 \in \Psi$, then $yN\underline{\psi}'_0 = \underline{\psi}'_1$. Thus the construction $\underline{\psi} \mapsto \underline{\psi}'$ provides a good approach for determining the stabilizer and solving constructive membership test modulo N.

To give an efficient implementation, it would be desirable to have an efficient procedure implementing a transformation which maps $\underline{\psi} \otimes |\underline{0}\rangle$ to $\underline{\psi} \otimes \underline{\psi}'$ or something similar. Recall that

$$\underline{\psi}' = \frac{1}{\sqrt{|N|}} \sum_{x \in N} x\underline{\psi}.$$

Note that it is easy to implement the map

$$\underline{\psi} \otimes |1_G\rangle \mapsto \frac{1}{\sqrt{|N|}} \sum_{x \in N} x\underline{\psi} \otimes |x\rangle.$$

This is very different form what we want. The point is that the result is not a tensor product of $\underline{\psi}'$ with some vectors in the other parts. (Physicists would say that the two parts are entangled.) The main idea of our method is a kind of disentangling the two parts using a constructive orbit membership algorithm. To demonstrate how this works, assume for a moment that the stabilizer intersects N trivially and we have a procedure for solving the constructive orbit membership over N using single instances (that is, the problem ORBIT-MEMBER$_1(N, \Psi, \underline{\psi}_0, \underline{\psi}_1)$. We further assume that as a part of input, we have two copies of $\underline{\psi}$. Then we can efficiently produce the vector

$$\underline{\psi} \otimes \frac{1}{\sqrt{|N|}} \sum_{x \in N} x\underline{\psi} \otimes |x\rangle.$$

We apply the inverse of the constructive orbit membership test to this superposition. As the membership test maps vectors of the form $\underline{\psi} \otimes x\underline{\psi} \otimes |0\rangle$ to $\underline{\psi} \otimes x\underline{\psi} \otimes |x\rangle$, the result after application of the inverse is

$$\underline{\psi} \otimes \frac{1}{\sqrt{|N|}} \sum_{x \in N} x\underline{\psi} \otimes |0\rangle.$$

This vector is already a tensor product of $\underline{\psi}'$ with other parts and can be used with the action of G/N.

Unfortunately, the approach outlined above does not work. The main problem that except for very special cases, ORBIT-MEMBER$_1(N, \Psi, \underline{\psi}_0, \underline{\psi}_1)$ cannot be solved at all. Actually, we can solve ORBIT-MEMBER$_K(N, \Psi, \underline{\psi}_0, \underline{\psi}_1)$ in a reasonable wide class of groups N with sufficiently small error only if K (the number of copies of $\underline{\psi}_0$ and $\underline{\psi}_1$ given in the input) is large enough.

To deal with the difficulty above, we will sometimes replace permutation actions with equivalent ones. Here equivalence is just the usual equivalence of permutation representations. That is, if G acts on $\Psi \subset \mathbb{C}^{2^t}$ and $\Psi' \subset \mathbb{C}^{2^{t'}}$ then the two actions are called equivalent if there is a bijection $\mathcal{F} : \Psi \to \Psi'$ such that $\mathcal{F}(x\underline{\psi}) = x\mathcal{F}(\underline{\psi})$ for every $x \in G$ and for every $\underline{\psi} \in \Psi$. Obviously, the stabilizer of $\mathcal{F}(\underline{\psi})$ coincides with that of $\underline{\psi}$, and, similarly, the solution of orbit membership problem for $\underline{\psi}_0$ and $\underline{\psi}_1$ is the same as for $\mathcal{F}(\underline{\psi}_0)$ and $\mathcal{F}(\underline{\psi}_1)$.

We use equivalent permutation actions in the following context. Let G act on $\Psi \subset \mathbb{C}^{2^t}$ and let K be a natural number. For $\underline{\psi} \in \Psi$ we set $\mathcal{F}_K(\underline{\psi})$ as the Kth tensor power of $\underline{\psi}$. That is,

$$\mathcal{F}_K(\underline{\psi}) = \underline{\psi}^{\otimes K} = \underline{\psi} \otimes \cdots \otimes \underline{\psi} \in \mathbb{C}^{2^{tK}}.$$

Let $\Psi' = \mathcal{F}_K(\Psi) \subset \mathbb{C}^{2^{tK}}$ be just the set consisting of the vectors $\mathcal{F}_K(\underline{\psi})$ where $\underline{\psi} \in \Psi$. The action of G on Ψ' is just the diagonal action obtained from the action on Ψ and the procedure for the diagonal action can be clearly implemented by K applications of the procedure for the original action.

The rest of this chapter is structured as follows. In Section 9.1 we describe an efficient quantum algorithm solving the stabilizer problem for abelian groups. Actually the method is a straightforward adaptation of the standard Fourier sampling method and we give the details just for convenience of the reader. In Section 9.2 we describe the quantum part of an approach solving the effective orbit membership problem over abelian groups. It is also based on the abelian Fourier sampling method and consists in a reduction to some classical statistical analysis. We solve the latter problem in polynomial time over abelian p-groups of constant exponent in Section 9.3. We put these ingredients together using a high-level reduction procedure following the lines sketched above in Section 9.4. This gives a proof for Theorem 9.1.

9.1 Abelian quantum stabilizer

In this section we describe a quantum algorithm for solving the quantum stabilizer problem STABILIZER$_K(G, \Psi, \underline{\psi})$ where G is a finite abelian group. It will be convenient to assume that G is presented as a direct sum of cyclic groups of prime power order. By this we mean that we are given prime powers $m_1, \ldots, m_n$ and an element of $G = \mathbb{Z}_{m_1} \oplus \cdots \oplus \mathbb{Z}_{m_n}$ is represented by a row vector $z_1, \ldots, z_n$ of integers where $0 \leq z_i < m_i$. We assume that such vectors are encoded by bit strings of length ℓ. We use the additive notation for G. Note that this presentation is very close to a special case of a refined polycyclic presentation. The set Ψ is an orthonormal set of unit vectors in $\mathbb{Z}^{2^t}$.

Because we want to use the algorithm as a quantum subroutine, we have to define the task accurately. A quantum algorithm solving STABILIZER$_K(G, \Psi, \underline{\psi})$ should implement a unitary transformation which maps vectors of the form

$$\underline{\psi}^{\otimes K} \otimes |\underline{0}\rangle \in \mathbb{C}^{2^{Kt}} \otimes \mathbb{C}^{2^{n\ell}},$$

to the vector

$$\underline{\psi}^{\otimes K} \otimes |H\rangle,$$

where H is the stabilizer of $\underline{\psi}$ and $|H\rangle$ stands for the computational basis vector of $\mathbb{C}^{2^{n\ell}}$ corresponding to the description of H. Note that actually one should give a more precise description which includes the workspace. Our procedure is "clean" in the sense of Subsection 2.5.2, and we do not include in the input/output defintion the workspace which contains both initially and finally zero qubits.

We use the standard abelian hidden subgroup algorithm – called Fourier sampling – adapted to our setting. The preparatory part consists of K rounds. In each round, we pick a copy of $\underline{\psi}$ form the input and take a register of size ℓ from the workspace. In that register we prepare the uniform superposition $\frac{1}{\sqrt{|G|}}\sum_{x\in G}|x\rangle$ of elements of G. Note that there are standard polynomial time methods making (approximations of) uniform superpositions over abelian groups, e.g., an application of the quantum Fourier transform, see Subsection 2.5.7.

We concentrate on the contents of the subsystem corresponding of the two registers we are working with. The actual state is a vector which is a tensor product of the contents of the present two-register part with the contents of the other registers and the zero qubits together with the possible garbage in the workspace. Our present pair of registers contain the vector

$$\frac{1}{\sqrt{|G|}}\sum_{x\in G}\underline{\psi}\otimes|x\rangle.$$

We apply the oracle for the G-action on Ψ and obtain the vector

$$\frac{1}{\sqrt{|G|}}\sum_{x\in G}x\underline{\psi}\otimes|x\rangle.$$

Let $x_1,\ldots,x_{|G:H|}$ be a system of coset representatives for H in G. Then the vector we have equals

$$\frac{1}{\sqrt{|G|}}\sum_{i=1}^{|G:H|}x_j\underline{\psi}\otimes\sum_{x\in H}|x_j+x\rangle.$$

Now we apply the quantum Fourier transform of G (see Subsection 2.5.7) to the second register. After the Fourier transformation, our pair of registers contains the vector

$$\frac{1}{|G|}\sum_{i=1}^{|G:H|}x_j\underline{\psi}\otimes\sum_{\chi\in\widehat{G}}\sum_{x\in H}\chi(x_j+x)|\chi\rangle=\frac{1}{|G:H|}\sum_{i=1}^{|G:H|}x_j\underline{\psi}\otimes\sum_{\chi\in\widehat{G}}\frac{\chi(x_j)}{|H|}\sum_{x\in H}\chi(x)|\chi\rangle$$

Here we used that $\chi(x_j+x)=\chi(x_j)\chi(x)$. By the orthogonality relations for the restriction of χ to H and the trivial character 1_H of H, we have

$$\frac{1}{|H|}\sum_{x\in H}\chi(x)=\begin{cases}1 & \text{if } \chi_{|H}=1_H,\\ 0 & \text{otherwise.}\end{cases}$$

Let $H^\perp \subseteq \widehat{G}$ stand for the set of characters of G whose restriction to H is the trivial character. It turns out that $H^\perp$ is isomorphic to the factor group G/H. Based on the above, the contents of our pair is

$$\frac{1}{\sqrt{|G:H|}} \sum_{\chi \in H^\perp} \underline{\psi}_\chi \otimes |\chi\rangle,$$

where

$$\underline{\psi}_\chi = \frac{1}{\sqrt{|G:H|}} \sum_{i=1}^{|G:H|} \chi(x_j) x_j \underline{\psi}$$

is a unit vector of $\mathbb{C}^{2^t}$ depending on χ. Notice that our pair contains the uniform superposition of the vectors $\underline{\psi}_\chi \otimes |\chi\rangle$. These vectors are pairwise orthogonal because for different characters χ_1 and χ_2 the vectors $|\chi_1\rangle$ and $|\chi_2\rangle$ are orthogonal.

We repeat this procedure K times, always with a distinct copy of $\underline{\psi}$. As a result, we obtain the vector

$$\left(\frac{1}{\sqrt{|G:H|}} \sum_{\chi \in H^\perp} \underline{\psi}_\chi \otimes |\chi\rangle \right)^{\otimes K}.$$

We apply a method following the lines given in Subsection 2.5.5. Here, based on the above observation, the distribution corresponding to the state $\frac{1}{\sqrt{|G:H|}} \sum_{\chi \in H^\perp} \underline{\psi}_\chi \otimes |\chi\rangle$ is the uniform distribution over $H^\perp$. We apply a (classical) procedure that, given a sequence $\chi_1, \ldots, \chi_K$ of characters of G computes (in a register from the workspace) the subgroup $H(\chi_1, \ldots, \chi_K)$ consisting of the of elements of G on which all the characters $\chi_1, \ldots, \chi_K$ take value 1. This can be done in polynomial time in general abelian groups. We do not give the details, only note that in the case where $G \cong \mathbb{Z}_p^n$ for some prime p, the characters of G are of the form χ_x, where θ is a primitive pth root of unity, $x \cdot y$ stands for the scalar product of vectors in $\mathbb{Z}_p$, and $\chi_x(y) = \theta^{x \cdot y}$. Therefore the subgroup we are looking for is the solution space of a system of homogeneous linear equations. The error of the quantum procedure is related to the probability of that $\chi_1, \ldots, \chi_K$ *do not* generate $H^\perp$ where χ_i are drawn independently and uniformly from $H^\perp$. For $K = O(\log |G|)$ this probability is at most one percent. By repeating the procedure $O(\log \frac{1}{\epsilon})$ times and taking the majority decision we obtain probability at most $\frac{1}{2}\epsilon^2$ and hence distance will be at most ϵ.

We achieved (an approximation of) the superposition

$$\underline{\psi}_{\chi_1} \otimes |\chi_1\rangle \otimes \cdots \otimes \underline{\psi}_{\chi_K} \otimes |\chi_K\rangle \otimes |H\rangle,$$

or, more precisely a tensor product of this vector with some garbage. As $|H\rangle$ is the unique computational basis element of $\mathbb{C}^{2^{n\ell}}$ representing the output we can copy it into the register designated to the output. Then we apply the usual cleanup: we do the reverse of the procedure carried out so far. We are left with (an approximation) of the desired output state. We have shown the following.

Proposition 9.4. *The problem* $\mathrm{STABILIZER}_K(G, \Psi, \underline{\psi})$ *over the abelian group G can be solved with error ϵ in time polynomial in $K\ell t$ if $K = (\log |G|)^{\theta(1)} \log \frac{1}{\epsilon}$.*

9.2 Constructive orbit membership test in abelian groups

In this section we describe the ”quantum part” of an approach to solving the constructive orbit membership problem over abelian groups. This is actually a reduction to a certain statistical problem, for which we give an efficient classical solution over abelian p-groups of constant exponent in the next section.

Like in the preceding section, we assume that the abelian group G is given as a direct sum of cyclic subgroups of prime power order and we consider a quantum permutation action on an orthonormal set $\Psi \subset \mathbb{C}^{2^t}$. The input state for an instance of $\mathrm{ORBIT\text{-}MEMBER}_K(G, \Psi, \underline{\psi}_0, \underline{\psi}_1)$ is the vector

$$\underline{\psi}_0^{\otimes K} \otimes \underline{\psi}_1^{\otimes K} \otimes |\underline{0}\rangle \in \mathbb{C}^{2^{2t+\ell}},$$

where the third part is reserved for the result.

Here we consider only instances where the stabilizers of $\underline{\psi}_0$ and $\underline{\psi}_1$ are trivial. This restriction is justified as follows. In the general case we can first use the algorithm of the preceding section to compute the stabilizers H_0 and H_1 of $\underline{\psi}_0$ and $\underline{\psi}_1$, respectively. If $H_1 \neq H_0$ then the two elements are in different orbits (recall that G is abelian). Otherwise we replace G with the factor group G/H_0. To implement the G/H_0-action, observe that any element of G/H_0 acts on the orbits of $\underline{\psi}_0$ and $\underline{\psi}_1$ as an arbitrary representative from the coset.

The outcome of a procedure solving an instance of $\mathrm{ORBIT\text{-}MEMBER}_K(G, \Psi, \underline{\psi}_0, \underline{\psi}_1)$ with trivial stabilizers should be the vector

$$\underline{\psi}_0^{\otimes K} \otimes \underline{\psi}_1^{\otimes K} \otimes |u\rangle \in \mathbb{C}^{2^{2t+\ell}},$$

where u is the unique solution: it is either ”none” or the group element carrying $\underline{\psi}_0$ to $\underline{\psi}_1$. Like in the preceding section we omit the workspace from the input and output state.

In order to exclude the ”degenerate” case where $\underline{\psi}_1 = \underline{\psi}_0$, we begin with an application of the swap test described in Subsection 2.5.5 (to the K copies of the pair) to decide if $\underline{\psi}_0 = \underline{\psi}_1$. We assume that K is large enough so that the error probability is at most $\frac{1}{18}\epsilon^2$ and the output of the test is in the first qubit of the workspace. These assumptions assure that after the swap test the distance of the vector we have from

$$\underline{\psi}_0^{\otimes K} \otimes \underline{\psi}_1^{\otimes K} \otimes |\underline{0}\rangle \otimes |b\rangle$$

is at most $\epsilon/3$, where $b = 1$ if $\underline{\psi}_1 = \underline{\psi}_0$ and $b = 0$ otherwise.

If the test returns 1 (representing ”*yes*”), we put 1_G in the output register and undo (do the reverse of) the swap test.

Otherwise the next phase of the algorithm consists of K rounds. In each round, one copy of $\underline{\psi}_0$, one copy of $\underline{\psi}_1$ are used together with ℓ qubits and a further qubit from the workspace. In the ℓ-qubit register we prepare the uniform superposition $\frac{1}{\sqrt{|G|}} \sum_{x \in G} |x\rangle$ (say, using the Fourier transform of G, like in the preceding section). In the one-qubit register we prepare the vector $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ using the Hadamard gate. Then, in superposition, we conditionally exchange the contents of the first two registers: We do nothing if the last qubit is zero and we apply the unitary transformation on the first two registers which exchanges their contents. As a result we obtain the following vector.

$$\frac{1}{\sqrt{2|G|}} \sum_{x \in G} \left(\underline{\psi}_0 \otimes \underline{\psi}_1 \otimes |x\rangle \otimes |0\rangle + \underline{\psi}_1 \otimes \underline{\psi}_0 \otimes |x\rangle \otimes |1\rangle \right).$$

(More precisely, this is just the part we work with in the current round. We actually have a tensor product of this vector with other parts.) Next we apply the permutation procedure to the third and the first registers; and its inverse to the third and the second registers. We obtain the vector

$$\frac{1}{\sqrt{2|G|}}\sum_{x\in G}\Big(x\underline{\psi}_0\otimes(-x)\underline{\psi}_1\otimes|x\rangle\otimes|0\rangle+x\underline{\psi}_1\otimes(-x)\underline{\psi}_0\otimes|x\rangle\otimes|1\rangle\Big).$$

We collect the terms according to the different values of in the first two registers. If $\underline{\psi}_1$ is not in the orbit of $\underline{\psi}_0$ then the set

$$\{x\underline{\psi}_0\otimes(-x)\underline{\psi}_1|x\in G\}\cup\{\underline{\psi}_1\otimes(-x)\underline{\psi}_0|x\in G\}$$

consists of $2|G|$ pairwise orthogonal vectors. If $\underline{\psi}_1=u\underline{\psi}_0$, then this set contains only n pairwise orthogonal vectors: $(x+u)\underline{\psi}_0\otimes(-x-u)\underline{\psi}_1$ coincides with $x\underline{\psi}_1\otimes(-x)\underline{\psi}_0$ and our vector can be written as

$$\frac{1}{\sqrt{2|G|}}\sum_{x\in G}x\underline{\psi}_1\otimes(-x)\underline{\psi}_0\otimes(|x+u\rangle\otimes|0\rangle+|x\rangle\otimes|1\rangle).$$

We apply the Fourier transform of G to the third register and the Hadamard transform to the fourth register. (We remark that these two transformations together give actually the Fourier transform of the direct product $G\times\mathbb{Z}_2$. The idea of applying transforms of this kind goes back to an exponential time algorithm [33] for solving the hidden subgroup problem in dihedral groups.) If $\underline{\psi}_1=u\underline{\psi}_0$ then we obtain

$$\frac{1}{2|G|}\sum_{\chi\in\widehat{G}}\sum_{x\in G}x\underline{\psi}_1\otimes(-x)\underline{\psi}_0\otimes(\chi(x+u)|\chi\rangle\otimes(|0\rangle+|1\rangle)+\chi(x)|\chi\rangle\otimes(|0\rangle-|1\rangle)).$$

In this case, for $\chi\in\widehat{G}$ and $i\in\{0,1\}$, put

$$\underline{\psi}_{\chi,i}=\frac{1}{\sqrt{|G|}}\sum_{x\in G}x\underline{\psi}_1\otimes(-x)\underline{\psi}_0$$

and

$$c_{\chi,i}=\frac{1}{2\sqrt{|G|}}\sum_{x\in G}(\chi(x+u)+(-1)^i\chi(x)).$$

In the case when $\underline{\psi}_0$ and $\underline{\psi}_1$ are in different orbits put

$$\underline{\psi}_{\chi,i}=\frac{1}{\sqrt{2|G|}}\sum_{x\in G}x\underline{\psi}_i\otimes(-x)\underline{\psi}_{1-i}$$

and

$$c_{\chi,i}=\frac{1}{\sqrt{2|G|}}.$$

Then for every character $\chi\in\widehat{G}$, $\underline{\psi}_\chi$ is a unit vector in $\mathbb{C}^{2^{2t}}$ and our vector can be written as the superposition

$$\sum_{\chi\in\widehat{G},i\in\{0,1\}}c_{\chi,i}\underline{\psi}_{\chi,i}\otimes|\chi\rangle\otimes|i\rangle.$$

After K rounds we have the tensor power

$$\left(\sum_{\chi\in\widehat{G},i\in\{0,1\}} c_{\chi,i}\underline{\psi}_{\chi,i}\otimes|\chi\rangle\otimes|i\rangle\right)^{\otimes K}.$$

Like in the preceding section we apply a method in the framework explained in Subsection 2.5.5: we pass this superposition to a classical algorithm which analyzes the sequence $S=(\chi_1,i_1),\ldots,(\chi_K,i_K)$ and tries to guess the subgroup U generated by u (or return "none" if $\underline{\psi}_1$ is not in the orbit of $\underline{\psi}_0$). The error analysis of the quantum procedure corresponds to the probabilistic error analysis of our classical procedure which analyzes a random sequence of pairs (χ,i) from $G\times\{0,1\}$ whose members are drawn independently with probability $|c_{\chi,i}|^2$. We assume that K is large enough that the probability (according to the distribution given by the numbers $|c_{\chi_j,i_j}|^2$) of that the classical statistical algorithm returns a wrong answer is at most $\frac{1}{18}\epsilon^2$. We copy the result into a new register from the workspace and undo the computations so far.

After the statistical procedure, the distance of the vector we have from

$$\underline{\psi}_0^{\otimes K}\otimes\underline{\psi}_1^{\otimes K}\otimes|U\rangle$$

is now at most $\frac{2}{3}\epsilon$. To finish the description of the quantum part, assume first that the procedure returned "none". Then we write "none" in the output register and do the reverse of the procedures executed so far. If a subgroup U is returned then, assuming that K is at least as large as r times the exponent of G where r is large enough such that the swap test on r copies of pairs errs with probability at most $\frac{1}{18}\epsilon^2$, we apply an exhaustive search for u over the subgroup U based on the swap test. We copy u to the output register and undo all the computations performed so far.

In any case, the distance of the vector we have at this point from the desired outcome is at most ϵ.

9.3 Solving random systems of linear disequations

In this section, based on the paper [57], we discuss a problem arising from the statistical analysis relegated from the preceding section and give a polynomial time solution over abelian p-groups of constant exponent. Recall that the input for the analyzing procedure is a sample of size K over pairs $(\chi,i)\in\widehat{G}\times\mathbb{Z}_2$, drawn independently according to a distribution where either

$$\mathbf{Pr}\left[(\chi,i)\right]=|c_{\chi,i}|^2=\frac{1}{2|G|}$$

or there is an element $u\in G\setminus\{0\}$ such that

$$\mathbf{Pr}\left[(\chi,i)\right]=|c_{\chi,i}|^2=\frac{|\chi(x+u)+(-1)^i\chi(x)|^2}{4|G|}=\frac{|\chi(u)+(-1)^i|^2}{4|G|}.$$

Observe that in the first case the distribution is uniform while in the second case the probability of any pair of the form $(\chi,1)$ is zero where χ is a character of G containing u in its kernel (that is $\chi(u)=1$). The task is to distinguish between the two distributions and provide information on the element u in the second case. This information will be the subgroup generated by u.

Notice that the probability of having 1 in the second register is

$$\mathbf{Pr}\left[(\cdot,1)\right] = \sum_{\chi\in\widehat{G}} \frac{|\chi(u)-1|^2}{4|G|} = \frac{1}{4|G|}\sum_{\chi\in\widehat{G}}(2-\chi(u)-\overline{\chi(u)}) = \frac{1}{2},$$

where the last equality follows from the orthogonality relations (for the columns of the character table of $|G|$) which give $\sum_{\chi\in\widehat{G}}\chi(u) = 0$ as $u \neq 0$. Therefore if we throw the pairs of the form $(\chi, 0)$ from the sample away then the rest – if we forget about the 1 in the second part – simulates a sample over $\widehat{G}$ according to the distribution where either

$$\mathbf{Pr}\left[\chi\right] = \frac{1}{|G|}$$

or there is an element $u \in G \setminus \{0\}$ such that

$$\mathbf{Pr}\left[\chi\right] = |c_{\chi,i}|^2 = \frac{|\chi(u)-1|^2}{2|G|}.$$

Actually we only notice the subgroup of $\widehat{G}$ generated by the character χ at hand. Equivalently, we can equalize the probability of characters that generate equal subgroups of $\widehat{G}$ as follows. If character χ occurs in the sample then we draw uniformly a number $0 < j < m$ which is prime to the exponent m of G and replace χ with χ^j. We show below that we obtain a distribution which is nearly uniform on the characters χ such that $\chi(u) \neq 1$.

Lemma 9.5. *Let ω be a primitive m_0th root of unity, let m be a multiple of m_0 and let m_1 be the product of the prime divisors of m. Then*

$$\sum_{0<j<m,(j,m)=1} \omega^j = \begin{cases} \mu(m_0)\frac{m}{m_1}\phi(\frac{m_1}{m_0}), & \text{if } m_0|m_1 \\ 0 & \text{otherwise,} \end{cases}$$

where ϕ is Euler's totient function and μ is the Möbius function.

Proof. For $k|m$ we define $f(k) = \sum_{1\leq j\leq k,(j,k)=1}\omega^{\frac{m}{k}j}$. Then for every $k|m$ we have $\sum_{j=1}^{k}\omega^{\frac{m}{k}j} = \sum_{d|k} f(d)$. (This follows from the fact that every positive integer $j \leq k$ can be uniquely written in the form $j = \frac{k}{d}\cdot j'$ where $d|k$, $1 \leq j' \leq d$ and $(j', d) = 1$.) Put $F(k) = \sum_{d|k} f(d)$ for $k|m$. Then, by the Möbius inversion formula, $f(m) = \sum_{d|m}\mu(\frac{m}{d})F(d)$. We know that $F(d) = d$ if $\omega^{\frac{m}{d}} = 1$ and $F(d) = 0$ otherwise. Hence the product $\mu(\frac{m}{d})F(d)$ is nonzero if and only if $m_0|\frac{m}{d}|m_1$. Therefore $f(m) = \sum_{\frac{m}{m_1}|d|\frac{m}{m_0}}\mu(\frac{m}{d})d = \frac{m}{m_1}\sum_{d'|\frac{m_1}{m_0}}\mu(\frac{m_1}{d'})d' = \mu(m_0)\frac{m}{m_1}\sum_{d|\frac{m_1}{m_0}}\mu(\frac{m_1/m_0}{d})d$, if $m_0|m_1$ and $f(m) = 0$ otherwise. We conclude by observing that if $\ell = p_1\cdots p_r$ where the p_is are pairwise distinct primes then $\sum_{d|\ell}\mu(\frac{\ell}{d})d = \sum_{I\subseteq\{1,\ldots,r\}}(-1)^{\ell-|I|}\prod_{i\in I}p_i = \prod_{i=1}^{r}(p_i-1) = \phi(\ell)$. □

Lemma 9.6. *Let $1 \neq \omega$ be an mth root of unity. Then*

$$\frac{1}{2} \leq \frac{1}{2\phi(m)}\sum_{0<j\leq m,(m,j)=1}|1-\omega^j|^2 \leq 2.$$

Proof. The upper bound is obvious. To see the lower one, let m_0 be the order of ω and let m_1 be the product of the prime divisors of m. Observe that $|1-\omega^j|^2 = 2-\omega^j-\omega^{-j}$. Therefore $\frac{1}{2\phi(m)}\sum_{0<j\leq m,(m,j=1)}|1-\omega^j|^2 = 1-\frac{1}{\phi(m)}\sum_{0<j\leq m,(m,j=1)}\omega^j$. By Lemma 9.5, the sum on the right hand side is zero unless $m_0|m_1$. If $m_0|m_1$ then that sum has absolute value $\frac{1}{\phi(m)}\frac{m}{m_1}\phi(\frac{m_1}{m_0})$. The assertion for $m_0>2$ follows from $\phi(m)=\frac{m}{m_1}\phi(m_1)=\frac{m}{m_1}\phi(m_0)\phi(\frac{m_1}{m_0})\geq 2\frac{m}{m_1}\phi(\frac{m_1}{m_0})$. If $m_0=2$ then $\omega=-1$ and the sum is 2. □

We say that a distribution over a finite set S is *nearly uniform* with a real tolerance parameter $c\geq 1$ over a subset $S'\subseteq S$ if $\mathbf{Pr}\left[s\right]=0$ if $s\in S\setminus S'$ and $\frac{1}{c}|S'|\leq \mathbf{Pr}\left[s\right]\leq c/|S'|$ for $s\in S'$. From Lemma 9.6 we immediately obtain that the new distribution remains uniform over $\widehat{G}$ if the original was, and it becomes nearly uniform with tolerance parameter 2 over the characters containing u in their kernels in the other case. To see the second part of the statement, let m stand for the exponent of G. Then the probability of χ in the resulting distribution is

$$\mathbf{Pr}\left[\chi\right]=\frac{1}{2\phi(m)|G|}\sum_{(j,m)=1}|1-\chi(u)^j|^2.$$

By Lemma 9.6, this probability is between $\frac{1}{2|G|}$ and $\frac{2}{|G|}$. Thus our original problem reduces to the following one with $c=2$.

Random Linear Disequations(G,c) - search version
Input: Sample from a distribution over $\widehat{G}$ which is
- either nearly uniform on characters not containing a fixed element u in their kernels.
- or nearly uniform on the whole $\widehat{G}$.
Task: Decide which is the case and in the second case return the set of elements u with the required property.

In the decision version of Random Linear Disequations the goal is merely distinguishing between the two cases. Note that if u is in the expected output of the search version then so is u^t where t is relatively prime to the order of u – these are the elements which generate the same cyclic subgroup as u. The output can be represented by any of such elements. We do not parametrize the problem Random Linear Disequations with the sample size. This simplification is can be justified as follows. We imagine that the input is an infinite sample. The running time of an algorithm solving the problem is obviously an upper bound for the actually required sample size. In the context where we apply such an algorithm this bound will be satisfactory.

The name Random Linear Disequations is justified as follows. Assume that $G=\mathbb{Z}_p^n$ where p is a prime number. Then fixing a p^{th} root of unity gives a one-to one correspondence between the characters of G and homomorphisms from G to the group $\mathbb{Z}_p$. If we consider G as a vector space over $\mathbb{Z}_p$ then these homomorphisms are actually the linear functions from G to $\mathbb{Z}_p$. The task is to find the elements u of G which fail to satisfy any the homogeneous linear equations corresponding to the functions.

In Subsection 9.3.1 we show that search problem Random Linear Disequations(G,c) is in time $(\log|G|+\exp(G))^{O(1)}$ reducible to the decision version – over subgroups of G and with slightly bigger tolerance parameter $c'=2c$.

The reduction is based on the following. If B is a subgroup of G and we restrict characters of G to B then we obtain a nearly uniform distribution characters of B not containing u in their kernels. If $u \notin B$ this is a nearly uniform distribution over all characters of B.

A possible solution of the decision problem could follow the lines below. If the distribution is uniform over all characters then the kernels of the characters from a sufficiently large sample will cover the whole B. Therefore a possible way to distinguish between the two cases is to collect a sufficiently large sample of characters and to check if their kernels cover the whole group B. Unfortunately, this test is coNP-complete already for $B = \mathbb{Z}_3^n$. Indeed there is a straightforward reduction for non-colorability of graphs by 3 colors to this problem.

In Subsection 9.3.2 we describe an algorithm which solves Random Linear Disequations in p-groups. The method is based on replacing the covering condition with a stronger but much more easily testable one which is still satisfied by not too many uniformly chosen characters. The running time is polynomial in $\log |G|$ if the exponent of G is constant.

9.3.1 Reductions

In this section we show that the search version of Random Linear Disequations is reducible to its decision version in abelian groups of the form $\mathbb{Z}_m^n$.

For a finite abelian group G we denote by $\widehat{G}$ its character group. Assume that H is a subgroup of G. Then taking restrictions of characters of G to H gives a homomorphism form $\widehat{G}$ onto $\widehat{H}$. The kernel of this map is the set of characters which contain H in their kernels. This set can be identified with the character group $\widehat{(G/H)}$. It follows that every character of H has exactly $\left|\widehat{(G/H)}\right|$ extensions to G. Therefore, if a distribution is nearly uniform on characters of G then restriction to H results in a nearly uniform distribution over characters of H with the same tolerance parameter.

The same holds in the reverse direction: taking uniformly random extensions of characters of H to G transforms a nearly uniform distribution over $\widehat{H}$ to a nearly uniform distribution over $\widehat{G}$ with the same parameter. And a similar statement holds for distributions nearly uniform on the characters of H which do not contain a specific $u \in H$ in their kernels.

For restricting characters of G not containing the element $u \in G$ in their kernel we have the following.

Lemma 9.7. *Let H be subgroup of a finite abelian group G, let χ be a character of H and let $u \in G$. Then the number of characters of G extending χ such that $\chi(u) \neq 1$ is*

$$\begin{cases} |G:H|(k-1)/k & \text{if } k_0 = k \\ |G:H| & \text{if } k_0 < k, \end{cases}$$

where k is the smallest positive integer such that $k \cdot u \in H$ and $\chi(k \cdot u) = 1$ and k_0 is the smallest integer such that $k_0 \cdot u \in H$.

Proof. If $k_0 < k$ then $\chi(k_0 u) \neq 1$ therefore $\psi(u) \neq 1$ for every ψ extending χ to G. Assume that $k_0 = k$. Let B be the subgroup of G generated by H and u and let $M = \{x \in H \mid \chi(x) = 1\}$. Then every character of G extending χ takes value 1 on M, therefore it is sufficient to consider the characters of B/M extending the characters of H/M. Equivalently, we may assume that $M = 1$, and k is the order of u. Then B is the direct product of the cyclic group generated by u and H. In this case there exists exactly one character of G extending χ which take value 1 on u. Thus there are $\frac{k-1}{k}|B/H|$ characters of B with the desired property extending χ and each of them has $|G/B|$ extensions to G. □

Assume that we have an instance of the search version of Random Linear Disequations(G, c) with solution $u \in G$. Then, by the lemma above, restricting characters of G to H gives an instance of the search version Random Linear Disequations$(H, 2c)$. This gives rise to the following.

Proposition 9.8. *Let G be an abelian group and let p be the largest prime factor of $|G|$. Then, for every number $c \geq 1$, the search version of* Random Linear Disequations(G, c) *is reducible to $O(p \cdot \text{polylog}\,|G|)$ instances of the decision version of* Random Linear Disequations$(H, 2c)$ *over subgroups H of G in time $(p \cdot \log |G|)^{O(1)}$.*

Proof. The first step of the reduction is a call to the decision version of Random Linear Disequations(G, c). If it returns that the distribution is nearly uniform over the whole $\widehat{G}$ then we are done. Otherwise there is an element $u \in G$ such that the probability of drawing $\chi \in \widehat{G}$ is zero if and only if $\chi(u) = 1$. We perform an iterative search for the subgroup generated by u using Random Linear Disequations over certain subgroups U of G. Initially put $U = G$. Assume first that U is not cyclic. Then we can find a prime q such that the q-Sylow subgroup Q of U (the subgroup consisting of elements of U of q-power order) is not cyclic. But then the factor group Q/qQ is not cyclic either and we can find two subgroups M_1 and M_2 of Q of index q in Q such that the index the intersection $M = M_1 \cap M_2$ in Q is q^2. This implies $Q/M \cong \mathbb{Z}_q^2$. Let Q' be the complement of Q in G. (Recall that Q' consists of the elements of G of order prime to q.) Let $N = M + Q'$. Then $M = N \cap Q$ and $G/N \cong Q/(N \cap Q) = Q/M \cong \mathbb{Z}_q^2$. The group $\mathbb{Z}_q^2$ has $q + 1$ subgroups of order q: these are the lines through the origin in the finite plane $\mathbb{Z}_q^2$. As a consequence, there are exactly $q + 1$ subgroups $U_1, \ldots, U_{q+1}$ with index q in G containing N. Furthermore, we can find these subgroups in time polynomial in $\log |G|$ and q. Note that $G = U_1 \cup \ldots \cup U_{q+1}$. Therefore, by an exhaustive search, using the decision version of Random Linear Disequations(U_i) for $i = 1, \ldots, q + 1$, we find an index i such that $u \in U_i$. Then we proceed with U_i in place of U. In at most $\log |G|$ rounds we arrive at a cyclic subgroup U containing the desired elements u. If U is cyclic then the maximal subgroups of U are $U_1, \ldots, U_l$ where the prime divisors of $|U|$ are $p_1, \ldots, p_l$ and $U_i = p_i U$. Again using the decision version of Random Linear Disequations(U_i) for $i = 1, \ldots, l$, we either find a proper subgroup U_i containing the solutions u or find that the solutions cannot be contained in any proper subgroup of U. In the latter case the required subgroup is U. □

Finally, for the decision problem we have the following.

Proposition 9.9. *Let $G = \mathbb{Z}_{m_1} \oplus \ldots \oplus \mathbb{Z}_{m_n}$ be a finite abelian group of exponent m. (So m is the least common multiple of $m_1, \ldots, m_n$.) Then, for every real number $c \geq 1$,* Random Linear Disequations(G, c) *is reducible to* Random Linear Disequations$(\mathbb{Z}_m^n, c)$ *in time* polylog $|G|$.

Proof. We can embed G into $B = \mathbb{Z}_m^n$ as $\frac{m}{m_1}\mathbb{Z}_m \oplus \ldots \oplus \frac{m}{m_n} Z_m$. We replace a character of G with a random extension to B. As every character of G has $|B/G|$ extensions, this transforms an instate of Random Linear Disequations(G, c) to Random Linear Disequations(B, c). □

9.3.2 An algorithm for p-groups

In this section we describe an algorithm which solves the decision version of Random Linear Disequations in polynomial time over groups of the form $\mathbb{Z}_{p^k}^n$, for every fixed prime power p^k.

For better understanding of the main ideas it will be convenient to start with a brief description of an algorithm which works in the case $k = 1$. This case is – implicitly – also solved in Section 3 of [36]. Here we present a similar method. The principal difference is that here we use polynomials rather than tensor powers. This – actually slight – modification of the approach makes it possible to generalize the algorithm to the case $k > 1$.

For the next few paragraphs we assume that $k = 1$, i.e., we are working on an instance of Random Linear Disequations over the group $G = \mathbb{Z}_p^n$. We choose a basis of G, and fix a primitive p^{th} root of unity ω. Then characters of G are of the form χ_x, where $x \in G$ and for $y \in G$ the value $\chi_x(y)$ is $\omega^{x \cdot y}$, where $x \cdot y = \sum_{i=1}^n x_i y_i$. (Here x_i and y_i are the coordinates of x and y, respectively, in terms of the chosen basis. Note that, as $\omega^p = 1$, it is meaningful to consider $x \cdot y$ as an element of $\mathbb{Z}_p$.)

Using this description of characters, we may – and will – assume that the input contains the index x rather than the character χ_x itself. We also consider G as an n-dimensional vector space over the finite field $\mathbb{Z}_p$ equipped with the scalar product $x \cdot y$ above. The algorithm will distinguish between a nearly uniform distribution over the whole group G and an arbitrary distribution where the probability of any vector orthogonal to a fixed vector $0 \neq u$ is zero.

We claim that in the case of a distribution of the latter type there exists a polynomial $Q \in \mathbb{Z}_p[x_1, \ldots, x_n]$ of degree $p-1$. such that for every x which occur with nonzero probability we have $Q(x) = 0$. Indeed, for any fixed u with the property above, $(\sum u_i x_i)^{p-1} - 1$ is such a polynomial by Fermat's little theorem.

On the other hand, if the distribution is nearly uniform over the whole group then, for sufficiently large sample size K, with high probability there is no nonzero polynomial $Q \in \mathbb{Z}_p[x_1, \ldots, x_n]$ of degree at most $p - 1$ such that $Q(a^{(i)}) = Q(a_1^{(i)}, \ldots, a_n^{(i)}) = 0$ for every vector $a^{(i)}$ from the sample $a^{(1)}, \ldots, a^{(K)}$.

This can be seen as follows. Let us consider the vector space W of polynomials of degree at most $p-1$ in n variables over the field $\mathbb{Z}_p$. Substituting a vector $a = (a_1, \ldots, a_n)$ into polynomials Q is obviously a linear function on W. Therefore for any $K_1 \leq K$, the polynomials vanishing at $a^{(1)}, \ldots, a^{(K_1)}$ is a linear subspace W_{K_1} of W. Furthermore, by the Schwartz–Zippel lemma (see Section 2.4), the probability of that a uniformly drawn vector a from $\mathbb{Z}_p^n$ is a zero of a particular nonzero polynomial of degree $p - 1$ (or less) is at most $(p - 1)/p$. This implies that with probability proportional to $1/cp$, the subspace W_{K_1+1} is strictly smaller than W_{K_1} unless W_{K_1} is zero. From this we infer that, if the sample size K is proportional to $p \cdot \dim W$ then with high probability, W_K will be zero. Also, we can compute W_K by solving a system of K linear equations over $\mathbb{Z}_p$ in $\dim W = \binom{n+p-1}{n} = n^{O(p)}$ variables.

As already mentioned in Section 2.4, the key ingredient of the argument above – the Schwartz-Zippel bound on the probability of hitting a nonzero of a polynomial – is also known from coding theory. Namely we can encode such a polynomial $Q(x) = Q(x_1, \ldots, x_n)$ with the vector consisting of all the values $P(a) = P(a_1, \ldots, a_n)$ taken at all the vectors $a = (a_1, \ldots, a_n)$ in $\mathbb{Z}_p^n$. This is a linear encoding of W and the image of W under such an encoding is a well known generalized Reed–Muller code. The relative distance of this code is $1/p$.

We turn to the general case: below we present an algorithm solving Random Linear Disequations in the group $G = \mathbb{Z}_{p^k}^n$ where k is a positive integer. Like in the case $k = 1$, the characters of the group $G = \mathbb{Z}_{p^k}^n$ can be indexed by elements of G when we fix a basis of G and a primitive p^kth root of unity ω: $\chi_x(y) = \omega^{x \cdot y}$, where $x \cdot y$ is the sum of the product of the coordinates of x and y in terms of the fixed basis. Again, we can consider $x \cdot y$ as an element of $\mathbb{Z}_{p^k}$. In view of this, it is sufficient to present a method that distinguishes between a nearly uniform distribution over $\mathbb{Z}_{p^k}^n$, and an arbitrary one where vectors which are orthogonal to a fixed vector $u \neq 0$ have zero probability.

The method is based on the idea outlined above for the case $k = 1$ combined with an encoding of elements of $\mathbb{Z}_{p^k}$ by k-tuples of elements of $\mathbb{Z}_p$. The encoding is the usual base p expansion, that is, the bijection $\delta : \sum_{j=0}^{k-1} a_j p^j \mapsto (a_0, \ldots, a_{k-1})$. We can extend this map to a bijection between $\mathbb{Z}_{p^k}^n$ and $\mathbb{Z}_p^{kn}$ in a natural way.

Obviously the image under δ of a nearly uniform distribution over $\mathbb{Z}_{p^k}^n$ is nearly uniform over $\mathbb{Z}_p^{kn}$. In the next few lemmas we are going to show that for every $0 \neq u \in \mathbb{Z}_{p^k}^n$ there is a polynomial Q of "low" degree in kn variables such that for every vector $a \in \mathbb{Z}_{p^k}^n$ not orthogonal to u, the codeword $\delta(a)$ is a zero of Q.

We begin with a polynomial expressing the *carry term* of addition of two base p digits.

Lemma 9.10. *There is a polynomial $C(x, y) \in \mathbb{Z}_p[x, y]$ of degree at most $2p - 2$ such that for every pair of integers $a, b \in \{0, \ldots, p-1\}$, $C(a, b) = 0$ if $a + b < p$ and $C(a, b) = 1$ otherwise.*

Proof. For $i \in \{0, \ldots, p-1\}$, let $L_i(z) \in \mathbb{Z}_p[z]$ denote the Lagrange polynomial $\prod_{0 \leq j < p : j \neq i} (z - j)/(i - j)$. We have $L_i(i) = 1$ and $L_i(j) = 0$ for $j \neq i$. Define $C(x, y) = \sum_{0 \leq i,j < p : i+j \geq p} L_i(x) L_j(y)$. □

Using the carry polynomial $C(x, y)$ we can also express the base p digits of sums by polynomials.

Lemma 9.11. *For every integer $T \geq 1$, there exist polynomials Q_i from the polynomial ring $\mathbb{Z}_p[y_{1,0}, \ldots, y_{1,k-1}, \ldots, y_{T,0}, \ldots, y_{T,k-1}]$, $(i = 0, \ldots, k-1)$ with $\deg Q_i \leq (2p-2)^i$ such that*

$$\delta\left(\sum_{t=1}^{T} a_t \mod p^k\right) = (Q_0(\delta(a_1), \ldots, \delta(a_T)), \ldots, Q_{k-1}(\delta(a_1), \ldots, \delta(a_T)))$$

for every $a_1, \ldots, a_T \in \mathbb{Z}_{p^k}$.

Proof. The proof is accomplished by induction on k. For $k = 1$ the statement is obvious: we can take $Q_0 = \sum_{t=1}^{T} y_{t,0}$. Now let $k > 1$. Again set $Q_0 = \sum_{t=1}^{T} y_{t,0}$ and for $t = 2, \ldots, T$ set $C_t = C\left((\sum_{j=1}^{t-1} y_{j,0}), y_{t,0}\right)$. Then for every $a_1, \ldots, a_T \in \mathbb{Z}_{p^k}$, the digits $s_0, \ldots, s_{k-1}$ of the sum $s = \sum_{t=1}^{T} a_t \mod p^k$ satisfy

$$\begin{aligned} s_0 &= Q_0(a_{1,0}, \ldots, a_{n,0}) \mod p, \\ \sum_{j=1}^{k-1} s_j p^{j-1} &= \sum_{t=1}^{T} \lfloor a_t / p \rfloor + \sum_{t=2}^{T} c_t \mod p^{k-1}, \end{aligned}$$

where $c_t = C_t(a_{1,0}, \dots, a_{t,0})$. In other words, the 0th digit of the sum s is a linear polynomial in $a_{t,0}$, and, for $1 \leq j \leq k-1$, the jth digit is the $(j-1)$th digit in the RHS term of the second equation. There we have a sum of $2T-1$ terms and each digit of each term is a polynomial of degree at most $2p-2$ in the $a_{t,j}$. Therefore we can conclude using the inductive hypothesis applied to that (longer) sum. □

Recall that we extended δ to $\mathbb{Z}_{p^k}^n$ in the natural way. To be specific, for $a = (a_1, \dots, a_n) \in \mathbb{Z}_{p^k}^n$ we define $\delta(a) \in \mathbb{Z}_p^{kn}$ as the vector $(a_{1,0}, \dots, a_{n,k-1}) \in \mathbb{Z}_p^{kn}$ where $a_{i,j}$ is the jth coordinate of $\delta(a_i) \in \mathbb{Z}_p^k$. We can express the digits of the scalar products of a vector from $\mathbb{Z}_{p^k}^n$ with a fixed one as follows.

Lemma 9.12. *For every $u \in \mathbb{Z}_{p^k}^n$, there exist polynomials $Q_i \in \mathbb{Z}_p[x_{1,0}, \dots, x_{n,k-1}]$ of total degree at most $(2p-2)^i$, for $i = 0, \dots, k-1$, such that $\delta(a \cdot u) = (Q_0(\delta(a)), \dots, Q_{k-1}(\delta(a)))$ for every $a \in \mathbb{Z}_{p^k}^n$.*

Proof. The statement follows from Lemma 9.11 by repeating u_i times the coordinate x_i, and taking the sum of all the terms obtained this way modulo p^k. □

In order to simplify notation, for the rest of this subsection we set $x_{jp+i} = x_{i,j}$ ($j = 0, \dots, k-1$, $i = 1, \dots, n$). For every positive integer D, let $\mathbb{Z}_p^D[x_1, \dots, x_{nk}]$ be the linear subspace of polynomials of $\mathbb{Z}_p[x_1, \dots, x_{nk}]$ whose total degree is at most D and partial degrees are at most $p-1$ in each variable.

Together with Fermat's little theorem, the previous lemma implies a polynomial characterization over $\mathbb{Z}_p$ of vectors in $\mathbb{Z}_{p^k}^n$ that are not orthogonal to a fixed vector $u \in \mathbb{Z}_{p^k}^n$.

Lemma 9.13. *Let $D = \frac{(p-1)((2p-2)^k-1)}{2p-3}$. For every $u \in \mathbb{Z}_{p^k}^n$, there exists a polynomial $Q_u \in \mathbb{Z}_p^D[x_1, \dots, x_{nk}]$ such that for every $a \in \mathbb{Z}_{p^k}^n$, $a \cdot u \neq 0 \mod p^k$ if and only if $Q_u(\delta(a)) = 0$.*

Proof. Let $Q = \prod_{j=0}^{k-1}(Q_j^{p-1} - 1)$, where the polynomials Q_j come from Lemma 9.12. This polynomial has the required total degree. To ensure that partial degrees are less than $p-1$, we replace x_i^p terms with x_i until every partial degree is at most $p-1$. Let Q_u be the polynomial obtained this way. Then Q_u and Q encode the same function over $\mathbb{Z}_p^{nk}$ and hence the polynomial Q_u satisfies the required conditions. □

It remains to show that if K is large then with high probability, for a sample $a_1, \dots, a_K$ taken accordingly to a nearly uniform distribution over $\mathbb{Z}_p^{nk}$, there is no nonzero polynomial in $\mathbb{Z}_p^D[x_1, \dots, x_{nk}]$ vanishing at all the points $a_1, \dots, a_K$ where D is as in Lemma 9.13. Furthermore, we also need an efficient method for demonstrating this.

To this end, for every $a \in \mathbb{Z}_p^{nk}$, we denote by ℓ_a the linear function over polynomials in $\mathbb{Z}_p^D[x_1, \dots, x_{nk}]$ that satisfies $\ell_a(Q) = Q(a)$. Deciding whether the zero polynomial is the the only polynomial in $\mathbb{Z}_p^D[x_1, \dots, x_{nk}]$ such that $\ell_{a_i}(Q) = 0$ amounts to determining the rank of the the $K \times \Delta$ matrix whose entries are $\ell_{a_i}(M)$ where M runs over the monomials in $\mathbb{Z}_p^D[x_1, \dots, x_{nk}]$. Here Δ stands for the dimension of $\mathbb{Z}_p^D[x_1, \dots, x_{nk}]$. Note that $\Delta \leq \binom{kn+D-1}{kn}$.

The image of the space $\mathbb{Z}_p^D[x_1, \dots, x_{nk}]$ under the linear map $L : Q \mapsto (\ell_a(Q))_{a \in \mathbb{Z}_p^{nk}}$ is known as a generalized Reed–Muller code with minimal weight at least $(p-s)p^{nk-r-1} \leq p^{nk-\lceil D/(p-1) \rceil}$, where r, s are integers such that $0 \leq s < p-1$ and $\mathrm{Max}\{D, (p-1)nk\} = r(p-1) + s$ cf. [2]. For $K_1 \leq K$, let W_{K_1} stand for the subspace of polynomials in

$\mathbb{Z}_p^D[x_1,\ldots,x_{nk}]$ vanishing at all the points $a_1,\ldots,a_{K_1}$. The minimal weight bound above gives that for $K_1 < K$,

$$\mathbf{Pr}\left[W_{K_1+1} < W_{K_1} | W_{K_1} \neq 0\right] \geq \frac{1}{c} \cdot p^{-\lceil D/(p-1)\rceil}.$$

Here c is the parameter of near uniformity. The formula above implies that if

$$K = \theta(cp^{\lceil D/p-1\rceil} \dim \mathbb{Z}_p^D[x_1,\ldots,x_{nk}]) = c(pnk)^{O((2p)^k)},$$

then with probability at least 2/3, W_K will be zero - provided that we have a nearly uniform distribution with parameter c. (In the second bound we have used that $D = \frac{(p-1)((2p-2)^k-1)}{2p-3} = O((2p)^k)$. Together with the remark on rank computation this gives the following.

Theorem 9.14. Random Linear Disequations$(\mathbb{Z}_{p^k}^n, c)$ *can be solved with (one-sided) error probability at most* $1/3$ *in time* $c(pnk)^{O((2p)^k)}$*. In particular, for every fixed prime power* p^k*, and for every fixed constant* c*,* Random Linear Disequations$(\mathbb{Z}_{p^k}^n, c)$ *can be solved in time polynomial in* n*.*

□

Note that with independent repetitions we can exponentially improve the error probability. Together with the quantum part described in Section 9.2 this implies the following.

Corollary 9.15. *Assume that we have a quantum permutation action of the group* $G = \mathbb{Z}_{p^k}^n$ *on* Ψ*. Then, for* $K = (pnk)^{\theta((2p)^k)} \log\frac{1}{\epsilon}$ ORBIT-MEMBER$_K(G, \Psi, \underline{\psi}_0, \underline{\psi}_1)$ *can be solved by a quantum algorithm in time* $K^{O(1)}$ *with error at most* ϵ*.*

9.4 A recursion for computing stabilizers

In this section we describe our algorithm which solves STABILIZER over our special class of solvable groups. We begin with the most important ingredient which makes the recursion described in the introductory part possible.

Theorem 9.16. *Let* G *be a group and assume that* N *is a normal subgroup of* G *isomorphic to* $\mathbb{Z}_p^n$ *for some prime* p*. Suppose that we have a quantum permutation action of* G *on the set* $\Psi \subseteq \mathbb{C}^{2^t}$*. Then, the vector* $\underline{\psi}^{\otimes 2K}$ *where* $\underline{\psi} \in \Psi$ *and* $K = (pn)^{\theta(p)} \log\frac{1}{\epsilon}$*, can be transformed by a quantum algorithm in time polynomial in* $((pn)^p \ell t)^{O(1)} \log\frac{1}{\epsilon}$ *to a vector* ϵ*-close to*

$$\underline{\psi}^{\otimes K} \otimes \frac{1}{\sqrt{|N|}} \sum_{y\in N} (y\underline{\psi})^{\otimes K} = \mathcal{F}_K(\underline{\psi}) \otimes \frac{1}{\sqrt{|N|}} \sum_{y\in N} y\mathcal{F}_K(\underline{\psi}).$$

Proof. We give a description as if all the ingredients worked exactly. The choice for K ensures that their cumulative error, that is the distance form the desired vector, will be at most ϵ. We begin with computing the stabilizer $N_{\underline{\psi}}$ of $\underline{\psi}$ by the method of Section 9.1, using the first K copies of $\underline{\psi}$. The (description of) $N_{\underline{\psi}}$ is computed in a register taken from the workspace. In the next step we compute a direct complement N_0 of $N_{\underline{\psi}}$ into a further register. Then we take an additional ℓ-qubit register from the workspace and prepare the superposition $\frac{1}{\sqrt{|N_0|}}\sum_{y\in N_0}|y\rangle$ in it using the quantum Fourier transform of N_0. We have the vector

$$\frac{1}{\sqrt{N}_0}\sum_{y\in N_0}\underline{\psi}^{\otimes K}\otimes\underline{\psi}^{\otimes K}\otimes|y\rangle.$$

(More precisely, we actually have a tensor product of this vector with the description of $N_{\underline{\psi}}$ and N_0 together with further 0 qubits in the workspace.) We apply the permutation action to the second K copies of $\underline{\psi}$ with the last register to obtain the vector

$$\frac{1}{\sqrt{|N_0|}}\sum_{y\in N_0}\underline{\psi}^{\otimes K}\otimes(y\underline{\psi})^{\otimes K}\otimes|y\rangle.$$

Next we apply in superposition the inverse of the algorithm for the orbit membership problem ORBIT-MEMBER$_K(N_0,\Psi,\underline{\psi},y\underline{\psi})$ to "uncompute" y in third register. We obtain

$$\frac{1}{\sqrt{|N_0|}}\sum_{y\in N_0}\underline{\psi}^{\otimes K}\otimes(y\underline{\psi})^{\otimes K}\otimes|0\rangle,$$

or, more precisely the tensor product of this with some other vectors. We undo the computations for N_0 and $N_{\underline{\psi}}$ and finally obtain the vector

$$\frac{1}{\sqrt{|N_0|}}\sum_{y\in N_0}\underline{\psi}^{\otimes K}\otimes(y\underline{\psi})^{\otimes K}=\underline{\psi}^{\otimes K}\otimes\frac{1}{\sqrt{|N|}}\sum_{y\in N}(y\underline{\psi})^{\otimes K}.$$

(More accurately – as usual – we have the tensor product of the vector above with the zero qubits of the workspace.) □

Now we are in a position to prove Theorem 9.1.

Proof of Theorem 9.1. By the assumptions G' has a series of subgroups $G'=N_1>\ldots>N_m>N_{m+1}=\{1\}$ such that for every index $1\le i\le m$, N_i is a normal subgroup of G and the factor group N_i/N_{i+1} is an elementary abelian p-group for a prime p bounded by a constant. Such a series can be efficiently computed as a refinement of the derived series of G'.

We show by induction on m that there is a constant $c=c(m)$ such that STABILIZER$_K$ can be solved in time polynomial in $K\ell t$ with error ϵ for $K=(\log|G|\log\frac{1}{\epsilon})^c$ (instead of $K=(\log|G|)^c\log\frac{1}{\epsilon})$ if $|G|$ is large enough. The desired dependence on ϵ can be obtained by the following standard technique. The result implies that a constant error (say 1/100) can be achieved with $K=(\log|G|)^{O(1)}$. By repeating the procedure with constant error $O(\log\frac{1}{\epsilon})$ times and taking the majority of the answers, the error can be made smaller than ϵ.

We assume that $c(m)$ is a monotone function of m and that $c(1)$ is large enough so that for every prime p below the bound for exponents, in any elementary abelian p-group P, $\log|P|^{c(1)}\log\frac{1}{\epsilon}$ copies of input pairs are sufficient to solve with error at most ϵ the constructive orbit membership problem, see Corollary 9.15. The initial case $m=0$ can be treated by the algorithm of Section 9.1. Assume that $m\ge 1$. We put $N=N_m$ and apply the induction hypothesis to the factor group G/N as follows. Put $c'=c(m-1)$. By the induction hypothesis, STABILIZER$_{K'}(G/N,\Psi',\underline{\psi}')$ can be solved with error $\epsilon/3$ in time polynomial in $K'\ell t'$, where $K' = (\log|G|\log\frac{3}{\epsilon})^{c'}$ and Ψ' is an orthonormal set in $\mathbb{C}^{2^{t'}}$. We apply this induction hypothesis in the context where

$$\underline{\psi}' = \frac{1}{\sqrt{|N|}}\sum_{y\in N} y\mathcal{F}_{K''}(\underline{\psi}),$$

$$\Psi' = \{\frac{1}{\sqrt{|N|}}\sum_{y\in N} xy\mathcal{F}_{K''}(\underline{\psi})|x\in G\},$$

and $K'' = (\log|G|\log\frac{3K'}{\epsilon})^d$, so that, by Theorem 9.16, from $\underline{\psi}^{\otimes 2K''}$ the vector $\underline{\psi}^{\otimes K''}\otimes\underline{\psi}'$ can be produced with error $\epsilon/(3K')$ in time polynomial in $K''\ell t$. (Here we assume that $d\geq 1$ is the constant implicit in Theorem 9.16.) Then, from $\underline{\psi}^{\otimes 2K'K''}$ an $\epsilon/3$-approximation of the vector $\underline{\psi}^{\otimes K'K''}\otimes\underline{\psi}'^{\otimes K'}$ can be produced in time polynomial in $K'K''\ell t$. By the induction hypothesis, the stabilizer of $\underline{\psi}'$ in G/N' can be computed with error $\epsilon/3$ in time polynomial in $K'K''\ell t$. Note that this stabilizer is HN/N where H is the stabilizer of $\underline{\psi}$. We make a copy of the description HN/N and undo the computations performed so far. Then we have the tensor product of the description of HN/N with the vector $\underline{\psi}^{\otimes 2K'K''}$. The cumulative error so far is at most $\frac{2}{3}\epsilon$. By Observation 9.3, we are done if we efficiently solve at most $\log|G|$ instances of orbit membership problem over a direct complement of $H\cap N$ in N with cumulative error at most $\epsilon/3$. As $K''K'\geq(\log|G|)^{c'+1}\log\frac{3}{\epsilon^2} > \log|G|^{c'}\frac{3\log|G|}{\epsilon}$, each individual instance can be solved with error $\epsilon/(3\log|G|)$ and hence the cumulative error is indeed at most $\epsilon/3$. We finish the procedure by putting the description of H into the output register and performing the usual cleanup.

As input, we need $K = 2K'K''$ copies of $\underline{\psi}$ (that is, $\underline{\psi}^{\otimes K}$). The running time is polynomial in $K\ell t$ and we have

$$K = 2K'K'' = 2(\log|G|\log\frac{3}{\epsilon})^{c'}(\log|G|(\log\frac{3}{\epsilon}+\log\log|G|+\log\log\frac{3}{\epsilon}))^d,$$

which is less than $(\log|G|)\log\frac{1}{\epsilon})^{c'+2d}$ if $|G|$ is large enough. Therefore

$$c = c(m) = c(m-1)+2d = c'+2d$$

is a good choice. □

9.5 Remarks

In the original paper [36] it is also shown that the constructive orbit membership problem – using polynomially many copies of the input states – can be solved in polynomial time in solvable groups of constant exponent and constant derived length. The proof is analogous to that of Theorem 9.1. We also remark that in [36], a self-reducibility theorem of a combination of the stabilizer problem and constructive orbit membership is proved based on a generalization of Theorem 9.16. The latter generalization is proved using an adapted version of a method of Watrous [92] originally designed for computing uniform superpositions over solvable black box groups. Here we restricted ourselves to the most important result (i.e., Theorem 9.1) and used a simplified approach to its proof.

In Section 9.3 we have shown that for any fixed prime power p^k, the problem Random Linear Disequations over the group $\mathbb{Z}_{p^k}^n$ can be solved in time which is polynomial in the rank n. Actually if we let the exponent p^k grow as well then our method runs in time polynomial in the rank n but exponential in the exponent p^k. Note that a brute force algorithm which takes a sample of size $O(knp^k \log p)$ (the kernels of that many random characters cover the whole group with high probability) and performs exhaustive search over all the the elements of $\mathbb{Z}_{p^k}^n$ runs in time $(p^{kn})^{O(1)}$ which is polynomial in the exponent p^k and exponential in n. It would be interesting to know if there exists a method which solves Random Linear Disequations in time polynomial in both n and p^k. Also, the method of Section 9.3 exploits heavily that the exponent of the group is a prime power. Existence of an algorithm for Random Linear Disequations in $\mathbb{Z}_m^n$ of complexity polynomial in n for fixed m having more than one prime divisors is an open problem, even in the smallest case $m = 6$.

Chapter 10

Efficient Testing of Groups

This chapter is based on parts of the paper [37], joint work with Katalin Friedl and Miklos Santha. Here we construct an efficient probabilistic algorithm that, given a finite set with a binary operation, tests if it is an abelian group. The complexity of the tester is polylogarithmic in the size of the set. The distance used is an analogue of the edit distance for strings. Previous testers used Hamming type distances and had superlinear query complexity. It is quite easy to construct a polylogarithmic quantum tester for abelian groups. Here we show that the power of quantum computers can be replaced by knowledge of a multiple of the order of elements.

Property testing deals with algorithms that can distinguish functions having some specific property from functions which are far away in a certain distance from functions having that property. Let $\mathcal{C}$ be a family of functions and let $\mathcal{F} \subseteq \mathcal{C}$ be the subset of functions possessing the property. Property testers are randomized procedures which receive as input an oracle for some function $f \in \mathcal{C}$, and after querying it at some number of points accept if $f \in \mathcal{F}$ and reject with high probability if f is far from $\mathcal{F}$. They are relaxations of the standard decision algorithms in the sense that they can give an arbitrary answer on functions which do not have the given property but are close to some function possessing it.

A property tester usually repeats some basic trials and its answer depends on the number of successful trials. Its complexity is mainly measured by the number of queries, which in turn is a function of the number of domain samples in the basic trials and the number of repetitions of the trials. This latter quantity – the query complexity – depends on the relations between the rejection probability of the basic trials and the distance from $\mathcal{F}$ of the function they reject. The computational complexity, the total number of operations performed be the tester is often but not always polynomially related to the query complexity. The main advantage of not requiring a correct answer on all the inputs is that for a large variety of problems we can have property testers which run in sublinear time, and thus do not even read the whole input data. Indeed, sublinear property testers have been constructed in recent years for example for problems in algebra, graph theory, geometry, string and set operations, optimization, probability theory and quantum computing. For surveys on property testing see for example [43, 84, 74].

Historically the first property testers were constructed for algebraic problems under the name of self-testers [15, 70]. Many of these testers dealt with group theoretical problems, and they were using a Hamming type distance. Two problems which are of particular interest for this chapter can be cast in the following general terms:

- Given a function from a group to another group, is it a homomorphism?
- Given a binary operation on a finite set, is it the multiplication operation for a group?

The first homomorphism tester for abelian groups was constructed by M. Blum, M. Luby and R. Rubinfeld [15]. To test if a function $f : G \to H$ is a homomorphism, where G and H are abelian groups, the following simple basic trial is repeated a constant number of times: pick two random elements $x, y \in G$ and verify that $f(x+y) = f(x)+f(y)$. This tester was extended to the case of non-abelian groups by M. Ben-Or, D. Coppersmith, M. Luby and R. Rubinfeld [12].

The obvious algorithm to decide if a binary operation on a finite set Γ is associative takes time $O(|\Gamma|^3)$. S. Rajagopalan and L. Schulman [81] gave an $O(|\Gamma|^2)$ randomized algorithm for this task, and they have also proved an $\Omega(|\Gamma|^2)$ lower bound. In the work which is the most closely related to ours, F. Ergün, S. Kannan, R. Kumar, R. Rubinfeld and M. Viswanathan [34] constructed a property tester for group multiplication which runs in time $O(|\Gamma|^{3/2}(\log|\Gamma|)^{O(1)})$. When the binary operation over Γ is guaranteed to be cancellative, the complexity of their tester is $O(|\Gamma|(\log|\Gamma|)^{O(1)})$. They measure the distance of two binary operations over the same finite set by the fraction of pairs of elements where the operations differ, and in their case the distance is undefined if the two operations act on different sets. The reason of the relatively high query complexity of their tester is that even in the cancellative case same basic tests are performed for every element of Γ.

Let Γ be a finite set. In this chapter we use the term *magma* for a finite set Γ equipped with a binary operation, often referred to as multiplication. Instead of the Hamming type distance discussed above we use a distance similar to the edit distance of strings. This will make it possible to correct magmas by changing the size of their ground sets. We define the edit distance so that it does not depend on the particular order of the elements.

A *table* of size k is a $k \times k$ matrix t whose element in row i and column j is denoted by t_{ij} for $1 \leq i, j \leq k$. We define three operations which transform a table t into a new table. An *exchange* operation at place (i, j) modifies the value t_{ij} and leaves unchanged the others. The cost of an exchange is 1. An *insert* operation at index i, where $1 \leq i \leq k+1$, transforms the table t into a table t' of size $k+1$ by inserting $2k+1$ elements to make a new row and a new column of index i, and by shifting down by one the rows of t of index at least i and by shifting to the right by one the columns of t of index at least i. The cost of an insert is $2k+1$. A *delete* operation at index i, where $1 \leq i \leq k$, transforms the table t into a table t' of size $k-1$ by deleting the ith row and the ith column and by shifting up by one the rows of index at least $i+1$ and by shifting to the left by one the columns of index at least $i+1$. The cost of a delete is $2k-1$. For two tables t and t' respectively of size k and k', the *(relative) edit distance* $\mathsf{Edit}(t, t')$ is the minimal cost of exchange, insert and delete operations, divided by $(\mathrm{Max}(k, k'))^2$, which transform t to t'.

Now we turn to the definition of the edit distance of magmas. Let Γ be a finite set of cardinality k equipped with multiplication $p : \Gamma \times \Gamma \to \Gamma$. We say that a table t of size k *represents* p if there exists a bijection $\pi : \{1, 2, \ldots, k\} \to \Gamma$ such that $t_{ij} = \pi^{-1}(p(\pi(i), \pi(j)))$ for $1 \leq i, j \leq k$. For two magmas S and S' with muliplications p resp. p' the *(relative) edit distance* of S and S', denoted by $\mathsf{Edit}(S, S')$, is defined as the minimum of $\mathsf{Edit}(t, t')$ where t represents p and t' represents p'.

Below we give a definition of property testers for magmas with respect to the edit type distance. Contrary to [34] our testers don't have to know the ground set of the magma to be tested, only an upper bound of its size and also suppose that random elements can be generated from the ground set. Let $\mathcal{S}$ stand for the family of finite magmas. Let $\mathcal{F} \subseteq \mathcal{S}$, and let $0 < \epsilon < 1$. An *ϵ–property tester* for $\mathcal{F}$ on $\mathcal{S}$ is a randomized algorithm T which can draw random elements from the ground set (independently and uniformly) and use an oracle for multiplying two elements, such that for every $S \in \mathcal{S}$, with ground set of size at most M and for every confidence parameter $0 < c < 1$:

- if $\mathsf{Edit}(S, \mathcal{F}) = 0$, then $\mathbf{Pr}\left[T^S(M, c)\ \mathtt{accepts}\right] = 1$,
- if $\mathsf{Edit}(S, \mathcal{F}) > \epsilon$, then $\mathbf{Pr}\left[T^S(M, c)\ \mathtt{rejects}\right] \geq 1 - c$,

where the probabilities are taken over the coin tosses of T. Here we assume that an upper bound on the size of the ground set (and the confidence parameter) are given to the algorithm. By T^S we mean that T is executed with the oracle for the multiplication of the specific S.

Our main result is the construction of a tester for the families of abelian groups whose exponents are divisors of a given number m. The complexity dependency of our tester on the size of the ground set is exponentially smaller than in the tester of [34]: the number of calls it makes to the oracle is only polylogarithmic in the size of the ground set.

Theorem 10.1. *Let $\mathcal{S}$ be the family of magmas and let $\mathcal{F}_m \subseteq \mathcal{S}$ be the family of finite abelian groups whose exponents divide m. For every $\epsilon > 0$, there exists an ϵ-tester T for $\mathcal{F}_m$ on $\mathcal{S}$ with distance* Edit *which for every $S \in \mathcal{S}$ with ground set Γ of size at most $M \geq m$, and for every confidence parameter c, uses $(\epsilon^{-1} \log M)^{O(1)} \log \frac{1}{c}$ calls to the oracle. With a large implicit constant, the total computational complexity of the tester is also $(\epsilon^{-1} \log M)^{O(1)} \log \frac{1}{c}$.*

In [34] associativity is tested in a clever, although rather direct way. Our approach is completely different. It originates in our idea of a quantum property tester for testing abelian groups. We briefly outline the main ingredients of the quantum tester. Let S be a magma with ground set Γ and a commutative multiplication $a \cdot b$.

First we pick sufficiently many random elements $\alpha_1, \ldots, \alpha_t$ from Γ so that if Γ is indeed a group then these elements generate Γ with high probability. We use the quantum algorithm of K. Cheung and M. Mosca [19] (as if Γ were a group) to find a basis $\gamma_1, \ldots, \gamma_s$ for the "subgroup" generated by $\alpha_1, \ldots, \alpha_t$, together with the orders $m_1, \ldots, m_s$ of the basis elements. If the algorithm fails we reject S. Otherwise let G denote the group $\mathbb{Z}_{m_1} \oplus \cdots \oplus Z_{m_s}$.

We have to assure that Γ is nearly isomorphic to G. For an element $\gamma \in \Gamma$, define its positive powers via fast exponentiation. Since we don't suppose that the binary operation is associative, we fix some method to parenthesize the terms. We define γ^0 as $\gamma_1^{m_1}$. Let $g : G \to \Gamma$ be the mapping given as $g(u_1, \ldots, u_s) = \gamma_1^{u_1} \cdot \gamma_2^{u_2} \cdots \gamma_s^{u_s}$, where again the multiplication is done according to some fixed way of parenthesizing. We test if g is almost a bijection between G and Γ as follows. We pick a random element $\gamma' \in \Gamma$, compute its "order" m' and solve the hidden subgroup problem for $G \oplus \mathbb{Z}_{m'}$ to find the "kernel" H of the map $(u, v) \mapsto g(u) \cdot \gamma'^{-v}$. We reject if H is not of order m'. Otherwise we see that the effective subgroup membership for γ' in the image of g can be solved. By repeating this procedure, we assure that the image of g is very close to Γ. If $g(u + v) = g(u) \cdot g(v)$ also holds with sufficiently high probability for random $u, v \in G$, then it is not difficult to show that g is close to an isomorphism between G and some group which approximates Γ.

We succeeded in extending the latter isomorphism test to a homomorphism test in the situation where a map similar to g above is far from being bijective. This extension makes it possible to substitute the quantum parts of the algorithm by the assumption on the knowledge of a multiple of the exponent.

For the classical probabilistic tester let S be a magma with ground set Γ, with binary operation $a \cdot b$ where a multiple m of the exponent is given. Again, we choose sufficiently many random elements $\gamma_1, \dots, \gamma_s$ from Γ. We can define 1 as the mth power of an arbitrary element of Γ. We consider the group $G = \mathbb{Z}_m^s$ and the map $g : G \to \Gamma$, where $g(u_1, \dots, u_s) = \gamma_1^{u_1} \cdot \gamma_2^{u_2} \cdots \gamma_s^{u_s}$. Here again products are defined according to some fixed way of parenthesizing and powers by fast exponentiation. If Γ is an abelian group then g is a homomorphism from G to Γ.

In Section 10.1 we consider maps from a not necessarily abelian group G to magmas. In Theorem 10.9 we establish that if $f : G \to \Gamma$ satisfies $f(uv) = f(u) \cdot f(v)$ with sufficiently high probability then it is close to a homomorphism from G to some group $\tilde{\Gamma}$.

We also would like to guarantee that the symmetric difference of $\tilde{\Gamma}$ and Γ is sufficiently small. We achieve this as follows. Let G_i be the subgroup $\mathbb{Z}_m^i \oplus \{0\}^{s-i}$ and assume that the restriction of g to G_i passes the homomorphism test with high probability. Then by Theorem 10.9, for $i = 1, \dots, s$ there exist groups $\tilde{\Gamma}_i$ such that the restriction of g to G_i is close to a homomorphism onto the group $\tilde{\Gamma}_i$. In Lemma 10.10 of Section 10.2 we will give further probabilistic conditions which guarantee that the size of $\tilde{\Gamma}_i$ grows exponentially with i with reasonably high probability until $\tilde{\Gamma}_i$ is close to Γ. The number of these conditions is polylogarithmic in the size of Γ which gives the bound on the query complexity of our tester.

In Section 10.3 the results proved in Theorem 10.9 and Lemma 10.10 are put together. We show that if a magma passes our tests with high probability then it is close to an abelian group, and Theorem 10.1 will follow immediately.

10.1 Approximate group homomorphisms

For the purpose of this section we fix a positive real number $\eta < 1/120$. Let Γ be a set equipped with a binary operation denoted by $a \cdot b$. Let G be a finite, not necessarily abelian group and let $f : G \longrightarrow \Gamma$ be a map from G to Γ such that

$$\mathop{\mathbf{Pr}}_{x,y \in G} [f(xy) = f(x) \cdot f(y)] \geq 1 - \eta. \tag{10.1}$$

If Γ is a group then a new function $\tilde{f}$ can be defined as

$$\tilde{f}(x) = \mathop{\mathrm{Maj}}_{y \in G} f(xy) \cdot f(y^{-1}),$$

and it can be shown that $\tilde{f}$ is a homomorphism close to f (see, for example, [74] for the abelian case). Here, if h is a function whose domain contains a finite set S, $\mathrm{Maj}_{y \in S}\, h(y)$ denotes the value of f taken most frequently on elements of S. If there are more than one most frequent values then we take the symbol "undefined".

Unfortunately the approach above does not work directly if Γ is not associative. However, we can construct a congruence relation on G using similar majority arguments. Recall that a congruence on G is an equivalence relation respecting the group operations or, equivalently, an equivalence relation where the classes are the cosets of a normal subgroup K. We shall identify a big part of the factor group G/K with a subset of Γ in a way so that f will be close to the natural homomorphism $G \to G/K$.

We define the binary relation $E \subseteq G \times G$ by

$$E(x,y) \Leftrightarrow \mathop{\mathbf{Pr}}_{u\in G}\left[f(xu) = f(yu)\right] \geq 1/2.$$

First we prove that the majority involved in the definition above is actually overwhelming.

Lemma 10.2. *Let $x, y \in G$. Then*

$$\mathop{\mathbf{Pr}}_{u\in G}\left[f(xu) = f(yu)\right] \geq 1 - 4\eta \quad \text{or} \quad \mathop{\mathbf{Pr}}_{u\in G}\left[f(xu) \neq f(yu)\right] \geq 1 - 4\eta.$$

Proof. Condition (10.1) implies that

$$\mathop{\mathbf{Pr}}_{u,w\in G}\left[f(xuw) = f(xu)\cdot f(w)\right] \geq 1 - \eta,$$
$$\mathop{\mathbf{Pr}}_{u,w\in G}\left[f(yuw) = f(yu)\cdot f(w)\right] \geq 1 - \eta.$$

Therefore

$$\mathop{\mathbf{Pr}}_{u,w\in G}\left[f(xu) = f(yu) \Rightarrow f(xuw) = f(yuw)\right] \geq 1 - 2\eta.$$

Rewriting uw as v we obtain

$$\mathop{\mathbf{Pr}}_{u,v\in G}\left[f(xu) = f(yu) \Rightarrow f(xv) = f(yv)\right] \geq 1 - 2\eta,$$

whence, by symmetry,

$$\mathop{\mathbf{Pr}}_{u,v\in G}\left[f(xu) = f(yu) \Leftrightarrow f(xv) = f(yv)\right] \geq 1 - 4\eta.$$

Rewriting the left hand side gives

$$\mathop{\mathbf{Pr}}_{u\in G}\left[f(xu) = f(yu)\right]^2 + \mathop{\mathbf{Pr}}_{u\in G}\left[f(xu) \neq f(yu)\right]^2 \geq 1 - 4\eta.$$

Let p stand for the smaller of $\mathbf{Pr}_{u\in G}\left[f(xu) = f(yu)\right]$ and $\mathbf{Pr}_{u\in G}\left[f(xu) \neq f(yu)\right]$. (As $1-4\eta > \frac{1}{2}$, they cannot be equal.) Then the last inequality can be written as $p^2 + (1-p)^2 \geq 1 - 4\eta$, whence $p(1-p) \leq 2\eta$. Assume by contradiction that $p > 4\eta$. Then, monotonicity of $p(1-p)$ for $p < \frac{1}{2}$ implies $4\eta(1-4\eta) < 2\eta$, or equivalently, $\eta(1-8\eta) < 0$. This is impossible as $0 < \eta < \frac{1}{120}$ and the statement follows. □

Lemma 10.3. *E is an equivalence relation on G.*

Proof. The facts that E is reflexive and symmetric are immediate consequences of the definition. For transitivity assume that $E(x,y)$ and $E(y,z)$. Then Lemma 10.2 implies that $\mathbf{Pr}_{u\in G}\left[f(xu) = f(zu)\right] \geq 1-8\eta \geq 1/2$. Therefore E is transitive, so it is an equivalence relation. □

We say that an element $x \in G$ is *well-behaving* if

$$\mathbf{Pr}_{u\in G}\left[f(xu) = f(x)\cdot f(u)\right] \geq 4/5, \tag{10.2}$$
$$\text{and} \quad \mathbf{Pr}_{u\in G}\left[(f(x)\cdot f(u))\cdot f(u^{-1}) = f(x)\right] \geq 4/5. \tag{10.3}$$

Next we show that on well-behaving elements the equivalence classes correspond to different values of f and also, that most of the elements are well-behaving.

Lemma 10.4. *Let $x_1, x_2 \in G$ well-behaving elements. Then $E(x_1, x_2)$ iff $f(x_1) = f(x_2)$.*

Proof. If $f(x_1) = f(x_2)$ then by (10.2) we have

$$\mathop{\mathbf{Pr}}_{u\in G}[f(x_1u) = f(x_2u)] \quad \geq \quad \mathop{\mathbf{Pr}}_{u\in G}[f(x_1u) = f(x_1)\cdot f(u) \text{ and } f(x_2u) = f(x_2)\cdot f(u)].$$

The latter probability is at least 3/5 and so $E(x_1, x_2)$ holds. On the other hand, when $E(x_1, x_2)$ then, by Lemma 10.2, $\mathbf{Pr}_{u\in G}[f(x_1u) = f(x_2u)] \geq 1 - 4\eta$. By assumptions (10.3) and (10.2), $\mathbf{Pr}_{u\in G}[f(x_i) = f(x_iu)\cdot f(u^{-1})] \geq 3/5$. Therefore

$$\mathop{\mathbf{Pr}}_{u\in G}[f(x_1) = f(x_2)] \geq 1 - 4\eta - 4/5 > 0.$$

□

Lemma 10.5. $\mathbf{Pr}_{x\in G}[x \text{ is not well-behaving }] \leq 15\eta$.

Proof. From (10.1) follows that the probability that (10.2) does not hold for an x is at most 5η. Regarding the other condition, notice that (10.1) implies the inequality $\mathbf{Pr}_{x,u\in G}[(f(x)\cdot f(u))\cdot f(u^{-1}) = f(x)] > 1 - 2\eta$, therefore the probability that an x violates (10.3) is at most 10η. □

Lemma 10.6. *There is a subgroup K of G such that $E(x, y)$ if and only if $Kx = Ky$.*

Proof. Let K be the equivalence class of the identity element. Then from the definition of E we obtain that

$$\begin{aligned} E(x,y) \Leftrightarrow \mathop{\mathbf{Pr}}_{u\in G}[f(xu) = f(yu)] \geq 1/2 \Leftrightarrow \\ \mathop{\mathbf{Pr}}_{u\in G}\left[f(xy^{-1}yu) = f(yu)\right] \geq 1/2 \Leftrightarrow \\ \mathop{\mathbf{Pr}}_{w\in G}\left[f(xy^{-1}w) = f(w)\right] \geq 1/2 \Leftrightarrow \\ E(xy^{-1}, 1) \Leftrightarrow xy^{-1} \in K. \end{aligned}$$

If $x, y \in K$, then by definition, $E(1, x)$ and $E(1, y)$ hold. By transitivity $E(x, y)$ follows, therefore $xy^{-1} \in K$. □

Lemma 10.7. *Let K denote the subgroup of G provided by Lemma 10.6. Then K is a normal subgroup of G.*

Proof. Define the relation E' by

$$E'(x,y) \Leftrightarrow \mathop{\mathbf{Pr}}_{u\in G}[f(ux) = f(uy)] \geq 1/2.$$

For E' all the preceding claims are true if we reverse the order of multiplication. In particular, there exists a subgroup H of G such that $E'(x, y)$ if and only if $xH = yH$. We say that an element $x \in G$ is *well-behaving on both sides* if x is well-behaving with respect to both orders of multiplication,

$$\begin{aligned} \mathop{\mathbf{Pr}}_{u\in G}[f(xu) = f(x)\cdot f(u)] \geq 4/5, \\ \mathop{\mathbf{Pr}}_{u\in G}[f(ux) = f(u)\cdot f(x)] \geq 4/5, \\ \mathop{\mathbf{Pr}}_{u\in G}\left[(f(x)\cdot f(u))\cdot f(u^{-1}) = f(x)\right] \geq 4/5, \\ \text{and } \mathop{\mathbf{Pr}}_{u\in G}\left[f(u^{-1}\cdot(f(u)\cdot f(x)) = f(x)\right] \geq 4/5. \end{aligned}$$

Application of Lemma 10.5 to both orders of multiplication gives

$$\mathop{\mathbf{Pr}}_{x\in G}[x \text{ is well-behaving on both sides }] \geq 1-30\eta.$$

Let $y \in G$ be an element such that

$$\mathop{\mathbf{Pr}}_{x\in Ky}[x \text{ is well-behaving on both sides }] > 1/2$$

holds. Notice that more than half of the y's are like this because $\eta < 120$.

Let $x \in Ky$ a well-behaving element on both sides. If $x' \in Ky$ is also a well-behaving element on both sides then, by Lemma 10.4, $f(x) = f(x')$, and again by Lemma 10.4 applied to the reverse order of multiplication, $xH = x'H$, so $x' \in xH$. Therefore, $|Ky \cap xH| > |K|/2$.

Consider the set $x^{-1}Ky = y^{-1}(yx^{-1}K)y$. Notice that $yx^{-1} \in K$ because of the choice $x \in Ky$. This ensures that $x^{-1}Ky = y^{-1}Ky$ is a subgroup. The facts that $y^{-1}Ky \cap H$ is a subgroup of $y^{-1}Ky$ and $|y^{-1}Ky\cap H| = |x^{-1}Ky\cap H| = |x(x^{-1}Ky\cap H)| = |Ky\cap xH| > |K|/2$ imply that $y^{-1}Ky \leq H$ from which $|K| \leq |H|$ also follows. Similarly, we can obtain that $|H| \leq |K|$, so $|K| = |H|$ and $y^{-1}Ky = H$ must hold for every y such that

$$\mathop{\mathbf{Pr}}_{x\in Ky}[x \text{ is well-behaving from both sides }] > 1/2.$$

Therefore $y^{-1}Ky = H$ holds for more than half of the elements $y \in G$. When $y^{-1}Ky = H$ and $z^{-1}Kz = H$ then yz^{-1} normalizes K, i.e., $(yz^{-1})^{-1}Kyz^{-1} = K$. It follows that the size of the normalizer group of K, $|N_G(K)| > |G|/2$, so $N_G(K) = G$ and the subgroup K is a normal subgroup in G. □

We say that an element $x \in G$ is *good* if it is well-behaving and $f(x) = \mathrm{Maj}_{y:E(x,y)}(f(y))$, and *bad* otherwise. A coset Kx is *good* if it contains a good element.

Lemma 10.8. $\mathbf{Pr}_{x\in G}[x \text{ is bad }] \leq 30\eta$

Proof. We show that

$$\mathop{\mathbf{Pr}}_{x\in G}[x \text{ is bad and well-behaving }] \leq \mathop{\mathbf{Pr}}_{x\in G}[x \text{ is not well-behaving }].$$

In fact we can prove the stronger statement, that this inequality holds when the probability is considered in an equivalence class. We know that the function f is constant on the well-behaving elements of an equivalence class (by Lemma 10.4). Since the bad well-behaving elements do not form a majority in the class, the number of non well-behaving elements must be greater. □

Theorem 10.9. *Let $\eta < 1/120$ and assume that the inequality*

$$\mathop{\mathbf{Pr}}_{x,y\in G}[f(xy) = f(x)\cdot f(y)] > 1-\eta$$

holds. Then there exists a group $\tilde{\Gamma}$ with multiplication $*$ and a homomorphism $\tilde{f} : G \to \tilde{\Gamma}$ such that

(i) $|\tilde{\Gamma} \setminus \Gamma| \leq 30\eta|\tilde{\Gamma}|$,

(ii) $\mathbf{Pr}_{\gamma_1,\gamma_2\in\tilde{\Gamma}}\left[\gamma_1 * \gamma_2 \neq \gamma_1 \cdot \gamma_2\right] \leq 91\eta$,

(iii) $\mathbf{Pr}_{x\in G}\left[\tilde{f}(x) \neq f(x)\right] \leq 30\eta$.

Proof. Let $\tilde{\Gamma} = G/K$ and let $\tilde{f} : G \to \tilde{\Gamma}$ be the natural homomorphism. Consider a good coset $Kx \in G/K$. We identify Kx with $f(y)$ where y is a good element of Kx. This identification is legitimate since the function f is constant on the good elements of Kx, and it ensures that when x is a good element of G then $\tilde{f}(x) = f(x)$. Now we prove the three claims.

(i) For every $x \in G$ observe that $\tilde{f}(x) \notin \Gamma \Leftrightarrow Kx$ is bad. If Kx is bad, then each of its elements is bad, so Lemma 10.8 implies the statement.

(ii) Let $\gamma_1, \gamma_2 \in \tilde{\Gamma}$. Then there exist $x, y \in G$, such that $\tilde{f}(x) = \gamma_1$ and $\tilde{f}(y) = \gamma_2$. Also $\gamma_1 * \gamma_2 = \tilde{f}(x) * \tilde{f}(y) = \tilde{f}(xy)$.

When $\gamma_1 \notin \Gamma$ or $\gamma_2 \notin \Gamma$ then $\gamma_1 \cdot \gamma_2$ is not defined. This happens with probability at most 60η. The probability that xy is bad is at most 30η. If xy is good then $\tilde{f}(xy) = f(xy)$. The probability in this case that $f(xy) \neq f(x) \cdot f(y)$ is at most η.

(iii) If $\tilde{f}(x) \neq f(x)$ then x is bad and the statement follows from Lemma 10.8. □

10.2 Growing subgroups

We keep the assumptions and notation introduced in the preceding section. Here we prove a lemma which will be used in Theorem 10.14. We establish further probabilistic conditions under which, if we apply Theorem 10.9 to groups $H < G$ in the context described in the outline given in the introductory part of this chapter, i.e. H is constructed from random elements of Γ and G is built by adding further random elements, yielding respectively groups $\tilde{\Pi}$ and $\tilde{\Gamma}$, then $\tilde{\Gamma}$ will be at least twice as large as $\tilde{\Pi}$ with reasonable probability unless $\tilde{\Pi}$ is close to Γ.

Lemma 10.10. *Let $\eta < 1/120$, let $f : G \to \Gamma$ be a function, let H be a subgroup of G and let z be an arbitrary element of G. Assume that*

$$\mathbf{Pr}_{u,v\in H}\left[f(uv) = f(u) \cdot f(v)\right] \geq 1 - \eta. \tag{10.4}$$

Let $\tilde{\Pi}$ be the group provided by Theorem 10.9 for H. Then there exists a subset $B = B(H)$ of Γ such that $|B| \leq |\tilde{\Pi}|$, and the following property holds: Assume also that f satisfies (10.1) and

$$\mathbf{Pr}_{u\in H,v\in G}\left[f(uv) = f(u) \cdot f(v)\right] \geq 1 - \eta, \tag{10.5}$$

$$\mathbf{Pr}_{u\in H,v\in G}\left[f(uv) \cdot f(v^{-1}) = f(u)\right] \geq 1 - \eta, \tag{10.6}$$

$$\mathbf{Pr}_{u\in H,v\in G}\left[f(zuv) \cdot f(v^{-1}) = f(xu)\right] \geq 1 - \eta, \tag{10.7}$$

$$\mathbf{Pr}_{u\in H}\left[f(zu) \cdot f(u^{-1}) = f(x)\right] \geq 1 - \eta. \tag{10.8}$$

Then either $|\tilde{\Gamma}| = |\tilde{\Pi}|$, or $|\tilde{\Gamma}| \geq 2|\tilde{\Pi}|$, where $\tilde{\Gamma}$ is the group provided by Theorem 10.9 for G. Moreover, if $|\tilde{\Gamma}| = |\tilde{\Pi}|$ then $f(z) \in B$.

Proof. We define the relation E_H on H by

$$E_H(x, y) \iff \mathop{\mathbf{Pr}}_{u\in H}\left[f(xu) = f(yu)\right] \geq 1/2.$$

For every $x \in H$, let $\beta(x) = \mathrm{Maj}_{u \in H} \left(f(xu) \cdot f(u^{-1})\right)$. We say that x is a *majority element in H* if

$$\mathop{\mathbf{Pr}}_{u \in H} \left[\beta(x) = f(xu) \cdot f(u^{-1})\right] > 1/2 + 4\eta.$$

Notice that a well-behaving element x is always a majority element in H. We define the set

$$B = \{\beta(x) \ : \ x \ \text{ is a majority element in } \ H\}.$$

Claim 10.11. $|B| \leq |\tilde{\Pi}|$.

Proof. It is sufficient to prove that the size of B is at most the number of equivalence classes of E_H. Let majority elements x, y belong to the same equivalence class of E_H. Then $\beta(x) = \beta(y)$ because

$$\mathop{\mathbf{Pr}}_{u \in H} \left[f(xu) \cdot f(u^{-1}) = f(yu) \cdot f(u^{-1})\right] > 1 - 4\eta.$$

□

Claim 10.12. *For $y_1, y_2 \in H$, $E_H(y_1, y_2) \iff E_G(y_1, y_2)$.*

Proof. Assume that $E_H(y_1, y_2)$. According to Lemma 10.2, $\mathbf{Pr}_{u \in H} [f(y_1 u) = f(y_2 u)] \geq 1 - 4\eta$. From (10.5) we infer that $\mathbf{Pr}_{u \in H, v \in G} [f(y_i uv) = f(y_i u) \cdot f(v)] \geq 1 - \eta$ because $H = y_i H$. So

$$\mathop{\mathbf{Pr}}_{w \in G} [f(y_1 w) = f(y_2 w)] \ = \ \mathop{\mathbf{Pr}}_{u \in H, v \in G} [f(y_1 uv) = f(y_2 uv)] \ \geq \ 1 - 6\eta \ > \ 1/2, \quad (10.9)$$

which means that $E_G(x, y)$.

To see the reverse implication, assume that $E_G(y_1, y_2)$. Then by Lemma 10.2

$$\mathop{\mathbf{Pr}}_{u \in H, v \in G} [f(y_1 uv) = f(y_2 uv)] \ = \ \mathop{\mathbf{Pr}}_{w \in G} [f(y_1 w) = f(y_2 (w)] \ \geq \ 1 - 4\eta. \quad (10.10)$$

By (10.6), $\mathbf{Pr}_{u \in G, v \in H} [f(y_i u) = f(y_i uv) \cdot f(v^{-1})] \geq 1 - \eta$ holds, therefore we have $\mathbf{Pr}_{u \in H} [f(y_1 u) = f(y_2 u)] \geq 1 - 6\eta > 1/2$, and so $E_H(y_1, y_2)$. □

Since the elements of $\tilde{\Pi}$ and $\tilde{\Gamma}$ are respectively the equivalence classes of E_H and E_G, it follows that $\tilde{\Gamma}$ has a subgroup isomorphic to $\tilde{\Pi}$, and therefore either $|\tilde{\Gamma}| = |\tilde{\Pi}|$ or $|\tilde{\Gamma}| \geq 2|\tilde{\Pi}|$.

Claim 10.13. *Let $y \in H$ such that $E_G(z, y)$. Then $f(z) = \beta(y)$ and y is a majority element in H.*

Proof. First observe that if u is a random element of H and v is a random element in G then uv is a random element in G. By Lemma 10.2,

$$\mathop{\mathbf{Pr}}_{u \in H, v \in G} [f(zuv) = f(yuv)] \geq 1 - 4\eta.$$

Therefore

$$\mathop{\mathbf{Pr}}_{u \in H, v \in G} \left[f(zuv) \cdot f(v^{-1}) = f(yuv) \cdot f(v^{-1})\right] \geq 1 - 4\eta.$$

From (10.7) and (10.6) follows that $\mathbf{Pr}_{u \in H} [f(zu) = f(yu)] \geq 1 - 6\eta$ and (10.8) implies that $\mathbf{Pr}_{u \in H} [f(z) = f(yu) \cdot f(u^{-1})] \geq 1 - 7\eta$. Because $1/2 + 4\eta < 1 - 7\eta$, it follows that y is a majority element in H and $f(z) = \beta(y)$. □

To finish the proof of Lemma 10.10 let us suppose that $f(z) \notin B$. Then by Claim 10.13, $E_G(z, y)$ does not hold for any $y \in H$. Therefore E_G has more equivalence classes than E_H and thus $|\tilde{\Gamma}| > |\tilde{\Pi}|$. □

10.3 The tester

Let S be a magma whose ground set is Γ. We denote by $\cdot$ the multiplication of S. For any $\gamma \in \Gamma$ and integer $0 < e \leq m$ we define the positive powers of γ as follows: $\gamma^1 = \gamma, \gamma^{2^i} = \gamma^{2^{i-1}} \cdot \gamma^{2^{i-1}}$ for $i > 0$, and for $e = \sum_{i=0}^{j} b_i 2^i$ where $b_i \in \{0, 1\}$, $\gamma^e = (\ldots(\gamma^{b_0 2^0} \cdot \gamma^{b_1 2^1}) \ldots) \cdot \gamma^{b_j 2^j}$, where multiplication by a factor of the form γ^0 is the identity map. We fix an arbitrary element a of Γ and define 1 as a^m. Then for every $\gamma \in \Gamma$, $\gamma^0 = 1$ and for an arbitrary integer e', $\gamma^{e'} = \gamma^e$, where e' is the smallest nonnegative number congruent with e' modulo m.

Let k and ℓ be integers which will be fixed later. We put $G = \mathbb{Z}_m^{k\ell}$ and for $i = 1, \ldots, k$, we define the subgroups $G^{(i)} = \bigoplus_{t=1}^{i} \bigoplus_{j=1}^{\ell} \mathbb{Z}_m \oplus \bigoplus_{t=i\ell+1}^{k\ell} \{0\}$ of G. For $i = 1, \ldots, k$ and $j = 1, \ldots, \ell$ put $x_{ij} = (0, \ldots, 0, 1, 0 \ldots, 0) \in G$ where the 1 is in the $(i-1)\ell+j$th coordinate.

Put $\bar{\Gamma} = \Gamma^{k\ell}$. Let $\bar{\gamma} = (\gamma_{11}, \ldots, \gamma_{1\ell}, \ldots, \gamma_{k1}, \ldots, \gamma_{k\ell}) \in \bar{\Gamma}$. For every such $\bar{\gamma} \in \bar{\Gamma}$ we define the functions $g_{ij} : \bar{\Gamma} \times \mathbb{Z}_m \to \Gamma$ by

$$g_{ij}(\bar{\gamma}, e) = \gamma_{ij}^e$$

and $g : \bar{\Gamma} \times G \to \Gamma$ by

$$g(\bar{\gamma}, e_{11}, \ldots, e_{k\ell}) = (\ldots(g_{11}(\bar{\gamma}, e_{11}) \cdot g_{12}(\bar{\gamma}, e_{12})) \ldots) \cdot g_{k\ell}(\bar{\gamma}, e_{k\ell}).$$

Theorem 10.14. *Let S be as above and let Γ be of size M. Let $0 < \epsilon < 1$. Set $k = \lceil \log_2 M \rceil$, $\ell = \lceil 9(\ln(2k))/\epsilon \rceil = \theta(\log\log M \cdot \frac{1}{\epsilon})$ and $\eta = \epsilon/300$. If the inequalities*

$$\mathop{\mathbf{Pr}}_{u,v \in G} [g(\bar{\gamma}, u+v) = g(\bar{\gamma}, u) \cdot g(\bar{\gamma}, v)] \geq 1 - \eta, \tag{10.11}$$

and for $i = 2, \ldots, k$

$$\mathop{\mathbf{Pr}}_{u,v \in G^{(i-1)}} [g(\bar{\gamma}, u+v) = g(\bar{\gamma}, u) \cdot g(\bar{\gamma}, v)] \geq 1 - \eta, \tag{10.12}$$

$$\mathop{\mathbf{Pr}}_{u \in G^{(i-1)}, v \in G^{(i)}} [g(\bar{\gamma}, u+v) = g(\bar{\gamma}, u) \cdot g(\bar{\gamma}, v)] \geq 1 - \eta, \tag{10.13}$$

$$\mathop{\mathbf{Pr}}_{u \in G^{(i-1)}, v \in G^{(i)}} [g(\bar{\gamma}, u+v) \cdot g(\bar{\gamma}, -v) = g(\bar{\gamma}, u)] \geq 1 - \eta, \tag{10.14}$$

and for $i = 2, \ldots, k$ and $j = 1, \ldots, \ell$

$$\mathop{\mathbf{Pr}}_{u \in G^{(i-1)}, v \in G^{(i)}} [g(\bar{\gamma}, x_{ij} + u + v) \cdot g(\bar{\gamma}, -v) = g(\bar{\gamma}, x_{ij} + u)] \geq 1 - \eta, \tag{10.15}$$

$$\mathop{\mathbf{Pr}}_{u \in G^{(i-1)}} [g(\bar{\gamma}, x_{ij} + u) \cdot g(\bar{\gamma}, -u) = g(\bar{\gamma}, x_{ij}))] \geq 1 - \eta', \tag{10.16}$$

are simultaneously satisfied by a random $\bar{\gamma} \in \bar{\Gamma}$ with probability at least 1/2 then for every such $\bar{\gamma}$ there exists an abelian group $A(\bar{\gamma})$ with operation $\circ$ which satisfies the following properties:

(i) $|A(\bar{\gamma}) \setminus \Gamma| \leq \epsilon |A(\bar{\gamma})|/9$,

(ii) $\mathbf{Pr}_{a,b \in \Gamma \cap A(\bar{\gamma})} [a \cdot b \neq a \circ b] \leq \epsilon/3$.

Moreover, there exists $\bar{\gamma} \in \bar{\Gamma}$ such that

(iii) $|\Gamma \setminus A(\bar\gamma)| \leq 2\epsilon|\Gamma|/9$,

(iv) $\mathsf{Edit}(S, A(\bar\gamma)) \leq \epsilon$.

Proof. Let $\bar\gamma$ satisfy (10.11)–(10.16). We apply Theorem 10.9 with $f = g(\bar\gamma, \cdot)$ and define $A(\bar\gamma)$ as the group $\tilde\Gamma$ provided by the result. Then (i) and (ii) of Theorem 10.9 imply respectively (i) and (ii).

We now turn to the proof of (iii). In fact we will prove that $|A(\bar\gamma)| \geq (1 - \epsilon/9))|\Gamma|$ which together with (i) implies (iii). For an integer $1 \leq t \leq k$ let $\mathcal{E}_t$ stand for the event that $\bar\gamma$ satisfies (10.12)–(10.16) for $1 \leq i \leq t$ and (10.12) also holds for $i = t+1$. Since $G^{(t)}$ satisfies (10.11) we can apply Theorem 10.9 with $G^{(t)}$ in place of G. Let $A_t(\bar\gamma)$ be the group provided by Theorem 10.9. Also, if $\mathcal{E}_{t+1}$ holds then inequalities (10.4)–(10.8) hold for $H = G^{(t)}, G = G^{(t+1)}$. Therefore Lemma 10.10 implies that either $|A_t(\bar\gamma)| = |A_{t+1}(\bar\gamma)|$ or $|A_{t+1}(\bar\gamma)| \geq 2 \cdot |A_t(\bar\gamma)|$, and, if the equality holds, then $\gamma_{t+1,1}, \ldots, \gamma_{t+1,\ell}$ are in B, a set of size at most $|A_t(\bar\gamma)|$.

For $1 \leq t \leq k$, let $\mathcal{E}'_t$ stand for the event that both $\mathcal{E}_t$ and $|A_t(\bar\gamma)| < (1 - \epsilon/9)|\Gamma|$ hold. Obviously $\mathcal{E}'_t$ implies $\mathcal{E}'_{t-1}$ for every $t > 1$. Let us now make the indirect assumption that $\mathcal{E}'_k = \mathcal{E}_k$. This means that if $\mathcal{E}_k$ holds for a $\bar\gamma$ then $|A_k(\bar\gamma)| < (1 - \epsilon/9)|\Gamma|$ also holds. Notice that for every $t < k$ we have $\mathbf{Pr}_{\bar\gamma}\left[\,|A_{t+1}(\bar\gamma)| = |A_t(\bar\gamma)| \;\mid \mathcal{E}_k\right] \leq \mathbf{Pr}_{\bar\gamma}\left[\,|A_{t+1}(\bar\gamma)| = |A_t(\bar\gamma)| \;\mid \mathcal{E}'_t\right] / \mathbf{Pr}_{\bar\gamma}\left[\mathcal{E}_k\right]$, because $\mathcal{E}'_k = \mathcal{E}_k$ implies $\mathcal{E}'_t$. We know that $|A_{t+1}(\bar\gamma)| = |A_t(\bar\gamma)|$ only can happen when $\gamma_{t+1,1}, \ldots, \gamma_{t+1,\ell}$ comes form a subset of size at most $|A_t(\bar\gamma)| \leq (1 - \epsilon/9)|\Gamma|$. Using the assumption that $\mathbf{Pr}_{\bar\gamma}\left[\mathcal{E}_k\right] \geq 1/2$ the right hand side can be estimated as

$$\mathop{\mathbf{Pr}}_{\bar\gamma}\left[\,|A_{t+1}(\bar\gamma)| = |A_t(\bar\gamma)| \;\mid \mathcal{E}'_t\right] / \mathop{\mathbf{Pr}}_{\bar\gamma}\left[\mathcal{E}_k\right] < 2(1 - \epsilon/9)^{\ell}$$

Clearly, if $|A_k(\bar\gamma)| < |\Gamma|$ then for some $t < k$ the equality $|A_{t+1}(\bar\gamma)| = |A_t(\bar\gamma)|$ must hold. Therefore,

$$\mathop{\mathbf{Pr}}_{\bar\gamma}\left[|A_k(\bar\gamma)| < |\Gamma| \mid \mathcal{E}_k\right] < 2k(1 - \epsilon/9)^{\ell}.$$

By the indirect assumption, the left hand side is 1 and because of the choice of k and ℓ the right hand side is less than 1, which gives a contradiction.

Finally we prove (iv) by evaluating the relative cost of transforming $A(\bar\gamma)$ to S. The cost of deletions by (i) is at most $2\epsilon/9$. The cost of exchange operations by (ii) is at most $\epsilon/3$. The cost of insertions by (iii) is at most $4\epsilon/9$, which finishes the proof. □

We are ready to finish the proof of our main result.

Proof of Theorem 10.1. We will set parameters k and ℓ as in Theorem 10.14, and use the notations before that theorem. The goal of the tester T will be to statistically verify if the following inequalities hold simultaneously with $\eta' = \eta/16k\ell$:

$$\mathop{\mathbf{Pr}}_{\bar\gamma\in\bar\Gamma,\; u,v\in G}\left[g(\bar\gamma, u+v) = g(\bar\gamma, u)\cdot g(\bar\gamma, v)\right] \geq 1 - \eta', \tag{10.17}$$

and for $i = 2, \ldots, k$

$$\mathop{\mathbf{Pr}}_{\bar\gamma\in\bar\Gamma,\; u,v\in G^{(i-1)}}\left[g(\bar\gamma, u+v) = g(\bar\gamma, u)\cdot g(\bar\gamma, v)\right] \geq 1 - \eta', \tag{10.18}$$

$$\mathop{\mathbf{Pr}}_{\bar\gamma\in\bar\Gamma,\; u\in G^{(i-1)}, v\in G^{(i)}}\left[g(\bar\gamma, u+v) = g(\bar\gamma, u)\cdot g(\bar\gamma, v)\right] \geq 1 - \eta', \tag{10.19}$$

$$\mathop{\mathbf{Pr}}_{\bar{\gamma}\in\bar{\Gamma},\ u\in G^{(i-1)}, v\in G^{(i)}} [g(\bar{\gamma}, u+v)\cdot g(\bar{\gamma}, -v) = g(\bar{\gamma}, u)] \qquad \geq \qquad 1-\eta', \quad (10.20)$$

and for $i = 2, \dots, k$ and $j = 1, \dots, \ell$

$$\mathop{\mathbf{Pr}}_{\bar{\gamma}\in\bar{\Gamma},\ u\in G^{(i-1)}, v\in G^{(i)}} [g(\bar{\gamma}, x_{ij}+u+v)\cdot g(\bar{\gamma}, -v) = g(\bar{\gamma}, x_{ij}+u)] \qquad \geq \qquad 1-\eta', \quad (10.21)$$

$$\mathop{\mathbf{Pr}}_{\bar{\gamma}\in\bar{\Gamma},\ u\in G^{(i-1)}} [g(\bar{\gamma}, x_{ij}+u)\cdot g(\bar{\gamma}, -u) = g(\bar{\gamma}, x_{ij}))] \geq 1-\eta'. \quad (10.22)$$

For all inequalities the tester will approximate the probabilities of the events on the left hand side by $O((1/\eta')\log(1/c))$ independent trials. It accepts if the frequency of the failure of every event is less than $\eta'/2$. The evaluation of g at any point requires $O(k\ell\log m)$ calls to the oracle. Therefore the total number of oracle calls is $O(k^2\ell^2\log M(1/\eta')\log(1/c))$ that, by our choice of the parameters $k = O(\log M)$, $\ell = O(\epsilon^{-1}\log\log M)$ and $\eta' = O(\epsilon k^{-1}\ell^{-1})$, is

$$O(\log^3 M\log m(\log\log M)^3\epsilon^{-4}\log(1/c)).$$

Clearly if S is an abelian group of exponent dividing m, there will be no failure, and the tester always accepts. Let us now suppose that $\mathsf{Edit}(S,\mathcal{F}) > \epsilon$. Then from (iv) of Theorem 10.14 it follows that for a random $\bar{\gamma}\in\bar{\Gamma}$ with probability at least 1/2 at least one of the inequalities (10.11)–(10.16) does not hold. Since the number of inequalities is less than $8kl$ and $\eta' = \eta/16kl$, Markov's inequality implies that at least one of the inequalities (10.17)–(10.22) does not hold either. It follows from standard Chernoff bound arguments [1] that T will find with probability at least $1-c$ a frequency of failure greater than $\eta'/2$ for the corresponding event, and therefore will reject with at least that probability. □

Remarks

As already mentioned, our result can be interpreted as in testing abelian groups, the quantum computers can be substituted by the knowledge of a multiple of the exponent of the group. In [37], an even stronger result is given: the power of quantum computers can be substituted by the assumption of the ability of taking inverses of elements.

The reason for that we considered edit distance instead of a Hamming type distance is the following. If we extend the multiplication table of a group with a few fake rows and columns then a polylogarithmic time quantum or classical randomized algorithm has a negligible chance to hit the fake part of the table and hence with high probability, it will recognize the table as a group multiplication table, unless its exact size is explicitly given.

There is heuristic evidence that the knowledge of the exact size of the ground set cannot help if we want to use a classical randomized algorithm. Namely, consider two n-bit primes $p_1 < p_2$ such that $p_2 - p_1 = O(\log n)$. We consider two tables on a ground set of size $p_1p_2^2$. The first one corresponds to the group $\mathbb{Z}_{p_1}\oplus\mathbb{Z}_{p_2}\oplus\mathbb{Z}_{p_2}$ while the other is created from the table of the group $\mathbb{Z}_{p_1}\oplus\mathbb{Z}_{p_1}\oplus\mathbb{Z}_{p_2}$ padded by fake rows and columns. Notice that in the relative Hamming distance, the second table is far away from any group multiplication table. A construction similar to the one in [7] shows that the two groups above, given as black box groups, can be distinguished by a classical randomized randomized in time polynomial in n only with exponentially small success probability. Based on this obstacle we expect that it is very difficult to distinguish classically the two tables corresponding to the two groups and hence a classical test for abelian groups with respect to the Hamming type distance is difficult as well.

Bibliography

[1] N. Alon, J. Spencer, The probabilistic method, *John Wiley & Sons, 1992.*

[2] E. F. Assmus, Jr, J. D. Key, Polynomial Codes and Finite Geometries, *V. S. Pless, W. C. Huffman, (eds), Handbook of Coding Theory, Vol. 2, 1269–1343, 1998.*

[3] L. Babai, R. Beals, J.-Y. Cai, G. Ivanyos, E. M. Luks, Multiplicative equations over commuting matrices, *Proc. 7th ACM-SIAM Symp. on Discrete Algorithms, 498–507, 1996.*

[4] L. Babai, R. Beals, D. Rockmore, Deciding finiteness for matrix groups in deterministic polynomial time, *Proc. ISSAC '93, 117–126, 1993.*

[5] L. Babai, K. Friedl, M. Stricker, Decomposition of *-closed algebras in polynomial time, *Proc. ISSAC'93, 86–94, 1993.*

[6] L. Babai, L. Rónyai, Computing irreducible representations of finite groups, *Mathematics of Computation 55, 705–722, 1990.*

[7] L. Babai, E. Szemerédi, On the complexity of matrix group problems I., *Proc. 25th IEEE FOCS, 229–240, 1984.*

[8] D. Bacon, A. Childs, and W. van Dam. From optimal measurement to efficient quantum algorithms for the hidden subgroup problem over semidirect product groups. *Proc. 46th IEEE FOCS, 469–478, 2005.*

[9] J. R. Bastida, Field Extensions and Galois Theory, *G-C. Rota (ed.), Encyclopedia of Mathematics and Its Applications, Vol. 22. Cambridge University Press and Addison-Wesley, 1984.*

[10] M. Batty, S. L. Braunstein, A. J. Duncan, S. Rees, Quantum algorithms in group theory, *In: A. V. Borovik, A. G. Myasnikov (eds.), Combinatorial and Experimantal Group Theory, AMS Contemporary Mathematics 349 (2004) 1-62.*
Preprint: arXiv:quant-ph/0310133 (http://arxiv.org/abs/quant-ph/0310133).

[11] R. Beals, L. Babai, Las Vegas algorithms for matrix groups, *Proc. 34th IEEE FOCS, 427–436, 1993.*

[12] M. Ben-Or, D. Coppersmith, M. Luby, R. Rubinfeld, Non-abelian homomorphism testing, and distributions close to their self-convolutions, *Proc. 8th RANDOM, Sringer LNCS Vol. 3122, 273–285, 2004.*

[13] E. Bernstein, U. Vazirani, Quantum complexity theory, *Proc. 25th ACM STOC, 11-20, 1993.*

[14] D. Bini, V. Pan, Polynomial and Matrix Computations, Vol. 1 (Fundamental Algorithms) *Birkhäuser, Basel, 1994.*

[15] M. Blum, M. Luby, R. Rubinfeld, Self–testing/correcting with applications to numerical problems, *J. of Computer and System Sciences 47, 549–595, 1993.*

[16] P. Brooksbank, E. M. Luks, Testing isomorphism of modules in polynomial time, *Manuscipt in preparation, 2007.*

[17] H. Buhrman, R. Cleve, J. Watrous, R. de Wolf, Quantum fingerprinting, *Phys. Rev. Lett., 87(16), Article 167902.*

[18] J. J. Cannon, W. Bosma (Eds.) Handbook of Magma Functions, *Edition 2.13, 2006, (http://magma.maths.usyd.edu.au/magma).*

[19] K. Cheung, M. Mosca, Decomposing finite Abelian groups, *Quantum Information and Computation 1, 26–32, 2001.*

[20] A. Chistov, G. Ivanyos, M. Karpinski, Polynomial time algorithms for modules over finite dimensional algebras, *Proc. 1997 Int. Symp. on Symbolic and Algebraic Computation, 68–74, 1997.*

[21] A. M. Cohen, G. Ivanyos, D. B. Wales, Finding the radical of an algebra of linear transformations, *Journal of Pure and Applied Algebra 117–118, 177–193 1997.*

[22] C. W. Curtis, I. Reiner, Representation Theory of Finite Groups and Associative Algebras, *Wiley, New York, 1962.*

[23] H. Derksen, E. Jeandel, P. Koiran, Quantum automata and algebraic groups, *J. Symb. Comp. 39, 357–371, 2005.*

[24] D. Deutsch, Quantum computational networks, *Proceedings of the Royal Society of London, Vol. A425, 73–90, 1989.*

[25] L. E. Dickson, Algebras and Their Arithmetics, *The University of Chicago Press, Chicago, 1923.*

[26] D. B. DiVincenzo, Two-bit gates are universal for quantum computation, *Phys. Rev. A 51, 1015–1022, 1995.*

[27] M. Domokos, Relative invariants of 3×3 matrix triples, *Linear and Multilinear Algebra 47, 175–190, 2000.*

[28] W. M. Eberly, Computations for Algebras and Group Representations, *PhD. thesis, Dept. of Computer Science, University of Toronto, 1989.*

[29] W. M. Eberly, Decomposition of algebras over finite fields and number fields, *Computational Complexity 1, 179–206, 1991.*

[30] W. M. Eberly, M. Giesbrecht, Efficient decomposition of associative algebras, *In: Proc. ISSAC'96, 170–178, 1996.*

[31] W. M. Eberly, M. W. Giesbrecht, Efficient decomposition of associative algebras over finite fields, *J. Symb. Comp. 37, 35–81, 2004.*

[32] J. Edmonds, System of distinct representatives and linear algebra, *Journal of Research of the National Bureau of Standards 718, 241–245, 1967.*

[33] M. Ettinger, P. Høyer, On quantum algorithms for noncommutative hidden subgroups, *Adv. in Appl. Math., 25(3), 239–251, 2000.*

[34] F. Ergün, S. Kannan, R. Kumar, R. Rubinfeld, M. Viswanathan, Spot-Checkers, *J. of Computer and System Sciences 60, 717–751, 2000.*

[35] K. Friedl, Decomposition of Matrix Groups and Algebras, *PhD. thesis, University of Chicago, 1994.*

[36] K. Friedl, G. Ivanyos, F. Magniez, M. Santha, P. Sen, Hidden translation and orbit coset in quantum computing, *Proc. 35th ACM STOC, 1–9, 2003.*

[37] K. Friedl, G. Ivanyos, M. Santha, Efficient testing of groups, *Proc. 37th ACM STOC, 157–166, 2005.*

[38] K. Friedl, L. Rónyai, Polynomial time solution of some problems in computational algebra, *Proc. 17th ACM STOC, 153–162, 1985.*

[39] The GAP Group, GAP – Groups, Algorithms, and Programming, *Version 4.4.9, 2006 (http://www.gap-system.org)*

[40] An algebraic matching algorithm, *Combinatorica 20, 61–70 (2000).*

[41] M. Giesbrecht, Nearly optimal algorithms for canonical matrix forms, *SIAM J. Comput. 24 948–969, 1995.*

[42] M. Giesbrecht, Factoring in skew-polynomial rings over finite fields, *J. Symbolic Comput. 26, no. 4, 463–486, 1998.*

[43] O. Goldreich, Combinatorial property testing — A survey, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science Vol. 43, 45–60, 1998.*

[44] W. A. de Graaf, Using Cartan subalgebras to calculate nilradicals and Levi subalgebras of Lie algebras, *J. Pure and Applied Algebra 139, 25–39, 1999.*

[45] W. A. de Graaf, G. Ivanyos, Finding maximal tori and splitting elements in matrix algebras, *In: F. van Oysteayen, M. Saorin (eds), Interaction between Ring Theory and Representations of Algebras, Lecture Notes in Pure and Applied Mathematics 210, Marcel Dekker, 95–105, 2000.*

[46] M. Grigni, L. Schulman, M. Vazirani, U. Vazirani, Quantum mechanical algorithms for the nonabelian Hidden Subgroup Problem, *Proc. 33rd ACM STOC, 68–74, 2001.*

[47] W. Greub, Multilinear Algebra, 2nd ed., *Springer-Verlag, New York, 1978.*

[48] R. M. Guralnick, P. H. Tiep, Decompositions of small tensor powers and Larsen's conjecture, *Represent. Theory 9, 138–208, 2005.*

[49] S. Hallgren, A. Russell, A. Ta-Shma, Normal subgroup reconstruction and quantum computation using group representations *SIAM J. Comp. 32, 916–934, 2003.*

[50] D. F. Holt, B. Eick, E. O'Brien, Handbook of computational group theory, *Chapman & Hall/CRC Press, 2005.*

[51] Hoefling, Efficient multiplication algorithms for finite polycyclic groups, *Preprint , 2004.*

[52] D. F. Holt, S. Rees, Testing modules for irreducibility. *J. Austral. Math. Soc. Ser. A 57, 1–16, 1994.*

[53] G. Ivanyos, Finding the radical of matrix algebras using Fitting decompositions, Journal of Pure and Applied Algebra 139, 159-182, 1999.

[54] G. Ivanyos, Fast randomized algorithms for the structure of matrix algebras over finite fields, *Proc. 2000 Int. Symp. on Symbolic and Algebraic Computation, 175–183, 2000.*

[55] G. Ivanyos, Deciding finiteness for matrix semigroups over function fields over finite fields, *Israel Journal of Mathematics 124, 185–188, 2001.*

[56] G. Ivanyos, Deciding universality of quantum gates, *J. Algebra 310, 49–56, 2007.*

[57] G. Ivanyos, On solving random systems of linear disequations, *Submitted, 2007.*

[58] G. Ivanyos, K. Lux, Treating the exceptional cases of the MeatAxe, *Experimental Mathematics 9, 373–381, 2000.*

[59] G. Ivanyos, F. Magniez, M. Santha, Efficient quantum algorithms for some instances of the non-Abelian hidden subgroup problem, *International Journal of Foundations of Computer Science 14, 723–739, 2003.*

[60] G. Ivanyos, L. Rónyai, Á. Szántó, Decomposition of algebras over $F_q(X_1, \ldots, X_m)$, *Applicable Algebra in Engineering, Communication and Computing 5, 71–90, 1994.*

[61] G. Ivanyos, L. Sanselme, M. Santha, An efficient quantum algorithm for the hidden subgroup problem in extraspecial groups, *Proc. STACS 2007, Springer LNCS Vol. 4393, 586–597, 2007.*

[62] G. Ivanyos, L. Sanselme, M. Santha, An efficient quantum algorithm for the hidden subgroup problem in nil-2 groups *Preprint arXiv:0707.1260 [quant-ph], 2007.*

[63] N. Jacobson, Lie Algebras, *Dover, New York, 1979.*

[64] E. Jeandel, Universality in quantum computation, *In: Proc. 31th ICALP, Springer LNCS 3142, 793–804, 2004.*

[65] E. Jeandel, Techniques algébriques en calcul quantique, *PhD. Thesis, ENS Lyon, 2005.*

[66] W. Keller-Gehrig, Fast algorithms for the characterisitic polynomial, *Theoretical Computer Science 36, 309–317, 1985.*

[67] A. Kertész, Lectures on Artinian rings, *Akadémiai Kiadó, Budapest, 1987.*

[68] A. Kitaev. Quantum measurements and the Abelian Stabilizer Problem. *Technical report arXiv:/quant-ph/9511026, 1995.*

[69] A. Y. Kitev, A. H. Shen, M. N. Vyalyi: Classical and quantum computation, *AMS Graduate Studies in Mathematics, Vol. 47, AMS, 2002.*

[70] R. Lipton, New directions in testing, *Series in Discrete Mathematics and Theoretical Computer Science, ACM/AMS, Vol. 2, 191–202, 1991.*

[71] L. Lovász, Singular spaces of matrices and their application in combinatorics, *Bulletin of the Brazilian Mathematical Society 20, 87–99, 1989.*

[72] D. E. Knuth, The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, Ed. 2, *Addison-Wesley, Reading, 1981.*

[73] D. Lazard, Résolution des systèmes d'équations algébriques, *Theoret. Comput. Sci. 15, 77–110, 1981*

[74] M. Kiwi, F. Magniez, M. Santha, Exact and approximate testing/correcting of algebraic functions: A survey, *Proc. 1st Summer School on Theoretical Aspects of Computer Science, Springer LNCS Vol. 2292, 30–83, 2000.*

[75] K. Lux, Algorithmic Methods in Modular Representation Theory, *Habilitation thesis, Rheinisch-Westfälische Technische Hochschule, Aachen, Germany, 1997.*

[76] A. Malcev, On the representation of an algebra as a direct sum of the radical and a semi-simple subalgebra, *Comptes Rendus (Doklady) Acad. Sci. URSS 36, 42–45, 1942.*

[77] C. Moore, D. Rockmore, A. Russell, L. Schulman, The power of basis selection in Fourier sampling: Hidden subgroup problems in affine groups, *Proc. 15th ACM-SIAM SODA, 1106–1115, 2004.*

[78] M. Mosca, Quantum Computer Algorithms, *PhD Thesis, University of Oxford, 1999.*

[79] R. A. Parker, The computer calculation of modular characters (the Meat-Axe). *In: Computational Group Theory, Academic Press, 267–274, 1984.*

[80] R. S. Pierce, Associative Algebras, *Springer-Verlag, 1982.*

[81] S. Rajagopalan, L. Schulman, Verification of Identities, *SIAM J. Computing 29, 1155-1163, 2000.*

[82] I. Reiner, Maximal Orders, *Academic Press, 1975.*

[83] D. N. Rockmore, K.-S. Tan, R. Beals, Deciding Finiteness for Matrix Groups over Function Fields, *Israel J. Math 109, 93-116, 1999.*

[84] D. Ron, Property testing (A tutorial), *Handbook on Randomized Computing, Kluwer Press, 2001.*

[85] L. Rónyai, Computing the structure of finite algebras, *J. Symbolic Computation 9, 355–373, 1990.*

[86] M. Rötteler, T. Beth, Polynomial-time solution to the Hidden Subgroup Problem for a class of non-abelian groups, *Technical report, arXiv:quant-ph/9812070.*

[87] J. T. Schwartz, Fast probabilistic algorithms for verification of polynomial identities, *Journal of ACM 27, 701–717, 1980.*

[88] A. Seidenberg, Constructions in algebra, *Trans. Amer. Math. Soc. 197, 273-313, 1974.*

[89] Á. Seress, An introduction to computational group theory, *Notices of the AMS, 44, 671-679, 1997.*

[90] P. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM J. on Computing 26, 1484–1509, 1997.*

[91] T. Toffoli, Reversible computing, *Proc. 7th ICALP, Springer LNCS Vol. 85, 632-644, 1980.*

[92] J. Watrous, Quantum algorithms for solvable groups, *Proc. 33rd ACM STOC, 60–67, 2001.*

[93] H. Weyl, The Classical Groups, *Princeton University Press, Princeton, 1946.*

[94] D. J. Winter, Abstract Lie Algebras, *The MIT Press, 1972.*

[95] A. Yao, Quantum circuit complexity, *Proc. 34th IEEE FOCS, 352-361, 1993.*

[96] R. E. Zippel, Probabilistic algorithms for sparse polynomials, *Proc. EUROSAM '79, Springer LNCS Vol. 72, 216–226, 1979.*