INTRODUCTION

This introduction begins with an account of the history of the subject and its relations with other areas of mathematics. This is followed by an overview of the whole book, and a brief description of its chapters.

1. The Weyl algebra.

The history of the Weyl algebra begins with the birth of quantum mechanics. The year was 1925. A number of people were trying to develop the principles of the mechanics that was to explain the behaviour of the atom. One of them was Werner Heisenberg. His idea was that this mechanics had to be based on quantities that could actually be observed. In the atomic model of Bohr, there was much talk about orbits; but these are only remotely connected with the things that are actually observed. In the words of Dirac: 'The things that are observed, or which are connected closely with the observed quantities, are all associated with two Bohr orbits and not with just one Bohr orbit: *two* instead of *one*. Now, what is the effect of that?' [Dirac **78**].

To Heisenberg's initial dismay, the 'effect of that' was the introduction of noncommutative quantities. Let Dirac continue with the story:

> Suppose we consider all the quantities of a certain kind associated with two orbits, and we want to write them down. The natural way of writing down a set of quantities, each associated with two elements, is in a form like this:
>
> $$\begin{pmatrix} \times & \times & \times & \times & \dots \\ \times & \times & \times & \times & \dots \\ \times & \times & \times & \times & \dots \\ \times & \times & \times & \times & \dots \\ . & . & . & . & \dots \end{pmatrix}$$
>
> an array of quantities, like this, which one sets up in terms of rows and columns. One has the rows connected with one of the states, the columns connected with the other. Mathematicians call a set of quantities like this a matrix.

The sequence of events is a little more intricate than Dirac's comments will have us believe. In 1925 matrices were not part of the toolkit of every physicist, as they now are. What Heisenberg originally introduced were quantum theoretical analogues of the classical Fourier series. These were supposed to describe the dynamical variables of the atomic systems; and they did not commute. Unable to go as far as he had hoped, Heisenberg decided to sum up his ideas in a paper that he presented to Max Born. Born was acquainted with matrices and was the first to realize that matrix theory offered the correct formalism for Heisenberg's ideas.

According to Born's approach, the dynamical variables (velocity, position, momentum) should be represented by matrices in quantum theory. Denoting the position matrix by q and the momentum matrix by p, one may write the equation for a system with one degree of freedom in the form

$$pq - qp = \imath\hbar.$$

The work on matrix mechanics began with Born and his assistant P. Jordan, soon to be joined by Heisenberg himself. Early on they understood that the fundamental equation above could not be realized by finite matrices; the matrices of quantum theory had to be infinite.

Matrix mechanics was soon followed by other formalisms. First came E. Schrödinger's wave mechanics. In this approach everything begins with a partial differential equation; a more familiar object to physicists. There was also Dirac's formalism. He chose the relations among the dynamical variables as his starting point. The dynamical variables subject to those relations were the elements of what he called *quantum algebra*.

From Dirac's point of view, one is interested in polynomial expressions in the dynamical variables momentum, denoted by p, and position, denoted by q. It is assumed that the variables satisfy the (normalized) relation $pq - qp = 1$. This is what we now call the first Weyl algebra. In particular, he showed how one could use the relation between p and q to differentiate polynomial expressions with respect to p and to q. The Weyl algebras of higher index appear when one considers systems with several degrees of freedom. The relations among the operators in this case had been established as early as September 1925 by Heisenberg. Dirac's point of view received a masterly presentation in H. Weyl's pioneering book *The theory of groups and quantum mechanics*, [Weyl **50**].

From that point onwards the mathematicians were ready to take over. Of very special interest is the paper [Littlewood **33**] of D. Littlewood. He begins by saying that although finite dimensional algebras had been intensively studied, the same was not true of algebras of infinite dimensions. One of the algebras he studies is Dirac's quantum algebra. Littlewood carefully constructs (infinite) matrices p and q for which the equation $pq - qp = 1$ is satisfied; see Exercise 1.4.10.

In his paper Littlewood established many of the basic properties of the Weyl algebra. He showed that the elements of the algebra have a canonical form (Ch. 1 ,§2) and that the algebra is a domain (Ch. 2, §1). He also showed that the relation $pq - qp = 1$ is not compatible with any other relation. Or, as we would now say, the only proper ideal of this algebra is zero (Ch. 2, §2).

The modern age in the theory of the Weyl algebra arrived when its connections with Lie algebras were realized. Suppose that $\mathbf{n}$ is a nilpotent Lie algebra over $\mathbb{C}$. Let $U(\mathbf{n})$ be its enveloping algebra. The quotient of $U(\mathbf{n})$ by a *primitive* ideal is always isomorphic to the Weyl algebra; see [Dixmier **74**, Théorème 4.7.9]. In [Dixmier **63**], the notation A_n was introduced for the algebra that corresponds to the physicist's system with n degrees of freedom. The name Weyl algebra was used by Dixmier as the title of [Dixmier **68**] following, as he says, a suggestion of I. Segal in [Segal **68**].

The increasing interest in noncommutative noetherian rings that followed A. Goldie's work and the intense development of the theory of enveloping algebras of Lie algebras contributed to keeping up the interest in the Weyl algebra. That is not the end of the story though: another theme was added in the seventies under the guise of D-module theory.

2. Algebraic D-modules.

Under the cryptic name of *D-module* hides a modest module over a ring of differential operators. The importance of the theory lies in its manifold applications, which span a vast territory. The representation theory of Lie algebras, differential equations, mathematical physics, singularity theory and even number theory have been influenced by D-modules.

One of the roots of the theory is the idea of considering a differential equation as a module over a ring of differential operators discussed in Ch.

6. This approach goes back at least to the sixties, when it was applied by B. Malgrange to equations with constant coefficients. A turning point was reached in 1971 with M. Kashiwara's thesis 'Algebraic study of systems of partial differential equations', where the same approach was systematically applied to equations with analytic coefficients. In this context, the theory is often called by the alternative name of *algebraic analysis*.

At the same time, in the Soviet Union, I. N. Bernstein was developing the theory of modules over the Weyl algebra. His starting point, however, was entirely different. I.M. Gelfand had asked in the International Congress of Mathematicians of 1954 whether a certain function of a complex variable, which was known to be analytic in the half plane $\Re(z) > 0$, could be extended to a meromorphic function defined in the whole complex plane. The problem remained open until 1968, when M.F. Atiyah and, independently, Bernstein and I.S. Gelfand gave affirmative answers. Both proofs made use of Hironaka's *resolution of singularities*, a very deep and difficult result.

Four years later, Bernstein discovered a new proof of the same result that was very elementary. The key to the proof was a clever use of the Weyl algebra. In his papers, Bernstein introduced many of the concepts that we will study in this book.

Of course the theory of D-modules is not restricted to the Weyl algebra. The theory has two branches: an analytic and an algebraic one; depending on whether the base variety is analytic or algebraic. Highly sophisticated machinery is required in the study of general D-modules, and the most important results cannot be introduced without derived categories and sheaves.

Perhaps the most spectacular result of the theory to date is the *Riemann-Hilbert correspondence* obtained independently by M. Kashiwara and Z. Mebkhout in 1984. This is a result of noble parentage; it can be traced to Riemann's memoir on the hypergeometric function. Its genealogical tree includes the work of Fuchs on differential equations with regular singular points and Hilbert's 21st problem of 1900. Very roughly speaking, the correspondence establishes an (anti-)equivalence between certain differential equations (described in terms of D-modules) and their solution spaces. Unfortunately, the correspondence requires deep results of category theory and cannot be included in an elementary book.

3. The book: an overview.

This book is about modules over the Weyl algebra, and the point of view is that of algebraic D-modules. It is also fair to say that this is a book about certain aspects of the representation theory of the Weyl algebra. So we will have a lot to say about irreducible modules and the dimension of modules, for example.

The book can be divided into two parts. The first part, which goes up to Ch. 11, is concerned with invariants of modules over the Weyl algebra; the most important being the dimension and the multiplicity.

The first two chapters deal with ring theoretic properties of the Weyl algebra itself. Most importantly, we show that it is a simple domain. Ch. 3 establishes the Weyl algebra as a member of the family of rings of differential operators. There is much talk of derivations in this chapter, and they are put to good use in the next chapter. Ch. 4 contains the first of our applications. It consists of using a *conjecture* about the Weyl algebra to derive the *Jacobian conjecture.*

With Ch. 5 we get into the realm of representation theory. The purpose of the chapter is to describe a few important examples of modules over the Weyl algebra. The relation of these modules with differential equations is found in Ch. 6, which also includes an elementary description of the module of microfunctions. These are used as generalized solutions of differential equations.

Chs. 7 to 9 form a sequence which culminates in the definition of the dimension and multiplicity of a module over the Weyl algebra and the study of their properties. Place of honour is given to Bernstein's inequality: the dimension of a module over the n-th Weyl algebra is an integer between n and $2n$. The modules of minimal dimension form such a nice category that they have a special name: holonomic modules. The whole of Ch. 10 is dedicated to them. Ch. 11 requires a smattering of algebraic geometry; it deals with the relation between D-modules and symplectic geometry. The key concept is another invariant of a D-module, the *characteristic variety.* This allows us to give a geometrical interpretation to the dimension previously defined.

The emphasis now shifts from invariants to operations, which are the theme of the second part. These operations are geometrical constructions which use a polynomial map to produce new D-modules. Since their definition depends on the use of the tensor product, which may not be familiar to some readers, we have included a discussion of them in Ch. 12.

Chs. 13 to 16 contain the definitions and examples of the three main operations that we apply to modules over the Weyl algebra: external products, inverse images and direct images. The results are served in homœopathic doses, since the calculations tend to be crammed with detail. Kashiwara's theorem is one of the pearls of the theory. A simple, rather meagre version is described in Ch. 17, but it will return in greater splendour in Ch. 18. In this chapter we also show that all the operations previously described map holonomic modules to holonomic modules. This is very mystifying since the dimension of a module is *not* preserved by some of these operations.

Finally we return to applications in Chs. 19 and 20. The former is concerned with the global stability of ordinary differential equations. We will discuss conditions under which a system of differential equations with polynomial coefficients has a global stability point. The key lemma has a very neat D-module theoretic proof due to van den Essen. It is discussed in detail in §2. The proof of the stability theorem is sketched in §3. Ch. 20 is about the work of Zeilberger and his collaborators on the automatic computation of sums and integrals. That is, how can D-modules help a computer to calculate a definite integral?

4. Pre-requisites.

All the algebra required in the book can be found in [Cohn **84**]. Like all general rules, this one has exceptions; these are Ch. 11 and Ch. 18, §3. We have already mentioned that one needs to know some algebraic geometry to follow Ch. 11. All the required results will be found in [Hartshorne **77**, Ch. 1]. In §3 of Ch. 18 we rewrite the results of Chs. 14 to 17 using categories. The subjects of these sections are not used anywhere else in the book, except in some exercises.

The book also requires a good knowledge of analysis to be properly appreciated. This is particularly true of examples and applications. However, most of the results we use are part of the standard undergraduate curriculum; and we will give several references to those which are not so well-known. Ch. 19, in particular, requires some results about ordinary differential equations. A good reference for our needs is [Arnold **81**].

The book is linearly ordered and the order is almost, though not quite, total. There are two obvious exceptions: Ch. 11 and §3 of Ch. 18 depend on everything that comes before them, but are not used anywhere else in the book. The only other exceptions are Chs. 4 and 6. We will need Ch. 4, §1 later on, but that will be only in Ch. 14.

CHAPTER 1

THE WEYL ALGEBRA

We will describe the main protagonist of this book, the Weyl algebra, in two different ways: as a ring of operators and in terms of generators and relations.

1. Definition

In this section the Weyl algebra is introduced as a ring of operators on a vector space of infinite dimension. Let us fix some notation. Throughout this book, K denotes a field of characteristic zero and $K[X]$ the ring of polynomials $K[x_1, \dots, x_n]$ in n commuting indeterminates over K.

The ring $K[X]$ is a vector space of infinite dimension over K. Its algebra of linear operators is denoted by $End_K(K[X])$. Recall that the algebra operations in the endomorphism ring are the addition and composition of operators. The Weyl algebra will be defined as a subalgebra of $End_K(K[X])$.

Let $\hat{x}_1, \dots, \hat{x}_n$ be the operators of $K[X]$ which are defined on a polynomial $f \in K[X]$ by the formulae $\hat{x}_i(f) = x_i \cdot f$. Similarly, $\partial_1, \dots, \partial_n$ are the operators defined by $\partial_i(f) = \partial f / \partial x_i$. These are linear operators of $K[X]$. The *n-th Weyl algebra* A_n is the K-subalgebra of $End_K(K[X])$ generated by the operators $\hat{x}_1, \dots, \hat{x}_n$ and $\partial_1, \dots, \partial_n$. For the sake of consistency, we write $A_0 = K$.

Note that for $n \geq m$, the action of the operators of A_m on $K[X]$ is well-defined. Thus A_m is a subalgebra of A_n in a natural way. We sometimes write $A_n(K)$ instead of A_n, if it is necessary to make explicit the field over which the algebra is defined.

According to our definition, the elements of A_n are linear combinations over K of monomials in the generators $\hat{x}_1, \dots, \hat{x}_n, \partial_1, \dots, \partial_n$. However, one has to be careful when representing the elements of A_n because this algebra is not commutative. This is easily checked, as follows. Consider the operator $\partial_i \cdot \hat{x}_i$ and apply it to a polynomial $f \in K[X]$. Using the rule for the differentiation of a product, we get $\partial_i \cdot \hat{x}_i(f) = x_i \partial f / \partial x_i + f$. In other words,

$$\partial_i \cdot \hat{x}_i = \hat{x}_i \cdot \partial_i + 1$$

where 1 stands for the identity operator. It is better to rewrite this formula using commutators. If $P, Q \in A_n$ then their *commutator* is the operator $[P, Q] = P \cdot Q - Q \cdot P$. The formula above becomes $[\partial_i, \hat{x}_i] = 1$. Similar calculations allow us to obtain formulae for the commutators of the other generators of A_n. These are summed up below:

$$[\partial_i, \hat{x}_j] = \delta_{ij} \cdot 1,$$
$$[\partial_i, \partial_j] = [\hat{x}_i, \hat{x}_j] = 0,$$

where $1 \leq i, j \leq n$. Recall that δ_{ij} is the Kronecker delta symbol: it equals 1 if $i = j$ and zero otherwise. A final observation. We have denoted the operator 'multiplication by x_i' by the symbol $\hat{x}_i$. From now on, we shall follow the standard convention and write x_i for both the variable and the corresponding operator. This tends to make the notation less cluttered. For the same reason we shall dispense with the subscripts for the generators of A_1, and write them simply as x and ∂.

2. Canonical form

In this section we construct a basis for the Weyl algebra as a K-vector space. This basis is known as the *canonical basis*. If an element of A_n is written as a linear combination of this basis then we say that it is in *canonical form*. Of course, to compare two elements in canonical form it is enough to compare the coefficients of their linear combinations, and that is easily done.

It is easier to describe the canonical basis if we use a multi-index notation. A *multi-index* α is an element of $\mathbb{N}^n$; say $\alpha = (\alpha_1, \ldots, \alpha_n)$. Now by x^α we mean the monomial $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$. The *degree* of this monomial is the *length* $|\alpha|$ of the multi-index α, namely $|\alpha| = \alpha_1 + \ldots + \alpha_n$. Notice that a pair (α, β) of multi-indices in $\mathbb{N}^n$ is itself a multi-index in $\mathbb{N}^{2n}$, so it makes sense to talk of its length.

2.1 Proposition. *The set* $\mathbf{B} = \{x^\alpha \partial^\beta : \alpha, \beta \in \mathbb{N}^n\}$ *is a basis of* A_n *as a vector space over* K.

The proof of this proposition uses a formula for the derivative of polynomials in terms of multi-indices that we state below. The factorial of a multi-index $\beta \in \mathbb{N}^n$ is defined by $\beta! = \beta_1! \dots \beta_n!$. The formula is written in terms of powers of the operators ∂_i. The proof is left to the reader.

2.2 LEMMA. *Let $\sigma, \beta \in \mathbb{N}^n$ and assume that $|\sigma| \leq |\beta|$. Then $\partial^\beta(x^\sigma) = \beta!$ if $\sigma = \beta$, and zero otherwise.*

PROOF OF THE PROPOSITION: It is easy to see that the elements of **B** generate the Weyl algebra as a vector space. Consider a monomial on the generators of A_n. Using the relations of §1, one shows that if $f \in K[x]$, then $\partial_i \cdot f - f \cdot \partial_i = \partial f / \partial x_i$. That allows us to bring all powers of x's to the left of all the ∂'s. By doing that, the monomial automatically ends up written as a linear combination of the elements of **B**.

Now to the uniqueness. Consider a finite linear combination of elements of **B**, say $D = \sum c_{\alpha\beta} x^\alpha \partial^\beta$. We must show that if some $c_{\alpha\beta}$ is non-zero then $D \neq 0$. But D is a linear operator of $K[X]$. Hence $D \neq 0$ if and only if there exists a polynomial f for which $D(f) \neq 0$. We construct such an f.

Let σ be a multi-index which satisfies $c_{\alpha\sigma} \neq 0$ for some index α, but $c_{\alpha\beta} = 0$, for all indices β such that $|\beta| < |\sigma|$. A straightforward calculation using Lemma 2.2 shows that $D(x^\sigma) = \sigma! \sum_\alpha c_{\alpha\sigma} x^\alpha$. This is non-zero since at least one of the coefficients $c_{\alpha\sigma}$ is non-zero by the choice of σ. Thus $f = x^\sigma$ is the required polynomial.

3. GENERATORS AND RELATIONS

Another way to define the Weyl algebra is by generators and relations. More precisely, we may write the Weyl algebra as a quotient of a free algebra in $2n$ generators. The ideal that we factor out is generated by the relations calculated in §1.

The *free algebra* $K\{z_1, \dots, z_{2n}\}$ in $2n$ generators is the set of all finite linear combinations of words in $z_1, \dots, z_{2n}$. Multiplication of two monomials is simple juxtaposition. We may define a surjective homomorphism $\phi : K\{z_1, \dots, z_{2n}\} \longrightarrow A_n$ by $\phi(z_i) = x_i$ and $\phi(z_{i+n}) = \partial_i$, for $i = 1, 2, \dots, n$.

Let J be the two-sided ideal of $K\{z_1, \dots, z_{2n}\}$ generated by $[z_{i+n}, z_i] - 1$ for $i = 1, 2, \dots, n$ and $[z_i, z_j]$ for $j \neq i + n$ and $1 \leq i, j \leq 2n$. It follows from the relations of §1 that $J \subseteq \ker \phi$. Thus ϕ induces a homomorphism of K-algebras $\hat{\phi} : K\{z_1, \dots, z_{2n}\}/J \longrightarrow A_n$.