Chapter 2

Non-commutative polynomial equations

We work over a base field k in this chapter.

This chapter discusses the fundamental problem of non-commutative algebra: classify the matrix solutions to a system of non-commutative polynomial equations. This is equivalent to classifying the finite dimensional modules over the associative algebra having those equations as its defining relations. The effectiveness of the theory of non-commutative algebra can be judged, in part, by how well it handles this problem.

The representation theories of groups and Lie algebras are historically important examples. In both cases the finite dimensional representation theory is equivalent to the theory of solutions to the system of non-commutative polynomial equations that arises when the group or Lie algebra is presented by generators and relations. The associative algebra determined by these equations is, in the first case, the group algebra, and, in the second case, the universal enveloping algebra.

Polynomials in non-commuting variables $x_1, \ldots, x_n$ are, by definition, elements of the free algebra $k\langle x_1, \ldots, x_n\rangle$. Section 2.1 shows that the set of $d \times d$ matrix solutions to a system of equations in a finite number of variables is an affine algebraic variety over k. In section 2.2 we show that solutions to such a system are in natural bijection with the finite-dimensional modules over the associative algebra having those equations as defining relations. This leads to the problem of understanding an algebra that is given to us by generators and relations. It can be difficult to find a basis for an algebra given by generators and relations: there are no universally effective methods, only ad hoc techniques. However, Bergman's Diamond Lemma will, in many situations, produce a basis for a finitely presented algebra. This method is presented in section 2.4. Finding a basis for an algebra tells us the "size" of the algebra in some sense—this topic is taken up in section 2.5 where we introduce Gelfand-Kirillov dimension. Its definition makes sense for any algebra but for a commutative algebra it agrees with some of the usual dimensions—if I is an ideal in the polynomial ring $R = k[x_1, \ldots, x_n]$ and $X \subset \mathbb{A}^n$ its zero locus, then

$$\mathrm{GKdim}(R/I) = \dim X.$$

In other words, in the commutative setting, the GK-dimension of the algebra gives the dimension of the space of solutions to the system of polynomials that defines the algebra. A similar, though less precise, result is true for non-commutative algebras too. The GK-dimension of an algebra gives us an upper bound on the "dimension" of the space of isomorphism classes of d-dimensional modules over that algebra.

2.1 Non-commutative polynomial equations

Definition 1.1 A polynomial in non-commuting variables $x_1, \ldots, x_n$ is an element of the free algebra $k\langle x_1, \ldots, x_n\rangle$. A $d \times d$ matrix solution to a system

$$f_1 = \cdots = f_r = 0$$

of non-commutative polynomial equations is an n-tuple $a_1, \ldots, a_n \in M_d(k)$ such that

$$f_1(a_1, \ldots, a_n) = \cdots = f_r(a_1, \ldots, a_n) = 0. \qquad \Diamond$$

To evaluate a non-commutative polynomial $f \in k\langle x_1, \ldots, x_n\rangle$ at an n-tuple $(a_1, \ldots, a_n)$ of matrices $a_i \in M_d(k)$ means to substitute, in the expression for f, the matrix a_i wherever x_i occurs and then evaluate the resulting expression using the arithmetic of $M_d(k)$. By Proposition 1.4.3, the process of replacing each x_i by a_i determines a unique k-algebra homomorphism

$$\Psi : k\langle x_1, \ldots, x_n\rangle \to M_d(k), \quad x_i \mapsto a_i \tag{1-1}$$

and $(a_1, \ldots, a_n)$ is a solution to the equation $f = 0$ if and only if $\Psi(f) = 0$.

The set of solutions to the system of equations $\mathcal{F} = 0$ depends only on the two sided ideal $(\mathcal{F})$ generated by $\mathcal{F}$ because $(a_1, \ldots, a_n)$ is a solution if and only if $(\mathcal{F}) \subset \ker \Psi$ where Ψ is defined by (1-1).

Definition 1.2 Let $\mathcal{F}$ be a set of non-commutative polynomial equations in a set of unknowns X. Thus $\mathcal{F} \subset k\langle X\rangle$. Let $\psi : X \to M_d(k)$ be a map and let $\Psi : k\langle X\rangle \to M_d(k)$ be the k-algebra homomorphism extending ψ. We say that ψ is a solution to the system of equations $\mathcal{F}$ if $\Psi(f) = 0$ for all $f \in \mathcal{F}$. $\Diamond$

Proposition 1.3 *Let $S \subset M_d(k)^n$ be the set of $d \times d$ solutions to a system of of equations $\mathcal{F} \subset k\langle x_1, \ldots, x_n\rangle$. Then S is an algebraic subvariety of $\mathbb{A}^{nd^2} = M_d(k)^n$.*

Proof. A $d \times d$ solution to $\mathcal{F} = 0$ is a point $(a_1, \ldots, a_n)$ in the nd^2-dimensional vector space $M_d(k)^n$.

Write each x_t as a $d \times d$ matrix of unknowns, say $x_t = (y_{ij}^{(t)})$ with $1 \leq, j \leq d$. The equation $f = 0$, $f \in \mathcal{F}$, in the n unknowns x_t can be expressed as d^2 equations, one for each matrix position, in the nd^2 unknowns $y_{ij}^{(t)}$. As f runs over $\mathcal{F}$, one obtains a collection, $\mathcal{G}$ say, of equations in the $y_{ij}^{(t)}$'s, and the

solutions in k^{nd^2} to the system $\mathcal{G} = 0$ are in natural bijection with solutions in $M_d(k)^n$ to the system $\mathcal{F} = 0$.

Since the $y_{ij}^{(t)}$s are required to take values in k, the equations in $\mathcal{G}$ may be viewed as elements of the commutative polynomial ring

$$k[y_{ij}^{(t)} \mid 1 \leq t \leq n, 1 \leq i, j \leq d].$$

Let J be the ideal generated by $\mathcal{G}$. Then solutions to $\mathcal{F} = 0$ are in bijection with the points of k^{nd^2} where J vanishes—so S is the affine algebraic variety that is the zero locus of J. □

Hilbert's Basis Theorem ensures that J is finitely generated, so the set of $d \times d$ solutions to $\mathcal{F} = 0$ is the same as the set of $d \times d$ solutions to some *finite* subset of $\mathcal{F}$.

The action of $\mathrm{GL}_d(k)$ on $\mathrm{sol}_d\mathcal{F}$. If $(a_1, \ldots, a_n) \in M_d(k)^n$ is a solution to a given system of equations so is $(ga_1g^{-1}, \ldots, ga_ng^{-1})$ for every $g \in \mathrm{GL}_d(k)$. Two solutions related in such a way are essentially the same, differing only by a choice of basis. This suggests the next definition.

Definition 1.4 Two solutions $\psi_i : X \to M_d(k)$ $(i = 1, 2)$ to the system of equations $\mathcal{F} \subset k\langle X \rangle$ are equivalent if there exists $g \in \mathrm{GL}_d(k)$ such that $\psi_2(x) = g\psi_1(x)g^{-1}$ for all $x \in X$. ◊

The solutions as a scheme. For many purposes it is better to view the set of $d \times d$ solutions to the system $\mathcal{F} = 0$ as a scheme, so we define as

$$\mathrm{sol}_d\mathcal{F} := \mathrm{Spec}\, k[y_{pq}^{(i)}]/J$$

where J is the ideal appearing in the proof of Proposition 1.3. The k-rational points of $\mathrm{sol}_d\mathcal{F}$ are the solutions in $M_d(k)$. There is an action of $\mathrm{GL}_d(k)$ on $\mathrm{sol}_d\mathcal{F}$ induced by its conjugation action on the matrices of variables $Y_t = (y_{ij}^{(t)})$. The orbits of the $\mathrm{GL}_d(k)$ action on the scheme $\mathrm{sol}_d\mathcal{F}$ give the equivalence classes of solutions. Hence one may bring the arsenal of Geometric Invariant Theory to bear on this problem.

Example 1.5 Consider the problem of finding all 2×2 solutions to the single equation

$$(xy - yx)^2z - z(xy - yx)^2 = 0.$$

We claim that every triple of 2×2 matrices is a solution. To see this observe that the trace of $xy - yx$ is zero, and that the square of a trace zero 2×2 matrix is a scalar multiple of the identity, so commutes with every matrix z.

This is an example of a polynomial identity of 2×2 matrices. If $f \in k\langle x_1, \ldots, x_n \rangle$ is such that *every* n-tuple $a_1, \ldots, a_n$ of $d \times d$ matrices satisfies $f(a_1, \ldots, a_n) = 0$, we call f a polynomial identity of $d \times d$ matrices. ◊

EXERCISES

1.1 Show that there are no matrix solutions to the equation $yx = qxy$ when $0 \neq q \in k$ and q is not a root of unity.

1.2 Show that there are no matrix solutions to the equation $xy - yx = 1$ when $\operatorname{char} k = 0$.

1.3 Let $k[t]$ be the polynomial ring. Let $P : k[t] \to k[t]$ be the k-linear map $P(f) = df/dt$. Let $Q : k[t] \to k[t]$ be the k-linear map $Q(f) = tf$, the product of f with t.

(a) Show that $PQ - QP$ is the identity operator. Thus, although the equation $xy - yx = 1$ has no solutions $d \times d$ solutions when $\operatorname{char} k = 0$, there are solutions in the ring of linear operators on an *infinite* dimensional vector space; for example, P and Q.

(b) Suppose that $\operatorname{char} k = 0$. Show that the only subspaces of $k[t]$ stable under the action of P and Q are 0 and $k[t]$.

1.4 Suppose that $\operatorname{char} k = p > 0$. Show that there exist $p \times p$ matrices such that $xy - yx = 1$. [*Hint:* the operators P and Q in the previous exercise still make sense when $\operatorname{char} k = p$, but $k[t]$ now has proper subspaces that are stable under P and Q. Look for such subspaces, and consider the action of P and Q on the quotient, which will be finite dimensional.]

2.2 Modules and solutions

The next two results say that finding solutions to a system of equations is equivalent to finding modules over a certain ring determined by those equations.

Proposition 2.1 *Let $\mathcal{F} \subset k\langle X\rangle$ be a system of non-commutative polynomial equations. Let*

$$\pi : k\langle X\rangle \to A := k\langle X\rangle/(\mathcal{F})$$

denote the quotient map. There is a bijection

$$\{\textit{solutions in } M_d(k) \textit{ to } \mathcal{F} = 0\} \longleftrightarrow \{k\textit{-algebra maps } \rho : A \to M_d(k)\}$$
$$\rho \circ \pi|_X \longleftrightarrow \rho.$$

Proof. If $\rho : A \to M_d(k)$ is an algebra map, then $\Psi := \rho \circ \pi$ satisfies $\Psi(f) = 0$ for all $f \in \mathcal{F}$, simply because $\mathcal{F} \subset \ker \pi$. Hence $\Psi|_X$ gives a solution. Conversely, if $\psi : X \to M_d(k)$ is a solution, its extension to $k\langle X\rangle$, Ψ say, vanishes on $\mathcal{F}$. Therefore $\Psi = \rho \circ \pi$ for some ρ. □

We will often write x for the image in A of $x \in X$.

Proposition 2.2 *Let $\mathcal{F} \subset k\langle X\rangle$ and define $A = k\langle X\rangle/(\mathcal{F})$. There is a bijection*

$$\{\textit{solutions in } M_d(k) \textit{ to } \mathcal{F} = 0\} \longleftrightarrow \{d\textit{-dimensional left } A\textit{-modules}\}. \quad (2\text{-}2)$$

The A-module corresponding to a solution $\psi : X \to M_d(k)$ is k^d with action given by $r.v = \rho(r)(v)$ for $r \in A$ and $v \in k^d$, where $\rho : A \to M_d(k)$ is the unique k-algebra homorphism such that $\rho|_X = \psi$.

Proof. Fix a basis for k^d, and use it to identify $M_d(k)$ with $\mathrm{Hom}_k(k^d, k^d)$.

Given a solution ψ, the algebra map $\rho : A \to M_d(k) = \mathrm{Hom}_k(k^d, k^d)$, associated to ψ as in Proposition 2.1 gives k^d a left A-module structure. Conversely, a left A-module structure on k^d is the same thing as a ring homomorphism $\rho : A \to \mathrm{Hom}_k(k^d, k^d)$. Therefore, if $\pi : k\langle X\rangle \to A$ is the quotient map, applying Proposition 2.1 to the composition $\rho \circ \pi : k\langle X\rangle \to M_d(k)$, yields a solution to $\mathcal{F} = 0$. □

Thus the problem of understanding the solutions to $\mathcal{F} = 0$ is equivalent to the problem of understanding the finite dimensional $k\langle X\rangle/(\mathcal{F})$-modules. The next result shows that equivalent solutions, in the sense of Definition 1.4, correspond to isomorphic modules, so the problem of finding matrix solutions to a system of equations may be rephrased as that of understanding finite dimensional $k\langle X\rangle/(\mathcal{F})$-modules up to isomorphism.

Proposition 2.3 *Two solutions $\psi_i : X \to M_d(k)\,(i = 1, 2)$ to $\mathcal{F} = 0$ are equivalent if and only if the corresponding A-modules are isomorphic.*

Proof. Let $\rho_i : A \to M_d(k)$ be the k-algebra maps corresponding to ψ_i, and let V_i be the corresponding A-modules.

The modules are isomorphic if and only if there is a k-linear isomorphism $\phi : V_1 \to V_2$ such that $\phi(x.v) = x.\phi(v)$ for all $x \in A$ and all $v \in V_1$. Since X generates A, it suffices to check this equality for all $x \in X$. Now $V_1 = V_2 = k^d$, so the linear isomorphism ϕ corresponds, after a choice of basis, to a $g \in \mathrm{GL}_d(k)$ such that $\phi(v) = g.v$ for all $v \in k^d$. Thus

$$\begin{aligned}
\phi(x.v) = x.\phi(v) \quad &\Leftrightarrow \quad g\rho_1(x)(v) = \rho_2(x)g.v \text{ for all } v \in V_1 \text{ and all } x \in X,\\
&\Leftrightarrow \quad g\rho_1(x) = \rho_2(x)g \qquad \text{for all } x \in X,\\
&\Leftrightarrow \quad g\rho_1(x)g^{-1} = \rho_2(x) \quad \text{for all } x \in X,\\
&\Leftrightarrow \quad g\psi_1(x)g^{-1} = \psi_2(x) \quad \text{for all } x \in X.
\end{aligned}$$

□

A finite dimensional A-module M has a composition series, i.e., a chain of submodules $0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$ such that each M_i/M_{i-1} is a simple module. The classification of finite-dimensional A-modules therefore breaks into two parts:

1. find all finite-dimensional *simple* A-modules—i.e., those modules M whose only submodules are itself and 0, and

2. determine how these modules can be joined together to make larger ones—i.e., given modules L and N, find/classify the modules M that fit into an exact sequence $0 \to L \to M \to N \to 0$.

These two problems are mostly separate from each other. The second problem is addressed as follows: there is a vector space $\mathrm{Ext}^1_A(N, L)$ that classifies the Ms. There is a whole family of "Ext-groups" $\mathrm{Ext}^i_A(N, L)$, $i \geq 0$, of which the simplest is $\mathrm{Ext}^0_A(N, L) = \mathrm{Hom}_A(N, L)$.

We will focus on the first step initially. That is, we will focus first on the problem of classifying the finite-dimensional simple modules. Thus, we follow in the grand tradition of representation theory of groups and Lie algebras.

If $(a_1, \ldots, a_n)$ is a $d \times d$ solution to a particular system of equations and $(b_1, \ldots, b_n)$ is an $e \times e$ solution to the same system it follows that

$$c_1 = \begin{pmatrix} a_1 & 0 \\ 0 & b_1 \end{pmatrix}, \ldots, c_n = \begin{pmatrix} a_n & 0 \\ 0 & b_n \end{pmatrix},$$

is a $(d+e) \times (d+e)$ solution to that same system of equations. Thus we have an operation that from two solutions produces a new solution, so we may ask what is the corresponding operation on the modules associated to the solutions? The answer is given in Exercise 4 below—the corresponding operation on modules is taking their direct sum.

2.2.1 Finitely presented algebras

Although what we have said applies to systems of infinitely many equations in infinitely many unknowns, finite systems of equations are the most common. This leads to the following definition.

Definition 2.4 Let A be a k-algebra. We say that A is finitely generated as a k-algebra if there is a finite subset $\{a_1, \ldots, a_n\} \subset A$ such that the smallest subring of A containing k and $a_1, \ldots, a_n$ is A itself.

We say that A is finitely presented if there is a surjective k-algebra map $f : k\langle x_1, \ldots, x_n\rangle \to A$ such that $\ker f$ is finitely generated as a two-sided ideal. The elements of $\ker f$ are called relations for A. If A is defined by saying it is isomorphic to $k\langle x_1, \ldots, x_n\rangle/(f_1, \ldots, f_r)$ we say it is given by generators and relations; the f_i are called defining relations; this is called a presentation of A. $\Diamond$

There is no uniqueness to a presentation of an algebra. For example, the polynomial ring $k[x]$ may be presented as $k\langle x\rangle$ and as $k\langle x, y\rangle/(y)$. The two-sided ideal (y) in $k\langle x, y\rangle$ is not finitely generated as a left ideal (Exercise 6), so it is important to note the requirement in the definition that the defining ideal be finitely generated as a *two-sided ideal*.

Generally speaking, it is not known how to use the hypothesis of finite presentation effectively. The structural consequences of having only a finite number of defining equations for an algebra are unknown despite there being various strange phenomena that have been observed for infinitely presented algebras but not for finitely presented ones.

2.2.2 The relation to commutative algebra

Non-commutative algebra concerns solutions in $M_d(k)$ for *arbitrary* d. The case $d = 1$ is classical; it is the commutative case—one is seeking solutions in k so the solution matrices automatically commute with one another. Therefore the set

of solutions in k to the original system, say $f_1 = \ldots = f_r = 0$, is the same as the set of solutions to the system of *commutative* polynomial equations $g_1 = \ldots = g_r = 0$ where g_j is obtained from f_j by assuming the x_i commute. To make this precise, suppose that each f_i is in $k\langle x_1, \ldots, x_n\rangle$. Let $k[t_1, \ldots, t_n]$ be the polynomial ring in n variables. Define the map $\Phi : k\langle x_1, \ldots, x_n\rangle \to k[t_1, \ldots, t_n]$ by $\Phi(x_i) = t_i$, and define $g_j = \Phi(f_j)$. Then $(a_1, \ldots, a_n) \in k^n$ is a solution to $f_1 = \ldots = f_r = 0$ if and only if it is a solution to $g_1 = \ldots = g_r = 0$.

EXERCISES

2.1 Let $A = k[x, y]$ be the commutative polynomial ring and, for each $n \geq 0$, let A_n be the subspace with basis $\{x^i y^j \mid i + j = n\}$. The linear operators $e = x\partial/\partial y, f = y\partial/\partial x, h = x\partial/\partial x - y\partial/\partial y$ act on A in the usual way. Each A_n is an invariant subspace. Use the given basis for A_n to represent each of e, f, h as a matrix.

2.2

(a) Show that the matrices in Exercise 1 are a solution to the system of equations $he - eh = 2e, ef - fe = h, hf - fh = -2f$.

(b) That solution is presented with respect to a particular ordered basis for k^d, say $\{v_1, \ldots, v_d\}$. Express the action of the ring $k\langle e, f, h\rangle$ on k^d by expressing each of $e.v_i$, $f.v_i$, $h.v_i$ as a linear combination of the v_j's.

(c) Can you find another 1×1 solution? What about other 2×2 and 3×3 solutions?

2.3 Find some/all solutions to the equation $xy + yx = 1$.

2.4 Let $\psi_i : X \to M_{d_i}(k)\,(i = 1, 2)$ be two solutions to the system of equations $\mathcal{F} \subset k\langle X\rangle$. The direct sum of ψ_1 and ψ_2 is $\psi_1 \oplus \psi_2 : k\langle X\rangle \to M_{d_1+d_2}(k)$ defined by

$$(\psi_1 \oplus \psi_2)(x) = \begin{pmatrix} \psi_1(x) & 0 \\ 0 & \psi_2(x) \end{pmatrix}$$

for all $x \in X$. Show that $\psi_1 \oplus \psi_2$ is also a solution to $\mathcal{F} = 0$, and that the corresponding $k\langle X\rangle/(\mathcal{F})$-module is the direct sum of the modules corresponding to ψ_1 and ψ_2.

2.5 Suppose that $\operatorname{char} k \neq 2$. Show that

$$\begin{pmatrix} k[y^2] & yk[y^2] \\ yk[y^2] & k[y^2] \end{pmatrix} \cong \begin{pmatrix} k[t] & (t) \\ k[t] & k[t] \end{pmatrix} \cong \frac{k\langle \sigma, x\rangle}{(\sigma^2 - 1, x\sigma + \sigma x)}.$$

Which of these ways of presenting the algebra makes it easier to understand?

2.6 Show that $k\langle x, y\rangle$ is not noetherian by showing that the ideal (y) in $k\langle x, y\rangle$ is not finitely generated as a left ideal.

2.3 The geometry of representations/solutions

In the previous section we saw that the problem of finding matrix solutions to a system of non-commutative polynomial equations is equivalent to finding the finite-dimensional modules over an algebra (up to isomorphism). We will tend to take the module point of view in what follows.

The first point we take up in this section is the $\mathrm{GL}_d(k)$-action on the solutions. If we consider the solutions as a scheme, then we can avail ourselves of the general machinery that has been developed to study algebraic actions of

algebraic groups on schemes. This well-developed machinery goes by the name of *Geometric Invariant Theory*, or G.I.T. for short; Mumford's book of that name is the bible. The word *git* has some unpleasant associations.

2.3.1 The space of representations

For any k-algebra A, we denote its space of d-dimensional representations (or modules) by

$$\mathrm{rep}_d A := \{\text{algebra homomorphisms } \phi : A \to M_d(k)\}.$$

By Proposition 2.2, $\mathrm{rep}_d A = \mathrm{sol}_d \mathcal{F}$ where $\mathcal{F}$ is a set of defining relations for A. Hence $\mathrm{rep}_d A$ is an affine scheme

$$\mathrm{rep}_d A = \mathrm{Spec}\, k[y_{ij}^{(t)} \mid 1 \le t \le n,\, 1 \le i, j \le d]/J$$

in the notation of Proposition 1.3. The algebraic group $\mathrm{GL}_d(k)$ acts as automorphisms of $\mathrm{rep}_d A$, the action being induced by the conjugation action of $\mathrm{GL}_d(k)$ on the matrices $(y_{ij}^{(t)})_{1 \le i,j \le d}$.

Proposition 3.1 *The orbits of* $\mathrm{GL}_d(k)$ *on* $\mathrm{rep}_d A$ *are the isomorphism classes of* d*-dimensional* A*-modules.*

Proof. This is a restatement of Proposition 2.3. □

Proposition 3.2 *A module* $M \in \mathrm{rep}_d A$ *is semisimple if and only if its orbit under the* $\mathrm{GL}_d(k)$ *action is closed.*

Proof. Let M be a d-dimensional A-module. If $0 = M_0 \subset M_1 \subset \cdots \subset M_\ell = M$ is a composition series for M we call

$$\mathrm{gr}\, M := \bigoplus_{j=1}^{\ell} M_j/M_{j-1}$$

a semisimplification of M. It is well-defined up to isomorphism. We show that the closure of the GL_d-orbit of M contains its semisimplification. It will follow from this that the closure of each orbit contains a semisimple module, so the only orbits that can be closed are those of the semisimple modules. It will then remain only to show that the orbit of a semisimple module is closed.

Write $A = k[x_1, \ldots, x_n]$.

Suppose that L is a submodule of M. We will show that $L \oplus M/L$ is in the closure of the GL_d-orbit of M.

Write $p = \dim_k L$ and $q = \dim_k M/L$; thus $p + q = d$. Pick a basis for L and extend it to a basis for M. With respect to this basis the action of x_i on M is given by

$$x_i \mapsto \begin{pmatrix} a_i & b_i \\ 0 & d_i \end{pmatrix}$$

where the $a_i \in M_p(k)$ give the action of x_i on L and the $d_i \in M_q(k)$ give the action of x_i on M/L. Write I_p and I_q for the $p \times p$ and $q \times q$ identity matrices; let $0 \neq \varepsilon \in k$; a computation shows that

$$\begin{pmatrix} I_p & 0 \\ 0 & \varepsilon^{-1} I_q \end{pmatrix} \begin{pmatrix} a_i & b_i \\ 0 & d_i \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & \varepsilon I_q \end{pmatrix} = \begin{pmatrix} a_i & \varepsilon b_i \\ 0 & d_i \end{pmatrix};$$

taking $\varepsilon \to 0$, we see that the closure of $\mathrm{GL}_d . M$ contains the module given by

$$x_i \mapsto \begin{pmatrix} a_i & 0 \\ 0 & d_i \end{pmatrix};$$

this module is isomorphic to $L \oplus M/L$. By induction, we see that the closure of $\mathrm{GL}_d . M$ contains a module isomorphic to $\operatorname{gr} M$.

It remains to show that the GL_d-orbit of a semisimple module is closed. □

We write

$$\mathrm{iss}_d A := \{\text{isomorphism classes of semisimple } d\text{-dimensional } A\text{-modules}\}.$$

The next result says that $\mathrm{iss}_d A$ can be realized as the closed points of an affine k-scheme.

Theorem 3.3 $\mathrm{rep}_d A /\!/ \mathrm{GL}_d(k) \cong \mathrm{iss}_d A$.

Proof. □

Example 3.4 Consider the d-dimensional representations of the polynomial ring $A = k[x]$. Then $\mathrm{rep}_d A = M_d(k)$ because for each matrix a there is a homomorphism $\phi : k[x] \to M_d(k)$ determined by $\phi(x) = a$. The $\mathrm{GL}_d(k)$ action on $\mathrm{rep}_d A$ is the conjugation action. The orbits are the conjugacy classes and these are in bijection with the isomorphism classes of d-dimensional $k[x]$-modules. By Proposition 3.1, the only conjugacy classes that are closed are those of the semisimple matrices. ◊

2.3.2 A functorial perspective

Schemes represent functors. A scheme S determines a functor

$$\begin{aligned} F : \{\text{Commutative Rings}\} &\to \mathsf{Sets} \\ R &\mapsto \text{Morphisms}(\operatorname{Spec} R, S) \end{aligned}$$

that sends a commutative ring R to the set of R-valued points of S.

Given a system of equations $\mathcal{F} \subset k\langle x_1, \ldots, x_n \rangle$, we can ask for solutions in $M_d(R)$ for any commutative k-algebra R. This gives a functor

$$\begin{aligned} S_d : \{\text{commutative } k\text{-algebras}\} &\to \mathsf{Sets} \\ R &\mapsto \{\text{solutions in } M_d(R) \text{ to } \mathcal{F} = 0\}. \end{aligned}$$

We prefer the module point of view: if $A = k\langle x_1, \ldots, x_n\rangle/(\mathcal{F})$, then the elements of $S_d(R)$ correspond to solutions in $d \times d$ matrices over R to the system of equations $\mathcal{F} = 0$. Accordingly, we consider a fixed k-algebra A and for each positive integer d consider the functor

$$\begin{aligned} S_d : \{\text{commutative } k\text{-algebras}\} &\to \mathsf{Sets} \\ R &\mapsto \mathrm{Hom}_{k-\mathrm{alg}}(A, M_d(R)). \end{aligned}$$

Our next result shows that this functor is represented by the scheme $\mathrm{rep}_d A$. To prove this we must introduce the ring of generic matrices.

Definition 3.5 Let $B = k[y_{ij}^{(t)}]$ be the commutative polynomial ring on the nd^2 indeterminates $\{y_{ij}^{(t)} \mid 1 \le t \le n, 1 \le i, j \le n\}$. The subring

$$k[Y_1, \ldots, Y_n]$$

of $M_d(B)$, generated by the matrices

$$Y_t = (y_{ij}^{(t)})_{1 \le i,j \le d}, \qquad 1 \le t \le n,$$

is called the ring of n generic $d \times d$ matrices over k. Each Y_t is called a generic matrix. ◊

The next lemma explains why we call these generic matrices.

Lemma 3.6 *Let R be a commutative k-algebra and $m_1, \ldots, m_n \in M_d(R)$. Let $k[Y_1, \ldots, Y_n]$ be the ring of n generic $d \times d$ matrices. There is an algebra homomorphism*

$$\Phi : k[Y_1, \ldots, Y_n] \to M_d(R)$$

such that $\Phi(Y_t) = m_t$ for $1 \le t \le n$.

Proof. For $1 \le t \le n$, write the $d \times d$ matrix m_t as $(r_{ij}^{(t)})_{1 \le i,j \le d}$ where each $r_{ij}^{(t)}$ is an element of R. There is of course a k-algebra homomorphism from the polynomial ring

$$\beta : B = k[y_{ij}^{(t)} \mid 1 \le t \le n,\ 1 \le i, j \le d] \to R, \qquad y_{ij}^{(t)} \mapsto r_{ij}^{(t)}.$$

By functoriality of $M_d(-)$ we obtain a ring homomorphism

$$M_d(\beta) : M_d(B) \to M_d(R)$$

and the image of Y_t under this is m_t. Hence the restriction of $M_d(\beta)$ to $k[Y_1, \ldots, Y_n]$ is the desired map Φ. □

Lemma 3.7 *Let B be a ring. Then every two-sided ideal in $M_d(B)$ is equal to $M_d(J)$ for some two-sided ideal J of B.*

Proof. □

Proposition 3.8 *Let A be a finitely generated k-algebra, and d a positive integer. There is a commutative ring A_d such that*

$$\mathrm{Hom}_{k-alg}(A_d, R) \cong \mathrm{Hom}_{k-alg}(A, M_d(R)) \tag{3-3}$$

for all commutative k-algebras R. In other words, the functor $S_d(-)$ defined above is represented by the affine scheme $\mathrm{rep}_d A$.

Proof. Fix a surjective k-algebra map $\psi : F = k\langle x_1, \ldots, x_n\rangle \to A$. Let

$$B = k[y_{ij}^{(t)} \mid 1 \leq t \leq n,\ 1 \leq i, j \leq d]$$

be the ring in Definition 3.5 and, as there, write Y_t, $1 \leq t \leq n$ for the generic matrices. Define $\theta : F \to M_d(B)$ by $\theta(x_t) = Y_t$. Let I be the two-sided ideal of $M_d(B)$ generated by $\theta(\ker \psi)$. There is an ideal, say J, of B such that $M_d(B)/I \cong M_d(B/J)$; let $\pi : B \to B/J$ be the natural map. Because $M_d(\pi) \circ \theta$ vanishes on $\ker \psi$, there is a unique k-algebra map $\rho : A \to M_d(B/J)$ such that

$$\begin{array}{ccc} F & \xrightarrow{\theta} & M_d(B) \\ {\scriptstyle\psi}\downarrow & & \downarrow{\scriptstyle M_d(\pi)} \\ A & \xrightarrow[\rho]{} & M_d(B/J) \end{array}$$

commutes.

Let $\phi : A \to M_d(R)$ be a k-algebra map. As in Lemma 3.6, there is a map $\beta : B \to R$ such that $M_d(\beta)(Y_t) = \phi(\psi(x_t))$, so

$$M_d(\beta) \circ \theta = \phi \circ \psi.$$

It follows that $M_d(\beta)$ vanishes on $\theta(\ker \psi)$, and hence on $I = M_d(J)$; hence $\beta(J) = 0$, and there is a unique map $\bar{\phi} : B/J \to R$ such that

$$\beta = \bar{\phi} \circ \pi.$$

A calculation now gives

$$\phi \circ \psi = M_d(\beta) \circ \theta = M_d(\bar{\phi}) \circ M_d(\pi) \circ \theta = M_d(\bar{\phi}) \circ \rho \circ \psi$$

whence $\phi = M_d(\bar{\phi}) \circ \rho$. □

Corollary 3.9 *The functor $A \mapsto A_d$, from k-algebras to commutative k-algebras, is left adjoint to the functor $R \mapsto M_d(R)$ from commutative k-algebras to k-algebras.*

Taking $R = A_d$, the adjoint isomorphism (3-3) gives us a distinguished map

$$j : A \to M_d(A_d)$$

from A to the ring of $d \times d$ matrices over a commutative ring. The map j has the following universal property: if R is a commutative k-algebra and $\phi : A \to M_d(R)$ is a k-algebra map, there is a unique k-algebra map $\bar{\phi} : A_d \to R$ such that the diagram

$$\begin{array}{ccc} A & \xrightarrow{j} & M_d(A_d) \\ {\scriptstyle =}\downarrow & & \downarrow{\scriptstyle M_d(\bar{\phi})} \\ A & \xrightarrow[\phi]{} & M_d(R) \end{array}$$

commutes.

Corollary 3.10 *Every d-dimensional A-module is annihilated by $\ker j$ where $j : A \to M_d(A_d)$ is the map above.*

Proof. The d-dimensional A-modules are given by maps $\phi : A \to M_d(k)$, so factor through $A \to A/\ker j \to M_d(k)$. □

Proposition 3.11 *There is an action of $G = \mathrm{GL}_d(k)$ as automorphisms of $M_d(A_d)$ such that $j(A) \subset M_d(A_d)^G$.*

Proof. Because G acts as automorphisms of $M_d(k)$ via conjugation it also acts as automorphisms of $M_d(k) \otimes_k A_d = M_d(A_d)$. Let $\pi_g : M_d(A_d) \to M_d(A_d)$ denote the action of $g \in G$. So $\pi_g \circ j : A \to M_d(A_d)$ and the universal property of j tells us there is a map $\psi_g : A_d \to A_d$ such that

$$\pi_g \circ j = M_d(\psi_g) \circ j.$$

If we now define the action of G on $M_d(A_d)$ by $\rho_g := M_d(\psi_g)^{-1} \circ \pi_g : M_d(A_d) \to M_d(A_d)$, then $\rho_g \circ j = j$ so $j(A)$ maps into the G-invariant part of $M_d(A_d)$ and the proof is complete *provided* $g \mapsto \rho_g$ really is a group homomorphism.

If $h \in G$, then

$$\pi_{gh} \circ j = \pi_g \circ \pi_h \circ j = \pi_g \circ M_d(\psi_h) \circ j;$$

but $\pi_g = \gamma_g \otimes \mathrm{id}_{A_d}$, where γ_g denotes conjugation by g, and $M_d(\psi_h) = \mathrm{id} \otimes \psi_h$, so $\pi_g \circ M_d(\psi_h) = M_d(\psi_h) \circ \pi_g$ and the computation continues

$$\pi_{gh} \circ j = M_d(\psi_h) \circ \pi_g \circ j = M_d(\psi_h) \circ M_d(\psi_g) \circ j = M_d(\psi_h \circ \psi_g) \circ j.$$

It follows that $\psi_{gh} = \psi_h \circ \psi_g$. A computation now shows that $\rho_g \rho_h = \rho_{gh}$ so G acts as automorphisms of A_d via ρ_g. □

When $\mathrm{char}\, k = 0$, there are reasonable hypotheses on A that ensure that $j(A)$ is equal to $M_d(A_d)^G$ [**?**, Thm. 2.6].

Our account in this section follows Procesi's book [**?**, Chapter IV] and his paper [**?**].

2.3.3 Polynomial Identity Rings

If we wish to classify, find, or understand, the d-dimensional A-modules we can replace A by $A/\ker j$ where $j : A \to M_d(A_d)$ is the universal map discussed in the previous section. Now $A/\ker j$ is a subalgebra of $M_d(A_d)$, and the latter ring is a finite module over its center. This does *not* imply that $A/\ker j$ is a finite module over its center.

The class of rings that are finite modules over their centers is not quite the "right" category to work with because, among other things, a subring of such a ring need not be finite over its center. The "right" category to work with is that of rings satisfying a polynomial identity.

Let k be a field and A a k-algebra. A polynomial identity for A is an non-zero element $f \in k\langle x_1, \ldots, x_n\rangle$ such that $f(a_1, \ldots, a_n) = 0$ for all $a_1, \ldots, a_n \in A$. For example, every commutative ring satisfies a polynomial identity, namely $xy - yx$. As we showed in Example 1.5, the ring of 2×2 matrices over any commutative ring satisfies the identity $(xy - yx)^2 z - z(xy - yx)^2$. Indeed, the ring of $d \times d$ over a commutative ring satisfies a polynomial identity. It is clear that subrings and quotients of rings satisfying polynomial identities satisfy polynomial identities.

The category of rings satisfying polynomial identities is a good class to work with. Rings in it are very close to being finite modules over their centers and behave in many ways as if they are.

One can construct a hierarchy of rings as follows. Define categories

$$\mathcal{C}_1 \subset \mathcal{C}_2 \subset \cdots$$

where $\mathcal{C}_d$ consists of the rings that satisfy all the polynomial identities satisfied by $M_d(\mathbb{Z})$. Thus $\mathcal{C}_1$ consists of the commutative rings, and each class of rings is slightly less commutative. If R is commutative, then $M_d(R)$ belongs to $\mathcal{C}_d$. If A belongs to $\mathcal{C}_d$ so does every subring of A and every quotient of it. Also $\mathcal{C}_d$ is closed under products. If A is any algebra, there is an ideal I_d minimal subject to the condition that $A/I_d \in \mathbb{C}_d$; take I_d to be the elements a that are contained in the kernel of every homomorphism $A \to M_d(R)$ as R ranges over all commutative rings. It follows that $I_d = \ker j$ where $j : A \to M_d(A_d)$ is the universal map above. One has $I_1 \supset I_2 \supset \cdots$ and hence surjective maps

$$\cdots A/I_3 \to A/I_2 \to A/I_1 = A/([A, A])$$

where $([A, A])$ denotes the ideal generated by all elements $[a, b] = ab - ba$.

There are important rings that do not belong to any $\mathcal{C}_d$, for example, $\mathbb{C}[x, y]$ with relation $xy - yx = 1$; the enveloping algebra of any non-abelian Lie algebra over a field of characteristic zero, the ring of differential operators on a smooth affine variety of dimension ≥ 1, the ring $\mathbb{C}[x, y]$ with relation $yx = qxy$ where $q \in \mathbb{C}$ is non-zero and not a root of unity, and so on.

However, the d-dimensional modules over a k-algebra A are the same as the d-dimensional modules over $A/\ker j \subset M_d(A_d)$, so $A/\ker j \subset \mathcal{C}_d$.

EXERCISES

3.1 Consider the $\mathrm{GL}_d(k)$ action on $\mathrm{rep}_d A$. Show that the stabilizer of M in $\mathrm{GL}_d(k)$ is isomorphic to $\mathrm{Aut}_A M$, the group of invertible A-module isomorphisms $M \to M$.

3.2 Show that although $M_2(\mathbb{R})$ is a finite module over its center, its subring

$$\begin{pmatrix} \mathbb{Q} & \mathbb{R} \\ 0 & \mathbb{Q} \end{pmatrix}$$

is not a finite module over its center.

3.3 Let $R = k[r_1, \ldots, r_n]$ be a finitely generated k-algebra and $I = Ra_1 + \cdots + Ra_m$ a finitely generated left ideal of R. Show that

$$A := \begin{pmatrix} k + I & R \\ I & R \end{pmatrix}$$

is a finitely generated k-algebra, generated by

$$\{e_{11}, e_{12}, a_i e_{21}, r_j e_{22} \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}.$$

If R is commutative show that the center of A is isomorphic to $k + I$.

3.4 Let $R = k[x, y]$ be the polynomial ring and $I = (x)$ the ideal generated by x. Show that $k + I$ is not noetherian, and hence not a finitely generated k-algebra, and that R is not a finitely generated module over $k + I$.

3.5 Let R and I be as in Exercise 4 and let A be as in Exercise 3. Show that A is not a finitely generated module over its center.

3.6 Continue the previous exercise. Show that $\mathrm{Fract}\, A = M_2(k(x, y))$.

3.7

2.4 Bergman's Diamond Lemma

In this section we describe an algorithm for finding a basis for an algebra

$$A = k\langle x_1, \ldots, x_d \rangle / (\mathcal{R})$$

when $\mathcal{R}$ is an explicitly given set of relations. The Diamond Lemma says that if the relations satisfy a particular condition then there is a set of words w_λ, $\lambda \in \Lambda$, such that $\{\text{words } w \mid \text{no } w_\lambda \text{ is a subword of } w\}$ is a basis for A. Unfortunately, a typical $\mathcal{R}$ will not satisfy this condition, and the algorithm need not terminate in a finite time.

As an indication of how little is known about algebras presented by generators and relations consider that although it is reasonable to expect that A is non-zero if $\mathcal{R}$ consists of a single relation of degree ≥ 1, this is only known in characteristic zero, and even then the proof is complicated [**?**], [**?**].

The idea behind the diamond lemma is apparent in the next example.

Example 4.1 Let $A = k[x, y]$ with defining relation $yx^2 = x^2y$. Thus $A = k\langle x, y \rangle / (yx^2 - x^2y)$. Our goal is to find a subset of the standard basis for $k\langle x, y \rangle$ that maps bijectively to a basis for A. To this end, we interpret the relation

as saying that any word in x and y having yx^2 as a subword, say $w = uyx^2v$, is not needed as a basis element for A since it becomes equal to ux^2yv in A. Hence A is spanned by (the images of) those words that do not contain yx^2 as a subword. It is reasonable to guess that the words that do not contain yx^2 as a subword give a basis for A—Theorem 4.5 tells us this is the case. The words that do not contain yx^2 as a subword are those that can be obtained from ll words in x and y by replacing a word of the form uyx^2v by the word ux^2yv, and repeating this process as often as possible. We call this process a 'replacement rule'.

Now suppose that $A = k[x, y]$ has two defining relations, $yx^2 = x^2y$ and $y^2x = xy^2$. Now there are two replacement rules, and either may be applied a word that contains y^2x^2 as a subword. This introduces a potential ambiguity, and some thought is required to see that the order of the replacement rules does not affect the final outcome. This section is devoted to explaining this. ◊

Let $F = k\langle x_1, \ldots, x_d\rangle$ be a free algebra, and let $\mathcal{B}$ be the standard basis consisting of the words in the x_i's. We define a total ordering on $\mathcal{B}$, the degree and dictionary ordering, by declaring that $u < v$ if either

- $\deg u < \deg v$, or
- $\deg u = \deg v$, say $u = x_{i_1} \cdots x_{i_n}$ and $v = x_{j_1} \cdots x_{j_n}$, and for some q, we have $i_1 = j_1, \cdots, i_q = j_q$, but $i_{q+1} < j_{q+1}$.

The only two properties of this ordering that will be used in the following analysis are

- if $u < v$, and $a, b \in \mathcal{B}$, then $aub < avb$, and
- for each $v \in \mathcal{B}$, $\{u \in \mathcal{B} \mid u < v\}$ is finite.

Let $\pi : F \to A$ be a surjective homomorphism. Suppose that $\ker \pi$ is generated as an ideal by

$$\mathcal{R} = \{r_\lambda \mid \lambda \in \Lambda\}.$$

let w_λ be the highest monomial (in the sense of $<$) occurring in r_λ. By replacing each r_λ by a suitable scalar multiple of itself we can assume that the coefficient of w_λ in r_λ is one, so we can write

$$r_\lambda = w_\lambda - s_\lambda$$

where all the words occuring in s_λ are $< w_\lambda$.

Lemma 4.2 *The relations r_λ may be chosen so that if λ and μ are distinct elements of Λ, then w_μ is not a subword of w_λ.*

Proof. If w_μ is a subword of w_λ, we may write $w_\lambda = aw_\mu b$. Hence we have the following equality of ideals:

$$(r_\mu, r_\lambda) = (w_\mu - s_\mu, w_\lambda - s_\lambda) = (w_\mu - s_\mu, s_\lambda - as_\mu b) = (r_\mu, s_\lambda - as_\mu b).$$

Each monomial occurring in $s_\lambda - as_\mu b$ is $< w_\lambda$. If we replace r_λ by $s_\lambda - as_\mu b$, then we still have a set of generators for $\ker \pi$. We repeat this process, and we will eventually stop because $(\mathcal{B}, <)$ has the descending chain property. When the process stops we will have obtained a set of defining relations satisfying the conditions in the lemma. □

From now on, assume that the relations $\mathcal{R} = \{r_\lambda \mid \lambda \in \Lambda\}$ are chosen as in Lemma 4.2. For each $\lambda \in \Lambda$ and each pair $(a, b) \in \mathcal{B} \times \mathcal{B}$, let

$$\rho_{ab}^{\lambda} : F \to F$$

be the linear map that on $\mathcal{B}$ is given by

$$\rho_{ab}^{\lambda}(w) = \begin{cases} as_\lambda b & \text{if } w = aw_\lambda b. \\ w & \text{otherwise.} \end{cases}$$

We call ρ_{ab}^{λ} a simple reduction. A composition of simple reductions, arising from various λ and various (a, b), is called a reduction. An element $x \in F$ is irreducible (with respect to $\mathcal{R}$) if $\rho(x) = x$ for all reductions. We define

$$\underline{\text{Irr}} = \{\text{irreducible elements}\}.$$

An element $x \in F$ is said to be reduction unique if there is a *unique* $y \in \underline{\text{Irr}}$ such that $\rho(x) = y$ for some reduction ρ; in this case we write $x_T = y$ (the T is for *terminal*).

Lemma 4.3 *Given the above situation,*

1. *$\underline{Irr}$ is a subspace of F with basis $\mathcal{B} \cap \underline{Irr} = \{w \in \mathcal{B} \mid$ no w_λ is a subword of $w\}$;*
2. *if $f \in F$, then $\rho(f) \in \underline{Irr}$ for some reduction ρ;*
3. *if ρ is a reduction, then $\rho(f) - f \in \ker \pi$ for all reductions ρ;*
4. *$F = \underline{Irr} + \ker \pi$;*
5. *if $f \in F$, and the reductions $\rho_1(f)$ and $\rho_2(f)$ are both irreducible, then $\rho_1(f) - \rho_2(f) \in \underline{Irr} \cap \ker \pi$;*
6. *the set of reduction unique elements is a subspace of F.*

Proof. (1) Since reductions are linear maps, $\underline{\text{Irr}}$ is a subspace. If $f \in \underline{\text{Irr}}$, then $\rho(f) = f$ for all simple reductions ρ. Since a simple reduction changes at most one monomial occurring in f, $\rho(w) = w$ for all monomials w occurring in f. Hence each such w is irreducible, and f is in the span of $\mathcal{B} \cap \underline{\text{Irr}}$.

(2) A simple reduction changes at most one monomial occurring in f, and replaces it with a strictly smaller monomial. Since $(\mathcal{B}, <)$ has the descending chain condition, repeating this process will eventually stabilize, thus giving an irreducible element.

(3) If ρ is a simple reduction, then $\rho(w) - w \in \ker \pi$ for all $w \in \mathcal{B}$. Summing over the words appearring in f, gives $\rho(f) - f \in \ker \pi$. Suppose, inductively, that if ρ' is a composition of n simple reductions, then $\rho'(f) - f \in \ker \pi$ for all f. If ρ is a simple reduction, then

$$\rho\rho'(f) - f = \rho\rho'(f) - \rho'(f) + \rho'(f) - f,$$

so this is in $\ker \pi$.

(4) This follows at once from (2) and (3).

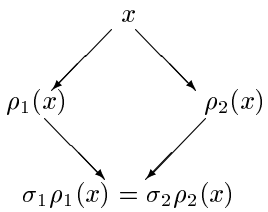
(5) We may write $\rho_1(f) - \rho_2(f) = \big(\rho_1(f) - f\big) - \big(\rho_2(f) - f\big)$, so the result follows from (1) and (3).

(6) A scalar multiple of a reduction unique element is reduction unique because reductions are linear maps. Now let x and x' be reduction unique. By (2), $\rho(x+x')$ is irreducible for some reduction ρ. Now $\rho(x)$ can be reduced to an irreducible element, but that element is also a reduction of x, so must equal x_T; say $\sigma\rho(x) = x_T$. Similarly, since x' is reduction unique, there is a reduction, τ say, such that $\tau\sigma\rho(x') = x'_T$. Notice that $\tau(x_T) = x_T$, because x_T is irreducible. Because $\rho(x + x')$ is irreducible,

$$\rho(x + x') = \tau\sigma\rho(x + x') = x_T + x'_T,$$

which proves the uniqueness of any irreducible element to which $x + x'$ can be reduced. In particular, $(x + x')_T = x_T + x'_T$. □

The word *"diamond"* refers to the idea of reduction uniqueness. If x is reduction unique, and $\rho_1(x)$ and $\rho_2(x)$ are reductions, then there exist reductions σ_1 and σ_2 such that $\sigma_1\rho_1(x) = \sigma_2\rho_2(x)$. We can draw a picture to illustrate this with the arrows going down to represent the idea that the reduction of an element is smaller with respect to $<$.



Definition 4.4 Let $a, b, c \in \mathcal{B}$. We call $[a, b, c]$ and ambiguity, and say that abc is ambiguous if $ab = w_\lambda$ and $bc = w_\mu$ for some $\lambda, \mu \in \Lambda$. This ambiguity is resolvable if there are reductions ρ and ρ' such that $\rho(s_\lambda c) = \rho'(as_\mu)$. If all ambiguities are resolvable, we say that $\{r_\lambda \mid \lambda \in \Lambda\}$ is a complete set of relations, or that the relations satisfy the diamond condition. ◊

The diagram (2.4) illustrates a resolvable ambiguity: there are two ways to reduce x, via ρ_1 and ρ_2, but we can apply further reductions, σ_1 and σ_2, to each of these and *get the same element*; the ambiguity is resolvable because the lower arrows converge at the same element thus forming a diamond.

Although the ambigous words are not the only ones that can be reduced in more than one way, they are the minimal such ones, and the next theorem says these are the only ones we need to worry about; this is good because if $\{r_\lambda \mid \lambda \in \Lambda\}$ is finite there are only finitely many ambiguities.

Theorem 4.5 (The Diamond Lemma) *[?] The following are equivalent:*

1. $\{r_\lambda \mid \lambda \in \Lambda\}$ *is a complete set of reductions;*
2. *all elements of* F *are reduction unique;*
3. $F = \underline{Irr} \oplus \ker \pi$*;*
4. $\{w \in \mathcal{B} \mid$ *no* w_λ *is a subword of* $w\}$ *is a basis for* A*.*

Proof. The equivalence of (3) and (4) is clear.

(1) $\Rightarrow$ (2) By Lemma 4.3(6), it suffices to prove that each $w \in \mathcal{B}$ is reduction unique. Suppose inductively that every word $< w$ is reduction unique. Let $\rho_1(w)$ and $\rho_2(w)$ be two distinct reductions. Each can be reduced to an irreducible element, and we must show that these two irreducible elements are the same. It suffices to do this when ρ_1 and ρ_2 are simple reductions. If $\rho_1 = \rho^\lambda_{**}$ and $\rho_2 = \rho^\mu_{**}$, then the proof depends on the relative positions of w_λ and w_μ in w.

Case 1. Suppose that $w = aw_\mu bw_\lambda c$ with $a, b, c \in \mathcal{B}$, so that $\rho_1(w) = \rho^\lambda_{bc}(w) = aw_\mu bs_\lambda c$ and $\rho_2(w) = \rho^\mu_{ab}(w) = as_\mu bw_\lambda c$. Since these elements are linear combinations of words $< w$, they are reduction unique by the induction hypothesis; both of them can be reduced to $as_\mu bs_\lambda c$, and hence to $(as_\mu bs_\lambda c)_T$. Thus $\rho_1(w)$ and $\rho_2(w)$ are reduction unique, and can be reduced to the same irreducible element, namely $(as_\mu bs_\lambda c)_T$.

Case 2. Suppose that $w = abcde$ with $a, b, c, d, e \in \mathcal{B}$, and $w_\mu = bc$ and $w_\lambda = cd$. Then $\rho_1(w) = \rho^\lambda_{be}(w) = abs_\lambda e$ and $\rho_2(w) = \rho(w)^\mu_{ad}(w) = as_\mu de$. As in case 1, these elements are reduction unique. By hypothesis (1), the ambiguity bcd is resolvable, so $s_\mu d$ and bs_λ can be reduced to a common element; hence so can $as_\mu de$ and $abs_\lambda e$; again, these elements are reduction unique, so we conclude that w is reduction unique.

(2) $\Rightarrow$ (3) By hypothesis (2), there is a well-defined linear map $\rho_T : F \to \underline{\mathrm{Irr}}$ that sends an element to the *unique* irreducible element to which it can be reduced. If $r \in \ker \pi$, then r is a linear combination of elements of the form $a(w_\lambda - s_\lambda)b$ with $a, b \in \mathcal{B}$. Since ρ^λ_{ab} kills this element, so does ρ_T. Thus $\ker \pi \subset \ker \rho_T$. If $f \in \ker \pi \cap \underline{\mathrm{Irr}}$, then $f = \rho_T(f) = 0$. Hence by Lemma 4.3(4), $F = \underline{\mathrm{Irr}} \oplus \ker \pi$.

(3) $\Rightarrow$ (1) Fix $w \in \mathcal{B}$. If $\rho_1(w)$ and $\rho_2(w)$ are irreducible reductions of w, then their difference is in $\underline{\mathrm{Irr}} \cap \ker \pi$ by Lemma 4.3(5) so, by hypothesis (3), this difference is zero. Hence w is reduction unique. □

If there is an ambiguity, say w, that can be reduced to distinct irreducible elements $\rho_1(w)$ and $\rho_2(w)$, then $\mathcal{R} = \{r_\lambda \mid \lambda \in \Lambda\}$ is not a complete set of relations. We therefore form enlarge our set of relations to

$$\mathcal{R}' := \mathcal{R} \cup \{r := \rho_1(w) - \rho_2(w)\};$$

we have $(\mathcal{R}') = (\mathcal{R}) = \ker \pi$, but the situation is better because w is a resolvable ambiguity with respect to $\mathcal{R}'$. Unfortunately, replacing $\mathcal{R}$ by $\mathcal{R}'$ will often introduce new ambiguities, and there is no guarantee that these new ambiguities are resolvable with respect to $\mathcal{R}'$. So, we keep repeating the process, enlarging the set of relations. Sadly, there is no guarantee that we will end up with a complete set of relations after a finite number of steps.

Example 4.6 Suppose that $\operatorname{char} k \neq 2$. Let $A = k[x, y, z]$ with defining relations

$$yz + zy = -y,\ xz + zx = x,\ xy + yx = z. \tag{4-4}$$

If we take the degree and dictionary ordering $x < y < z$, then

$$\begin{aligned} zy &= -y - yz, \\ zx &= x - xz, \\ yx &= z - xy \end{aligned}$$

gives the relations in the form $w_\lambda = s_\lambda$. The only ambiguity is zyx, and if we repeatedly apply the simple reductions to this we get

$$\begin{aligned} (zy)x &= -yx - yzx = -z + xy - y(x - xz) = -2z + 2xy + z^2 - xyz, \\ z(yx) &= z^2 - zxy = z^2 - xy + xzy = z^2 - 2xy - xyz. \end{aligned}$$

Thus the ambiguity zyx is not resolvable; taking the difference of the irreducible elements just produced leads to a new relation $xy = \frac{1}{2}z$ that must be added to the original three relations. Upon doing this, the ambiguity zyx becomes resolvable. But we get three new ambiguities, namely xyx, yxy, and zxy. Applying simple reductions to zxy gives

$$\begin{aligned} (zx)y &= xy - xzy = xy - x(-y - yz) = 2xy + xyz = z + \tfrac{1}{2}z^2, \\ z(xy) &= \tfrac{1}{2}z^2. \end{aligned}$$

Thus, zxy is a non-resolvable ambiguity, and are forced to adjoin the further relation $z = 0$. Now one sees from the relations (4-4), together with the relation $z = 0$, that A is defined by the three relations $x = 0$, $y = 0$, $z = 0$. The three relations $x = y = z = 0$ show us at once that $A \cong k$. $\Diamond$

To get a sense of how a "small" change in the relations can have a big effect, contrast the previous example with the next one.

Example 4.7 Let $B = k[x, y, z]$ with defining relations

$$yz + zy = y,\ xz + zx = x,\ xy - yx = z. \tag{4-5}$$

If we take the degree and dictionary ordering $x < y < z$, then

$$\begin{aligned} zy &= y - yz, \\ zx &= x - xz, \\ yx &= xy - z \end{aligned}$$

gives the relations in the form $w_\lambda = s_\lambda$. The only ambiguity is zyx, and if we repeatedly apply the simple reductions to this we get

$$(zy)x = yx - yzx = yx - y(x - xz) = yxz = xyz - z^2$$
$$z(yx) = zxy - z^2 = (-xz + x)y - z^2 = -x(-yz + y) + xy - z^2 = xyz - z^2.$$

Thus the ambiguity zyx is resolvable. It follows that $\{x^p y^q z^r \mid p, q, r \geq 0\}$ is a basis for B. $\Diamond$

EXERCISES

4.1 Find a basis for the algebra $k[x, y]$ with relation $xy = qyx$ where $0 \neq q \in k$.

4.2 Find a basis for the algebra $A = \mathbb{C}[x, \sigma]$ with relations $s^n = 1$ and $\sigma x = \xi x \sigma$ where ξ is a primitive n^{th} root of unity. Is this algebra a quotient of one of those in the previous exercise?

4.3 What is the center R of the algebra in the previous exercise? Find a basis for A as an R-module.

4.4 With the same notation, view A as a free left R-module and use the action of A on itself by left multiplication to obtain a ring homomorphism $A \to \operatorname{End}_R A$ and hence an embedding

$$A \to M_n(R).$$

Describe the image of this map as a subalgebra of $M_n(R)$.

4.5 Let R be a commutative ring and A a subalgebra of $M_n(R)$. Show that if I is a two-sided ideal of R, then $M_n(I)$ is a two-sided ideal of $M_n(R)$ and that

$$M_n(R)/M_n(I) \cong M_n(R/I).$$

In particular, $A/A \cap M_n(I)$ is a subalgebra of $M_n(R/I)$ and if $\mathfrak{m}$ is a maximal ideal of R, the composition

$$A \to M_n(R) \to M_n(R/\mathfrak{m}) \cong M_n(k)$$

produces an n-dimensional A-module. Hence, we obtain a family of n-dimensional A-modules parametrized by the closed points of $\operatorname{Spec} R$.

4.6 Use the idea in the previous exercise to find some modules for the algebra A in Exercise 2. Are the modules you found simple?

4.7 Fix $a, b, c \in \mathbb{C}$ and let $A = \mathbb{C}[x, y, z]$ have defining relations

$$axy + byx + cz^2 = ayz + bzy + cx^2 = azx + bxz + cy^2 = 0.$$

Do enough resolving of ambiguities and increasing the relations to drive yourself nuts. What do you suspect?

4.8 Find a basis for the algebra $A = \mathbb{C}[x, y, z, \sigma]$ with defining relations

$$\sigma^2 = 1,$$
$$\sigma x + x\sigma = \sigma y + y\sigma = \sigma z - z\sigma = 0,$$
$$xz - zx = yz - zy = 0,$$
$$xy - yx = \sigma z^n.$$

4.9 Fix a point $(\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$. Find a basis for the algebra $A = \mathbb{C}[x, y, \sigma]$ with defining relations

$$\sigma^{n+1} = 1,$$
$$\sigma x + x\sigma = \sigma y + y\sigma = 0,$$
$$xy - yx = \lambda_1 \sigma + \lambda_2 \sigma^2 + \cdots + \lambda_n \sigma^n.$$

2.5 Gelfand-Kirillov dimension

In this section k is a field and all rings are k-algebras.

Definition 5.1 Let A be a finitely generated k-algebra. A frame for A is a subspace V, containing 1, which generates A as a k-algebra. ◊

A frame V for A gives subspaces

$$k = V^0 \subset V \subset V^2 \subset V^3 \subset \cdots \subset \bigcup_{n=0}^{\infty} V^n = A$$

where V^n denotes the k-linear span of all words $v_1 \dots v_n$ where $v_i \in V$.

The next definition is a way of measuring the size of an algebra; it really measures how fast the algebra "grows", but it can be seen that this is a good measure of size by looking at a commutative polynomial ring. The idea is this. The polynomial ring $A = k[x_1, \dots, x_d]$ is generated by $V = k + kx_1 + \cdots + kx_d$ and a simple calculation shows that

$$\dim_k V^n = \binom{n+d}{d} = \frac{1}{d!} n^d + \text{lower degree terms},$$

so the number d can be obtained be saying that $\dim_k V^n$ grows like a polynomial of degree d.

Definition 5.2 The Gelfand-Kirillov dimension of a finitely generated k-algebra A is

$$\operatorname{GKdim} A := \limsup \frac{\log \dim_k V^n}{\log n}$$

taken over all frames V for A. ◊

If $\dim_k V^n$ grows like a polynomial of degree d, i.e., $\dim_k V^n \sim cn^d$, then

$$\frac{\log \dim_k V^n}{\log n} \sim d + \frac{\log c}{\log n}$$

so $\operatorname{GKdim} A = d$.

Example 5.3 If $m \geq 2$, then $\operatorname{GKdim} k\langle x_1, \dots, x_m \rangle = \infty$. If $V = k + kx_1 + \cdots + kx_m$, then $\dim_k V^n = m^n$, so as $n \to \infty$,

$$\frac{\log \dim_k V^n}{\log n} = \frac{n \log m}{\log n} \to \infty.$$

◊

The GK-dimension can be computed from any frame because if V and W are frames there are integers a and b such that $V \subset W^a$ and $W \subset V^b$, whence $V^n \subset W^{an}$ and

Lemma 5.4 *Let $\phi : A \to B$ be a homomorphism between finitely generated k-algebras. Then*

$$\operatorname{GKdim} A \text{ is } \begin{cases} \leq \operatorname{GKdim} B & \text{if } \phi \text{ is injective} \\ \geq \operatorname{GKdim} B & \text{if } \phi \text{ is surjective} \end{cases}$$

The following fundamental result of Bergman is proved in [**?**, Theorem ??].

Theorem 5.5 (Bergman) *If R is a finitely generated k-algebra, then*

- $\operatorname{GKdim} R \in \{0, 1\} \cup [2, \infty)$;
- *if* $\operatorname{GKdim} R = 1$*, then for each frame V there is a constant c such that* $\dim_k V^n \leq cn$ *for all n.*

Lemma 5.6 *Let A be a k-algebra that is a domain. If $Ax \cap Ay = 0$ for some non-zero $x, y \in A$, then A contains a copy of the free algebra $k\langle x, y\rangle$.*

Proof. Let $\varphi : F = k\langle X, Y\rangle \to A$ be the algebra map defined by $\varphi(X) = x$ and $\varphi(Y) = y$. Choose $0 \neq \alpha + aX + bY \in \ker \varphi$ of least degree, with $\alpha \in k$. Write $\bar{a} = \varphi(a)$ and $\bar{b} = \varphi(b)$. Then

$$\alpha y + y\bar{b}y = -y\bar{a}x \in Ax \cap Ay = 0,$$

so $\bar{a} = \alpha + y\bar{b} = 0$, whence $\alpha = \beta = 0$. But a and b have smaller degree than $\alpha + aX + bY$; this is a contradiction, so we conclude that φ is injective. □

Lemma 5.7 *Let A be a k-algebra that is a domain. If* $\operatorname{GKdim} A < \infty$*, then $Ax \cap Ay \neq 0$ for all non-zero $x, y \in A$.*

Proof. This follows from Lemma 5.6 and the fact that $\operatorname{GKdim} k\langle x, y\rangle = \infty$. □

Theorem 5.8 *Let A be a k-algebra and a domain. If* $\operatorname{GKdim} A < \infty$*, then A embeds in a division algebra D such that every element of D is of the form xy^{-1} with $x, y \in A$.*

Proof. Suppose first that A embeds in some division ring D. Then every non-zero element $x \in A$ has an inverse $x^{-1} \in D$.

Claim: if $x, y \in A$ with $y \neq 0$, then $x^{-1}y$ is equal to cd^{-1} for some $c, d \in A$ with $d \neq 0$. By Lemma 5.7, $xA \cap yA \neq 0$, so there are elements $c, d \in A$ such that $yc = xd$; hence, in D, $x^{-1}y = dc^{-1}$.

We shall show first $\{wx^{-1} \mid w, x \in A,\ x \neq 0\}$ is a subring of D. Certainly this subset contains A so will be a k-subalgebra if it is closed under multiplication and addition. It is closed under multiplication because $wx^{-1}.yz^{-1} = wdc^{-1}z^{-1} = (wd)(zc)^{-1}$ for some $c, d \in A$. It is closed under addition: given $wx^{-1} + yz^{-1}$, $xA \cap zA \neq 0$, so there are non-zero elements $a, b \in A$ such that $xa = zb = t$, whence $wx^{-1} + yz^{-1} = (wa + yb)z^{-1}$.

□

We call the division algebra in Theorem 5.8 the ring of fractions of R and denote it by Fract R.

The condition that a ring be a domain is somewhat too restrictive in the non-commutative setting. The domain condition is appropriate in the commutative setting because coordinate rings of irreducible varieties are domains. In the non-commutative setting we make a small change to the definition of a prime ideal (see chapter ???) and prime rings are the appropriate non-commutative analogue of domains (for example, the matrix algebras $M_n(k)$ are prime rings).

A result of fundamental importance in non-commutative algebra is Goldie's theorem—a prime noetherian ring R has a ring of fractions: there is a division ring D and an injective homomorphism $R \to M_n(D)$ for some n such that every element of $M_n(D)$ is of the form wx^{-1} for some $w, x \in R$ with x regular (i.e., x is not a zero-divisor). The proof of Goldie's Theorem is quite different from that of Theorem 5.8.

2.5.1 GK-dimension of modules

Definition 5.9 Let A be a k-algebra generated by a finite dimensional subspace V that contains 1. Let M be a left A-module that is generated by a finite dimensional subspace W. We define

$$\operatorname{GKdim} M := \limsup \frac{\log \dim_k V^n W}{\log n}$$

◊

This is compatible with the previous definition in the sense that the GK-dimension of A when computed by viewing A as a left A-module is the same as its GK-dimension when computed as an algebra. Indeed, if $\phi : A \to B$ is a homomorphism of k-algebras such that B is becomes a finitely generated A-module, then the GK-dimension of B as a left A-module is the same as $\operatorname{GKdim} B$ computed by viewing B as a k-algebra.

The following is a consequence of this observation.

Proposition 5.10 *If B is a finitely generated commutative k-algebra, then*

$$\operatorname{GKdim} B = \operatorname{Kdim} B = \dim X,$$

where $\operatorname{Kdim} B$ *denotes the Krull dimension of B and X is the affine variety determined by B.*

Lemma 5.11 *Let $\phi : M \to N$ be a homomorphism between finitely generated A-modules. Then*

$$\operatorname{GKdim} M \text{ is } \begin{cases} \leq \operatorname{GKdim} N & \text{if } \phi \text{ is injective} \\ \geq \operatorname{GKdim} N & \text{if } \phi \text{ is surjective} \end{cases}$$

It follows that $\mathrm{GKdim}(M \oplus N) = \max\{\mathrm{GKdim}\, M, \mathrm{GKdim}\, N\}$. Hence the GK-dimension of a non-zero finitely generated free A-module is equal to $\mathrm{GKdim}\, A$, and $\mathrm{GKdim}\, M \leq \mathrm{GKdim}\, A$ for all finitely generated A-modules M.

For many reasonable algebras if $0 \to L \to M \to N \to 0$ is an exact sequence of modules one has

$$\mathrm{GKdim}\, M = \max\{\mathrm{GKdim}\, L, \mathrm{GKdim}\, N\}.$$

Lemma 5.12 *Let R and S be k-algebras, and M an R-S-bimodule. If ${}_RM$ is finitely generated, then*

1. $\mathrm{GKdim}\, M_S = \mathrm{GKdim}\, S/\ \mathrm{Ann} M$;
2. $\mathrm{GKdim}\, M_S \leq \mathrm{GKdim}\, {}_RM$;
3. *if M_S is finitely generated, then* $\mathrm{GKdim}\, M_S = \mathrm{GKdim}\, {}_RM$.

Proof. [?, Lemma 5.3] □

It follows from Lemma 5.12 that Morita equivalent k-algebras have the same GK-dimension. This allows us to unambiguously define the dimension of an affine nc-space as the GK-dimension of any one of its cooordinate rings.

EXERCISES

5.1 Let k be an algebraically closed field. Fix $\lambda \in k\backslash\{0,1\}$ and consider the ring $A = k\langle x, y\rangle/(yx - \lambda xy)$ corresponding to the equation $yx = \lambda xy$. Recall that A has basis $\{x^i y^j \mid i, j \geq 0\}$ (Exercise 1).

(a) Let M be a left A-module. Show that $\{m \in M \mid xm = 0\}$ is a submodule.

(b) Show that if λ is not a root of unity every finite dimensional simple A-module is 1-dimensional. [*Hint:* consider determinants.]

(c) Assume that λ is not a root of unity, and find all 1-dimensional A-modules.

(d) Assume that λ is not a root of unity. Show that if M is a d-dimensional A-module then $(xy)^d M = 0$.

(e) Fix $\alpha \in k$, and define the A-module $M(\alpha)$ as follows: as a k-vector space $M = k[t]$ is the polynomial ring, and the A-module action on $f = f(t) \in M$ is defined by

$$x.f = \alpha t f \quad \text{and} \quad y.f = tf(\lambda t).$$

Show that the ideals (t^r) are A-submodules.

(f) Show that $M(\alpha) \cong A/A(x - \alpha y)$.

5.2 Continue the previous exercise, but assume that λ is a primitive n^{th} root of 1.

(a) Show that the module $M(\alpha)$ in Exercise 1 is isomorphic to its submodule (t^n) via the A-module map defined by $\varphi(f) = t^n f$.

(b) Since $\varphi \in \mathrm{End}_A\, M(\alpha)$ so too is $1 - \beta\varphi \in \mathrm{End}_A\, M(\alpha)$ for all $\beta \in k\backslash\{0\}$. The image of this is a submodule, so

$$N(\alpha, \beta) := M(\alpha)/(1 - \beta\varphi)(M(\alpha))$$

is an n-dimensional A-module. In particular, the actions of x and y on this yield solutions in $M_n(k)$ to $yx = \lambda xy$. What are the matrices for x and y with respect to the basis for $N(\alpha, \beta)$ given by the images of $1, t, \ldots, t^{n-1}$?

(c) Show that $N(\alpha, \beta)$ is simple if and only if $\alpha \neq 0$.

5.3 Continue the previous exercise. Let $Z(A)$ denote the center of A.

(a) Show that the subalgebra $k[x^n, y^n]$ generated by x^n and y^n lies in $Z(A)$.

(b) Show that A is a free module over $k[x^n, y^n]$ of rank n^2.

(c) Show that $Z(A) = k[x^n, y^n]$.

(d) Show that if $\mathfrak{m}$ is a maximal ideal of $Z(A)$, then $\dim_k(A/\mathfrak{m}A) = n^2$.

5.4 Continue the previous exercise.

(a) If M is a finite dimensional simple A-module, show that $\operatorname{End}_A M = k$, and hence that M is annihilated by a maximal ideal of $Z(A)$.

(b) Show that every finite dimensional simple A-module is of dimension $\leq n$. *Hint:* If M is a simple A-module of dimension d, then $A/\operatorname{Ann} M \cong M_d(k)$.

(c) What is the maximal ideal of the center that annihilates $N(\alpha, \beta)$?

(d) Show that the 1-dimensional modules together with the modules $N(\alpha, \beta)$ with $\alpha\beta \neq 0$ are all the finite dimensional simple A-modules.

5.5 [Irving-Small] Let $A = k\langle x, y\rangle/(x)^2$. Use the Diamond lemma to find a basis for A. Observe that A is a graded ring with $\deg x = \deg y = 1$, and show that $\dim_k A_n = n + 1$ (as for the polynomial ring in two variables). Show that A contains an infinite direct sum of left ideals, so is not noetherian.