Modifying Lax Equations
and the Second Hamiltonian Structure

B.A. Kupershmidt[1] and George Wilson[2]

[1] University of Michigan, Ann Arbor, MI 48109, USA
[2] Mathematical Institute, 24–29 St Giles, Oxford OX1 3LB, England

1. Introduction

The main result of this paper concerns the second Hamiltonian structure for the Lax equations based on a scalar n^{th} order differential operator

$$(1.1) \qquad L = \xi^n + u_{n-2}\,\xi^{n-2} + \ldots + u_1\,\xi + u_0, \qquad \xi \equiv \partial/\partial x$$

(see [1, 5]). We show that this complicated-looking structure arises by 're-duction' from a vastly simpler one (essentially just $\partial/\partial x$) on an appropriate space of 'modified' variables.

We begin by explaining this in the simplest case $n = 2$, so that L is the Schrödinger operator $\xi^2 + u$; here u is a function of two variables x and t. By a Lax equation formed from L we mean an equation of the form

$$(1.2) \qquad L_t = [P_+, L]$$

where P_+ is a differential operator whose coefficients are differential polynomials in u, that is, polynomials in u and its x-derivatives $u^{(j)}$. (The subscript $+$ may be ignored at this point: we introduce it so as not to conflict with the notation in the main body of the paper.) Since L_t is an operator of order zero, for (1.2) to make sense P_+ must be chosen so that the commutator on the right has order zero too; (1.2) is then equivalent to an evolution equation for u, that is, an equation of the form

$$u_t = f(u, u_x, u_{xx}, \ldots)$$

(no t-derivatives on the right). The construction of such operators P_+ is now well understood, and is reviewed in Sect. 3. The simplest non-trivial example is the operator

$$P_+ = 4\,\xi^3 + 6\,u\,\xi + 3\,u_x;$$

the corresponding Lax equation (1.2) is the Korteweg-de Vries (KdV) equation

(1.3) $$u_t = u_{xxx} + 6 u u_x.$$

Equation 1.2 is called the 'Lax representation' of the KdV equation.

Among the many remarkable properties of the KdV equation, we shall be specially concerned with the following.

(i) *The equation has an infinite sequence* $H_1, H_2, \ldots$ *of conserved densities; they are differential polynomials in* u, *and we have* $H_1 = 4 u$, $H_2 = u^2$, $H_3 = \frac{1}{2} u^3 - \frac{1}{4} u_x^2, \ldots$.

(ii) *The conserved densities satisfy the recursion relation (due to A. Lenard)*

(1.4) $$(\tfrac{1}{2} \partial^3 + u \partial + \partial u) \frac{\delta H_q}{\delta u} = 2 \partial \frac{\delta H_{q+1}}{\delta u}.$$

(iii) *The equation can be written in Hamiltonian form*

$$u_t = 2 \partial \frac{\delta H_3}{\delta u}.$$

First a few words of explanation for non-expert readers. In (ii) and (iii), ∂ means $\partial / \partial x$: our principle is that in differential operators we write $\partial / \partial x$ as ξ when we are thinking of the operator as an algebraic object in its own right, as in (1.1), but as ∂ when it is actually going to operate on something, as in (ii) and (iii). In (ii), ∂u means the operator: multiply by u, then differentiate. 'Conserved densities' means that we have identities

$$\frac{\partial H_q}{\partial t} = \partial J_q$$

which follow formally from (1.3); the J_q are also differential polynomials in u. For example, for $H_2 = u^2$ we have

$$\partial / \partial t (u^2) = 2 u (u_{xxx} + 6 u u_x) = \partial (2 u u_{xx} - u_x^2 + 4 u^3).$$

Under suitable analytic circumstances, for example if all the $u^{(j)}$ vanish as $x \to \pm \infty$, if follows at once that the integrals of the H_q will be constants (independent of t). In (ii) and (iii), $\delta / \delta u$ denotes the (formal) variational derivative (Euler-Lagrange operator)

$$\frac{\delta H}{\delta u} = \sum_{i \geqq 0} (-\partial)^i \frac{\partial H}{\partial u^{(i)}}.$$

We refer to Sect. 5 for a precise explanation of the term 'Hamiltonian' in (iii), but the idea is as follows: for any differential polynomial H we think of the equation

$$u_t = 2 \partial \frac{\delta H}{\delta u}$$

(or rather the corresponding derivation $\partial/\partial t$) as the Hamiltonian vector field corresponding to H; we are thus thinking of H as a function on phase space, $\delta H/\delta u$ as its gradient, and the skew operator 2∂ is playing the role of the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ in the usual form of Hamilton's equations. The term 'Hamiltonian' then refers to the following property of the operator 2∂: if we define Poisson brackets in the usual way, then this map (functions)$\rightarrow$(vector fields) takes Poisson brackets to commutators.

Let us look more closely at the Lenard relations (1.4). Given an arbitrary differential polynomial H, we can form the expression $(\frac{1}{2}\partial^3 + u\,\partial + \partial u)\,\delta H/\delta u$, but in general there is no reason why this expression should lie in the image of the operator $2\partial\,\delta/\delta u$; it is thus by no means obvious that, given H_q, we can find any differential polynomial H_{q+1} satisfying (1.4). However, that is in fact the case for the whole infinite sequence of conserved densities H_q. The papers [6, 10] throw considerable light on this miracle: it is closely connected with the fact that the operator $\frac{1}{2}\partial^3 + u\partial + \partial u$ on the left of (1.4) is Hamiltonian in the same sense as indicated above for the operator 2∂ on the right. (Perhaps we should emphasize that this property is by no means shared by all skew-adjoint differential operators: for example, the operator $\frac{1}{2}\partial^3 + u^2\,\partial + \partial u^2$ is not Hamiltonian; nor are any of the operators $\alpha\,\partial^{2r+1} + u\partial + \partial u$ for $r > 1$, α a non-zero constant.) The KdV equation can thus be written in *two* Hamiltonian forms:

$$u_t = 2\partial \frac{\delta H_3}{\delta u} = (\tfrac{1}{2}\partial^3 + u\,\partial + \partial u) \frac{\delta H_2}{\delta u}.$$

Furthermore, these two Hamiltonian structures are compatible in the sense that for any constants α, β, the operator $\alpha(2\partial) + \beta(\frac{1}{2}\partial^3 + u\partial + \partial u)$ is Hamiltonian. (That is not automatic, since the requirement of being Hamiltonian imposes a quadratic, not linear, condition on the skew operator.) In [6] (see also [10]) it is shown that this is sufficient to ensure that the equation has an infinite sequence of conserved densities satisfying (1.4).

The main question, then, is: why is the operator $\frac{1}{2}\partial^3 + u\partial + \partial u$ Hamiltonian? Our explanation for this involves the relationship between the KdV equation and the 'modified' KdV equation

(1.5) $$v_t = v_{xxx} - 6v^2\,v_x.$$

This relationship played an important role in the original discovery of the properties of the KdV equation (see, for example, [8]). The connexion between the two equations, due to Miura, is the following: let

(1.6) $$u = v_x - v^2.$$

Then if v satisfies (1.5), u satisfies (1.3).

Now, the modified KdV equation can be written in the Hamiltonian form

(1.7) $$v_t = (-\tfrac{1}{2}\partial) \frac{\delta H}{\delta v}$$

where $H = u^2 = (v_x - v^2)^2$. As we have seen, the KdV equation can be written in the form

(1.8) $$u_t = (\tfrac{1}{2}\partial^3 + u\partial + \partial u)\frac{\delta H}{\delta u}$$

with the same Hamiltonian $H = u^2$. This connexion between Eqs. 1.7 and 1.8 is not an accident involving the particular Hamiltonian $H = u^2$, but reflects a relationship between the two skew operators $-\tfrac{1}{2}\partial$, $\tfrac{1}{2}\partial^3 + u\partial + \partial u$. If v satisfied an Eq. (1.7) for any other Hamiltonian $H(u, u_x, \ldots)$, then $u = v_x - v^2$ would satisfy the corresponding Eq. (1.8). That is a simple consequence of the easily verified identity

(1.9) $$\tfrac{1}{2}\partial^3 + u\partial + \partial u = D(-\tfrac{1}{2}\partial)D^*$$

where $D = \partial - 2v$ is the 'Fréchet Jacobian' of u with respect to v, and $D^* = -\partial - 2v$ is its (formal) adjoint. It follows automatically from (1.9) that the operator $\tfrac{1}{2}\partial^3 + u\partial + \partial u$ is Hamiltonian (given that $-\tfrac{1}{2}\partial$ is). We say that the Hamiltonian structure defined by the operator $\tfrac{1}{2}\partial^3 + u\partial + \partial u$ is obtained by restriction of that defined by $-\tfrac{1}{2}\partial$ (from functions of v to functions of u). (The term 'reduction' would be appropriate if we were thinking in a geometric, rather than algebraic, context.) We refer to Sect. 6 for a detailed explanation of all this.

We shall generalize everything we have said so far to all values of n. Here we just point out the two main clues as to how the generalization is to be done. First, if we formally factorize the Schrödinger operator

$$\xi^2 + u = (\xi - v)(\xi + v)$$

then u and v are related by the 'Miura transformation' (1.6). We are not sure to whom this observation should be credited, but it is certainly to be found in the paper of Adler and Moser [16]. Secondly, the modified KdV equation has a matrix Lax representation in which

$$L = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\xi + \begin{pmatrix} 0 & v \\ -v & 0 \end{pmatrix}.$$

In the general case we shall introduce 'modified' variables $v_1, \ldots, v_{n-1}$ by splitting the operator (1.1) into linear factors, and the modified Lax equations will be based on a first order operator with $n \times n$ matrix coefficients.

Here are some remarks to orient the reader among the various sections. The modified Lax equations are defined in Sect. 4; the definition depends on some very simple matrix algebra involving 'circulants', which is set out in Sect. 2. Our main theorem on the second Hamiltonian structure is in Sect. 8. The intervening sections are of a trivial and/or expository nature. Section 3 reviews the general theory of Lax equations (following [15]) and discusses 'specialization' of the basic operator L. The point is that in the general theory the coefficients of L are supposed to be independent variables; but in practice we often want to consider operators for which that is not the case. Some parts of the general theory carry over automatically to a specialized L, but others cause problems. Sections 5–7 could be regarded as a quick introduction to the Hamiltonian formalism.

Section 5 contains definitions and examples. We have formulated the definitions in the minimum generality adequate for our purposes; for example, there would be no difficulty in working with, say, C^∞ functions rather than polynomials, which would put us essentially in the framework of [11], Chap. 1. Section 6 discusses the process of restricting Hamiltonian structures. We should emphasize that a structure obtained by restriction is automatically Hamiltonian (preserves brackets); in Sect. 8 we show that the second Hamiltonian structure for Lax equations arises in this way, so we obtain a proof that this structure actually is Hamiltonian; this proof is quite different from that of Gel'fand and Dikii [5]. The compatibility of the two Hamiltonian structures for Lax equations (see [5, 6, 10]) is a trivial consequence of the fact that the second structure is Hamiltonian (see Sect. 5, Example 6), so we have a new proof of that too. Section 7 explains why Lax equations have these two Hamiltonian structures. The exposition essentially follows [11], with appropriate modifications to take care of the matrix case (which we need).

Sections 9 and 10 give the proofs of two comparatively technical results used in Sect. 8. Then finally, in Sect. 11 we consider what happens if we abandon the assumption that the second coefficient u_{n-1} of the operator (1.1) should vanish. The restriction $u_{n-1} = 0$ is a natural one for many purposes, but for the discussion of the second Hamiltonian structure the general case $u_{n-1} \neq 0$ is actually easier (compare Examples 4 and 5 in Sect. 5). In the case $u_{n-1} \neq 0$, our main result can be formulated very simply as follows. Recall that we are dealing with Hamiltonian structures on spaces of coefficients of differential operators. Then the second Hamiltonian structure is characterized by the following properties:

(i) for a first order operator $\xi + x$, it is just $-\partial$
(ii) the composition of operators is a 'canonical transformation'.

We mention briefly two other applications of the idea of 'modification'. The first concerns 'Bäcklund transformations': as is well known (see [4]), in the KdV case $(n=2)$ these are closely connected with the fact that the modified KdV equation (1.3) is invariant under $v \mapsto -v$. For general n, our modified equations are invariant under $v_i \mapsto \omega^i v_i$, where ω is an nth root of unity. Second, one can discuss systematically the hierarchies of Lax equations based on certain specialized operators L obtained by imposing relations among the 'roots' of the basic operator (1.1). The simplest examples would be when $n=3$; then we should obtain hierarchies of Lax equations based on the operators

$$L = \xi^3 + u\,\xi + \alpha\,u_x, \qquad \alpha = 0, \tfrac{1}{2} \text{ or } 1.$$

(The three values of α correspond to the three roots of L vanishing. The equations with $\alpha = 0$ or 1 coincide, so there are only two hierarchies.)

We should like to draw the reader's attention to two recent short papers [17, 18] which are closely related to our work. We saw these papers only during the revision of our manuscript (August 1980).

2. Circulants

Let A be an associative, but not necessarily commutative, algebra over $\mathbb{C}$ (in our application A will be an algebra of formal pseudo-differential operators). Let $M_n = M_n(A)$ be the algebra of all $n \times n$ matrices with entries in A. We consider the map $\varphi: M_n \to M_n$ defined by

$$(\varphi(X))_{ij} = X_{i+1, j+1}, \qquad X \in M_n.$$

(Here the subscripts are to be read $\bmod n$.) It is easy to check that φ is an algebra automorphism: indeed, φ is just the inner automorphism of M_n induced by cyclic permutation of the standard basis for A^n. The matrices left fixed by φ are called *circulants* (see [2]). For example, a 3×3 circulant is a matrix of the form

$$\begin{pmatrix} a & b & c \\ c & a & b \\ b & c & a \end{pmatrix}.$$

More generally, we want to consider the decomposition of M_n according to the different characters of the cyclic group of order n generated by φ (obviously, φ^n is the identity). To make this explicit, we fix a primitive n^{th} root of unity, say

$$\omega = \exp(2\pi i/n).$$

(2.1) *Definition.* We say a matrix $X \in M_n$ is an ω^k-*circulant* if

$$\varphi(X) = \omega^k X.$$

More explicitly still, X is an ω^k-circulant if it can be obtained from a circulant by multiplying the ith row (or column) by ω^{ik} (we index the rows and columns by the numbers $0, 1, \ldots, n-1$). For example, a 3×3 ω-circulant is a matrix of the form

$$\begin{pmatrix} a & b & c \\ \omega c & \omega a & \omega b \\ \omega^2 b & \omega^2 c & \omega^2 a \end{pmatrix}.$$

(2.2) **Proposition.** (i) *Each $X \in M_n$ has a unique decomposition*

$$X = X_0 + X_1 + \ldots + X_{n-1}$$

with X_k an ω^k-circulant.

(ii) *The product of an ω^k-circulant and an ω^l-circulant is an ω^{k+l}-circulant.*

Briefly, we have a $\bmod n$ grading on M_n (as indicated above, the grading really takes values in the character group of the cyclic group generated by φ).
We introduce the following notation:

$$X = \omega^k\text{-circ}(x_0, x_1, \ldots, x_{n-1})$$

means that X is the (unique) ω^k-circulant whose first row is the vector indicated. Let L be an ω-circulant, say

$$L = \omega\text{-circ}(v_0, v_1, \ldots, v_{n-1}).$$

For $k = 0, 1, \ldots, n-1$ we define circulants L_k by

$$L_k = \text{circ}(v_0, \omega^k v_1, \omega^{2k} v_2, \ldots, \omega^{(n-1)k} v_{n-1}).$$

If Ω denotes the diagonal matrix

(2.3) $$\Omega = \text{diag}(1, \omega, \omega^2, \ldots, \omega^{n-1})$$

it is easy to check that

$$L = \Omega L_0, \qquad L_k = \Omega^{-k} L_0 \Omega^k.$$

Using this, and the fact that Ω^n is the identity, we get the next proposition.

(2.4) **Proposition.** *We have*

$$L^n = L_{n-1} \ldots L_2 L_1 L_0.$$

This can be regarded as a non-commutative version of the well known factorization for the determinant of a circulant in the case where A is commutative (see [2]).

Finally, if $X = \text{circ}(x_0, \ldots, x_{n-1})$ is a circulant, we set $s(X) = \sum_0^{n-1} x_k$.

(2.5) **Proposition.** *The map*

$$s: \text{(circulants)} \to A$$

is a homomorphism of algebras.

Indeed, $s(X)$ is just the eigenvalue of X corresponding to the eigenvector $(1, 1, \ldots, 1)^t$.

3. Lax Equations and the Problems of Specialization

Let A now be an associative differential algebra over $\mathbb{C}$; that is, an algebra together with a derivation $\partial: A \to A$. We denote by $A[\xi]$ the algebra of (ordinary) differential operators with coefficients in A: thus each element of $A[\xi]$ has a unique representation in the form $\sum_0^r a_i \xi^i$, $a_i \in A$, and the rules for multiplying two such expressions follow from the basic one

(3.1) $$\xi a = a \xi + (\partial a), \qquad a \in A.$$

If A is an algebra of functions of x and $\partial = \partial/\partial x$, then $A[\xi]$ is just the usual algebra of differential operators. We denote by $A[\xi, \xi^{-1}]$ the algebra of formal

pseudo-differential operators obtained from $A[\xi]$ by formally inverting ξ. Each element X of $A[\xi, \xi^{-1}]$ has a unique representation in the form

(3.2) $$X = \sum_{-\infty}^{r} x_i \xi^i, \qquad x_i \in A.$$

The rules for multiplying these expressions also follow from (3.1): for example, multiplying (3.1) on the right and left by ξ^{-1}, we easily find

$$\xi^{-1} a = \sum_{i=0}^{\infty} (-1)^i \partial^i a \, \xi^{-i-1}, \qquad a \in A.$$

(This rule explains why it is essential to allow infinite sums in $A[\xi, \xi^{-1}]$.) It is easy to check that the multiplication in $A[\xi, \xi^{-1}]$ is associative (see, for example, [11]). Each element $X \in A[\xi, \xi^{-1}]$ has a unique decomposition

$$X = X_+ + X_-$$

where X_+ is a differential operator and X_- is an 'integral operator' (involving only negative powers of ξ). If we write X in the form (3.2), then we have

$$X_+ = \sum_{0}^{r} x_i \xi^i, \qquad X_- = \sum_{-\infty}^{-1} x_i \xi^i.$$

Finally, the coefficient of ξ^{-1} in the expansion (3.2) is called the *residue* of X, written res X.

Throughout this paper we shall be working with differential algebras of the form

$$B = \mathbb{C}[w_1^{(j)}, \ldots, w_N^{(j)}], \qquad j \geqq 0$$

(polynomials in independent variables w_i and their 'derivatives'; the derivation is defined by $\partial w_i^{(j)} = w_i^{(j+1)}$, and as usual $w_i \equiv w_i^{(0)}$, so that $w_i^{(j)} = \partial^j w_i$). When it is necessary to indicate the variables we shall write this algebra as $B(w_1, \ldots, w_N)$, or sometimes just $B(w)$. We write $M_l(B)$ for the algebra of $l \times l$ matrices with entries in $B(w)$; the derivation ∂ is extended to $M_l(B)$ so as to act on each entry separately.

The theory of Lax equations starts out from a differential operator $L \in M_l(B)[\xi]$:

(3.3) $$L = u_n \xi^n + u_{n-1} \xi^{n-1} + \ldots + u_1 \xi + u_0, \qquad u_i \in M_l(B).$$

We assume that the two leading coefficients satisfy the following conditions.

(3.4)(i)   *The first coefficient u_n is a constant diagonal matrix*

$$u_n = \operatorname{diag}(c_1, \ldots, c_l)$$

where the c_α are all non-zero, and $c_\alpha \neq c_\beta$ if $\alpha \neq \beta$.

(3.4)(ii) *The diagonal entries in the second coefficient vanish:*

$$u_{n-1,xx} = 0.$$

The reason for these assumptions (apart from the fact that they are satisfied in all the examples of physical interest) will become clear below. The assumption that the c_x be distinct is in fact quite inessential; we make it just to introduce some verbal simplifications in what follows. The case when some or all of the c_x are equal is treated in [15].

In the general theory of Lax equations (for example in [15]) it is assumed in addition that the (non-constant) entries $u_{i,x\beta}$ in L are differentially independent, that is, that there are no polynomial relations among the $u_{i,x\beta}^{(j)}$. In that case one might as well start off from the algebra $B(u_{i,x\beta})$, as is done in [15]. We shall refer to an operator (3.3) in which the $u_{i,x\beta}$ are differentially independent as a 'general' L. On the other hand, in practice we often want to form Lax equations from an operator L in which the $u_{i,x\beta}$ are not differentially independent; we shall refer to such an operator as a 'specialization' of the general L. (We prefer this to the commonly used term 'reduction'.) Formally, we could define a specialization to be a homomorphism of differential algebras

$$\varphi : B(u_{i,x\beta}) \to B(w_1, \ldots, w_N),$$

but since such a homomorphism is determined by its values on the $u_{i,x\beta}$, and these may be chosen arbitrarily, giving φ is equivalent simply to writing down an operator of the form (3.3). Perhaps the simplest example of a 'specialized' L is the operator

(3.5)
$$L = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \xi + \begin{pmatrix} 0 & v \\ -v & 0 \end{pmatrix}.$$

This L gives rise to the hierarchy of modified KdV equations. But since the entries v, $-v$, are not differentially independent, the theory of [15] does not apply immediately.

Let us review the general theory, paying attention to any problems caused by specialization. Given an operator L of the form (3.3), the associated Lax equations have, by definition, the form

(3.6)
$$\partial_t L = [P_+, L] = [L, P_-],$$

where $P \in M_l(B)[\xi, \xi^{-1}]$ is an operator that commutes with L, so that the two sides of (3.6) are indeed equal. To describe all such equations we have therefore to determine the centralizer $Z(L)$ of L in the algebra $M_l(B)[\xi, \xi^{-1}]$. That is done in [15] for the general L, and the argument is unaffected by specialization. It goes as follows: we find an 'integral operator' K of the form $K = 1 + \chi_1 \xi^{-1} + \ldots$ such that $K^{-1}LK = u_n \xi^n$. It is easy to see that the centralizer of $u_n \xi^n$ consists of the operators P_0 each of whose coefficients is a constant diagonal matrix. Thus $Z(L)$ consists of the operators $P = KP_0 K^{-1}$. The operators P with P_0 a monomial, $P_0 = p\xi^r$ (p a constant diagonal matrix), are called *homogeneous*. The reason is that in the general case ($u_{i,x\beta}$ differentially independent) one can

introduce a grading on $M_l(B)[\xi, \xi^{-1}]$ in which ξ has degree 1 and $u_{i,\alpha\beta}^{(j)}$ has degree $n - i + j$, so that L is homogeneous of degree n; then the operators P just described are indeed homogeneous with respect to this grading. (Strictly speaking, we are using the word 'grading' a bit loosely, because there are infinite sums involved.) In the case of a specialized L there might in general be no such grading, that is, if $\varphi: B(u) \to B(w)$ is the homomorphism defining the specialization, there might not be any grading on $B(w)$ making φ a graded homomorphism; however, for the specializations of interest in practice there nearly always is such a grading. For the operator (3.5), for example, we give v degree 1. Thus the reader may safely think of 'homogeneous' as referring to a natural grading determined so as to make L homogeneous of degree n.

(3.7) *Remark.* The idea of finding an operator K that conjugates L into its leading term plays, in some guise or other, a basic role in most work on this subject. We can now understand the conditions (3.4) better: given (3.4)(i), the condition (3.4)(ii) is just what is needed to ensure that we can find the desired K. We take this opportunity to correct the incautious remark 2 in Sect. 6 of [15]: the condition that all the c_α be non-zero is quite essential for the construction sketched above. That is clear from a glance at the equations determining the coefficients χ_i of K (Eqs. (5.4) in [15]).

(3.8) *Remark.* In the sketch above, we have suppressed the main difficulty, namely that the entries in the coefficients of K do not lie in our original algebra B, but in a larger one. It is thus not obvious that the entries in the operators $P = K P_0 K^{-1}$ lie in B: but it is one of the main results of [15] that this is in fact the case.

To sum up, then: for each constant diagonal matrix p and each integer r, $Z(L)$ contains a unique homogeneous operator P with leading term $p\xi^r$; and $Z(L)$ consists of the linear combinations of these. For each $P \in Z(L)$, we can try to form the Lax equation (3.6). It is at this point that specialization makes a difference. Since P and L commute, we have

$$[P_+, L] = [L, P_-]$$

which shows that this expression is a differential operator of order at most $n - 1$, and also that the diagonal entries in the coefficient of ξ^{n-1} vanish. This shows that in the general case each $P \in Z(L)$ gives us a sensible Lax equation (3.6). But if L is a specialization, these equations will in general be inconsistent. In the case of the operator (3.5), for example, the general theory assures us that $[P_+, L]$ will always be an operator of order zero of the form $\begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$ but it does not guarantee that $b = -a$, which we need if Eq. (3.6) is to make sense. In general, whatever relations there may be among the coefficients of L, we shall not be happy with Eq. (3.6) unless the same relations are reproduced among the coefficients of $[P_+, L]$. More formally, we make the following definition.

(3.9) *Definition.* Given a specialized L, we say an operator $P \in Z(L)$ (or the corresponding equation (3.6)) *survives the specialization* if there is an evolutionary (that is, commuting with ∂) derivation ∂_t of B so that (3.6) holds.

We then have the following basic problem.

(3.10)   *Problem.* (First problem of specialization.) Determine which operators P survive a given specialization.

Next we consider the conservation laws. In the general case (see [15], (4.1)) one shows that for each $Q \in Z(L)$, the trace $\operatorname{tr}\operatorname{res}Q$ is a conserved density for all the Eqs. (3.6), that is, $\partial_t \operatorname{tr}\operatorname{res}Q \in \operatorname{Im}\hat{c}$. It follows automatically that for a specialized L, the $\operatorname{tr}\operatorname{res}Q$ will still be conserved densities for any Eqs. (3.6) that survive the specialization; the problem now is that they may be trivial; that is, we may have $\operatorname{tr}\operatorname{res}Q \in \operatorname{Im}\hat{c}$. This is perhaps best viewed in terms of the map $B(u) \to B(w)$ defining the specialization: there is no reason why the induced map $B(u)/\operatorname{Im}\hat{c} \to B(w)/\operatorname{Im}\hat{c}$ should be injective. So we have another problem.

(3.11)   *Problem.* (Second problem of specialization.) Determine which conserved densities $\operatorname{tr}\operatorname{res}Q$ survive a given specialization (that is, do not become trivial).

In studying this problem it is of course helpful to know that (for Q of positive order) the conserved densities $\operatorname{tr}\operatorname{res}Q$ are not trivial to begin with, for the general L (except when Q is an integral power of L). This is not proved anywhere in the literature (it was conjectured in [15]), but it is a simple consequence of the fact that the general Lax equations (3.6) can be written in Hamiltonian form. We shall give the argument at the end of Sect. 7.

Naturally, one can formulate many more 'problems' associated with specialization, but the two above will do for the moment. We do not know any general approach to these problems, but we can solve them for most of the specializations that interest us in practice. Sections 4 and 8 will illustrate this.

4. The Modified Equations

We shall work over the algebra $B = B(v_1, \ldots, v_{n-1})$ of differential polynomials in $n-1$ variables v_i. Combining the trains of thought in Sects. 2 and 3, we consider the (specialized) first order differential operator

(4.1) $L = \omega\text{-circ}(\xi, v_1, \ldots, v_{n-1}).$

Thus if $n = 2$, L is the operator (3.5). If $n = 3$, then we have

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix} \xi + \begin{pmatrix} 0 & v_1 & v_2 \\ \omega v_2 & 0 & \omega v_1 \\ \omega^2 v_1 & \omega^2 v_2 & 0 \end{pmatrix}$$

As in Sect. 3, let $Z(L)$ be the centralizer of L in the algebra $M_n(B)[\xi, \xi^{-1}]$.

(4.2)   **Proposition.** *If L is given by* (4.1), *then $Z(L)$ contains a unique homogeneous element P of each order r with the properties*

(i) $P = \operatorname{Id}.\xi^r + $ *(lower order terms)*

(ii) *P is a circulant*

(iii) *the Lax equation formed from P is consistent; that is, there exists an evolutionary derivation ∂_t of $B(v)$ such that* (3.6) *holds.*

Proof. We recall from Sect. 3 that 'homogeneous' refers to the natural $\mathbb{Z}$-grading such that ξ has degree 1 and $v_i^{(j)}$ has degree $j + 1$. We know that there is a unique

homogeneous $P \in Z(L)$ satisfying (i), so we have only to see that this P satisfies (ii) and (iii). Now, there is also the $\mathrm{mod}\, n$ grading given by the decomposition into ω^k-circulants (see (2.2)); in fact it is easy to see that we have a bigrading, that is, each operator in $M_n(B)[\xi, \xi^{-1}]$ has a unique decomposition as a sum of components that are homogeneous in both senses. Since L is bihomogeneous, when we decompose an element of $Z(L)$ in this way each component will still lie in $Z(L)$. Now, an operator is clearly an ω^k-circulant if and only if each of its coefficients is; and a $\mathbb{Z}$-homogeneous element of $Z(L)$ is uniquely determined by its leading term. Hence a $\mathbb{Z}$-homogeneous element of $Z(L)$ is an ω^k-circulant if and only if its leading term is. Since the identity matrix is a circulant, it follows that the homogeneous element $P \in Z(L)$ satisfying (i) also satisfies (ii). Finally, to prove (iii), we note that the consistency condition for the Lax equation (3.6) formed from P is just that the right hand side $[P_+, L]$ should be an ω-circulant (as is the left hand side $\partial_t L$); that is clearly the case if P (hence also P_+) is a circulant. That proves (iii). Indeed, $[P_+, L]$ is an ω-circulant if and only if any non-circulant part of P contributes zero to the Lax equation, so we have solved the 'first problem of specialization' (3.10) for this L: only circulant operators $P \in Z(L)$ give non-trivial consistent Lax equations.

Remarks. (i) Of course, if r is a multiple of n, the operator P described in (4.2) is just a power of L, so the corresponding Lax equation is trivial.

(ii) Instead of using the theory of [15], we could also construct the necessary operators using the technique of 'fractional powers': the operators P of (4.2) are just the admissible fractional powers of L^n in the sense of [11].

The Eqs. (3.6) formed from the circulant operators of (4.2) constitute our hierarchy of 'modified Lax equations'. To see why we call them that, define a differential operator $\tilde{L}$ with scalar coefficients by $\tilde{L} = s(L^n)$, where s is the summation map of Sect. 2. Of course L^n is a circulant, by (2.2)(ii); its leading coefficient is the identity, hence its second coefficient vanishes, so that $\tilde{L}$ is an operator of the form

$$\tilde{L} = \xi^n + u_{n-2}\xi^{n-2} + \ldots + u_1\xi + u_0$$

with $u_i \in B(v)$. (To avoid confusion with the notation of Sect. 3, note that these u_i are scalars, not matrices.)

(4.3) **Proposition.** (i) *Let P be the circulant of order r described in (4.2); set $\tilde{P} = s(P)$. Then $\tilde{P}$ is the operator of order r in the usual hierarchy formed from $\tilde{L}$ (in particular, the coefficients of $\tilde{P}$ are differential polynomials in the u_i).*

(ii) *If the variables v_i satisfy the modified Lax equation (3.6), then the variables u_i satisfy the (scalar) Lax equation*

$$\partial_t \tilde{L} = [\tilde{P}_+, \tilde{L}].$$

Proof. Since s is a homomorphism (see (2.5)), the equation $[P, L^n] = 0$ implies that $[\tilde{P}, \tilde{L}] = 0$; since $\tilde{P}$ is homogeneous of degree r with leading term ξ^r, (i) follows. Part (ii) is also a trivial consequence of the fact that s is a homomorphism.

Finally, we apply the homomorphism s to the factorization of L^n into n circulants (see (2.4)) to obtain the relationship between the 'original' and 'modified' variables in a more convenient form.

(4.4) **Proposition.** *The operator* $\tilde{L}$ *splits into linear factors*

$$\tilde{L} = (\xi + y_{n-1}) \dots (\xi + y_1)(\xi + y_0)$$

where

$$y_k = \omega^k v_1 + \omega^{2k} v_2 + \dots + \omega^{(n-1)k} v_{n-1}.$$

Thus the variables u_i are some kind of 'non-commutative elementary symmetric functions' in the variables v_i.

Examples. In the case $n = 2$, the variables u and v are related by the Miura transformation (1.6). The circulant P_+ of degree 3 described in (4.2) is

$$P_+ = \text{circ}(\xi^3 - \tfrac{3}{2}v^2\xi - \tfrac{3}{2}vv', \tfrac{3}{2}v'\xi + \tfrac{3}{4}v'').$$

The Lax equation formed from $4P$ is the modified KdV equation (1.5). Also

$$\tilde{P}_+ = s(P_+) = \xi^3 + \tfrac{3}{2}(v' - v^2)\xi + \tfrac{3}{4}v'' - \tfrac{3}{2}vv' = \xi^3 + \tfrac{3}{2}u\xi + \tfrac{3}{4}u';$$

the scalar Lax equation formed from $4\tilde{P}$ is the KdV equation (1.3).

In the case $n = 3$, explicit formulas are already quite unenlightening. For example, the relationship between the two sets of variables is

$$u_1 = (1 - \omega^2)v_1' + (1 - \omega)v_2' - 3v_1v_2$$
$$u_0 = v_1'' + v_2'' + (v_1 + v_2)[(\omega - 1)v_1' + (\omega^2 - 1)v_2'] + v_1^3 + v_2^3.$$

5. Hamiltonian Structures

Let $B = B(w_1, \dots, w_N)$ be a differential algebra of our usual kind (see Sect. 3). We first recall the definition of adjoint operators. The adjoint Q^* of a (scalar) differential operator $Q \in B[\xi]$ is characterized by the two properties (i) $*$ is an anti-automorphism of $B[\xi]$ ('anti-' means that $(QR)^* = R^* Q^*$) (ii) if $f \in B$, then $(f\xi)^* = -\xi f$. If l is a matrix of differential operators, then l^* is defined by $(l^*)_{ij} = l_{ji}^*$. The main property of l^* is the following: if F, G are column vectors of elements of B, then

(5.1) $(l^*F)^t G \equiv F^t(lG)$ $\mod \text{Im } \partial$.

(As usual, the superscript t denotes the matrix transpose.) Intuitively: the 'integrals' of the two expressions in (5.1) are equal.

Now let l be an $N \times N$ skew matrix of differential operators ($l^* = -l$). We are going to use l to assign to each $f \in B$ an evolutionary derivation ∂_f of B. We write w for the column vector $(w_1, \dots, w_N)^t$, and similarly $\delta f/\delta w$ for $(\delta f/\delta w_1, \dots, \delta f/\delta w_N)^t$. Here $\delta/\delta w_i$ is the partial variational derivative

$$\frac{\delta f}{\delta w_i} = \sum_j (-\partial)^j \frac{\partial f}{\partial w_i^{(j)}}.$$

Then given $f \in B$, we define ∂_f to be the evolutionary derivation of B whose values on $w_1, \dots, w_N$ are given by

$$\partial_f w = l \frac{\delta f}{\delta w}.$$

The value of ∂_f on any function $g \in B$ is then given explicitly by the chain rule:

$$\partial_f g = \sum_{i,j} (\partial_f w_i)^{(j)} \frac{\partial g}{\partial w_i^{(j)}} \tag{5.2}$$

We think of ∂_f as the Hamiltonian vector field corresponding to f, though the term 'Hamiltonian' is not justified unless l satisfies the extra condition formulated below (5.5).

Given $f, g \in B$, we define their Poisson bracket (with respect to l) by

$$\{f, g\} = \left(\frac{\delta g}{\delta w} \right)^t \left(l \frac{\delta f}{\delta w} \right). \tag{5.3}$$

Since the variational derivatives vanish on the image of ∂, the derivations ∂_f and the brackets $\{f, g\}$ depend only on the class of f, g in $B/\partial B$. We regard the brackets too as taking values in $B/\partial B$; it then follows from (5.1) that the bracket is skew-symmetric. The Definition (5.3) is justified by the next lemma, which follows at once from (5.2) and 'integration by parts'.

(5.4) **Lemma.** *In $B/\partial B$, we have*

$$\{f, g\} = \partial_f g = -\partial_g f.$$

Now we can formulate our main definition.

(5.5) *Definition.* We say the skew operator l is *Hamiltonian* if for all $f, g \in B$ we have

$$\partial_{\{f, g\}} = [\partial_f, \partial_g].$$

Before giving examples, we should like to comment on the intuitive meaning of our set-up. We think of B, or possibly $B/\partial B$, as being like the algebra of smooth functions on a manifold; we then think of ∂_f as a vector field and the map $f \mapsto \delta f/\delta w$ as the exterior derivative or 'gradient'. It is then natural to think, more generally, of any N-tuple $x = (x_1, \ldots, x_N)^t$ of elements of B as a 1-form, and assign to each 1-form x the vector field ∂_x defined by

$$\partial_x w = lx.$$

Thus the basic structure we are studying is a map from 1-forms to vector fields (satisfying certain conditions). Our skew matrix l should therefore be thought of as analogous to a skew form on the cotangent bundle (not the tangent bundle) of a manifold.[1] This is something rather more general than what is usually encountered in Hamiltonian mechanics; there one is given a 2-form, that is, a skew form on the tangent bundle, and it is necessary to assume that it is non-

---

[1]   That is, l is like a skew tensor *of the opposite type* from a 2-form. It is for this reason that we avoid the term 'symplectic'

degenerate in order to invert the corresponding map (vector fields) $\rightarrow$ (1-forms). The best finite-dimensional example of our situation is provided by the dual of a Lie algebra (see [7], Chap. 2, Sect. 15). There one has a natural skew form on the cotangent bundle, and it is not non-degenerate. As is well known, this form induces a symplectic structure in the usual sense on each co-adjoint orbit, but for our present purposes that is an unnecessary refinement. We refer to [6] for a more detailed discussion along these lines.

Now we give some examples of Hamiltonian structures.

Example 1. Let l be any skew matrix of differential operators with constant coefficients. Then l is Hamiltonian (see, for example, [11], Chap. 1, (7.13)(a)).

The rest of the examples are motivated by the theory of Lax equations (see (7.13) below).

Example 2. (First Hamiltonian structure for Lax equations, case $u_{n-1} \neq 0$). We take $B = B(u_0, \ldots, u_{n-1})$. Let L be the differential operator

$$L = \xi^n + u_{n-1} \xi^{n-1} + \ldots + u_0.$$

We assign to each '1-form' $x = (x_0, \ldots, x_{n-1})$ the (formal) integral operator

$$X = \xi^{-1} x_0 + \xi^{-2} x_1 + \ldots + \xi^{-n} x_{n-1}.$$

Then the vector field ∂_x corresponding to x is defined by

$$(5.6) \qquad\qquad \partial_x L = [L, X]_+$$

where as usual ∂_x is understood to act coefficient-wise on L. It is clear that the map $x \mapsto \partial_x$ is given by a certain matrix of differential operators l, which we shall not write out explicitly. This matrix is Hamiltonian.

This Hamiltonian structure has the following, rather silly, properties: (i) ∂_x does not depend on the last 'co-ordinate' x_{n-1} (ii) we always have $\partial_x u_{n-1} = 0$; intuitively, all the Hamiltonian vector fields are tangent to the 'sub-manifold' $u_{n-1} = 0$. It is thus more sensible (as is usually done) to restrict this structure to the submanifold $u_{n-1} = 0$.

Example 3. (First Hamiltonian structure for Lax equations, case $u_{n-1} = 0$). We take $B = B(u_0, \ldots, u_{n-2})$. Let L be the differential operator

$$L = \xi^n + u_{n-2} \xi^{n-2} + \ldots + u_0.$$

We assign to each 1-form $x = (x_0, \ldots, x_{n-2})$ the integral operator

$$X = \xi^{-1} x_0 + \ldots + \xi^{-(n-1)} x_{n-2}.$$

The vector field ∂_x is defined by the same formula as before (5.6). The matrix l implementing this map $x \mapsto \partial_x$ is Hamiltonian.

The fact that the operators of examples 2 and 3 are Hamiltonian has been proved many times, originally by horrifying computations (see [11]): better proofs can be found in [1, 6, 9]. The present paper incidentally provides yet another proof, because the fact that the second structure is Hamiltonian implies trivially that the first is too (see the discussion of Example 6 below).

Example 4. (Second Hamiltonian structure for Lax equations, case $u_{n-1} \neq 0$.) We take B, L and X as in Example 2, but now ∂_x is defined by

(5.7) $\partial_x L = (LX)_+ L - L(XL)_+ = -(LX)_- L + L(XL)_-$.

(The first expression shows that $\partial_x L$ is a differential operator, the second that it has order at most $n-1$, so the formula makes sense.) Again it is clear that the map $x \mapsto \partial_x$ is given by some matrix l of differential operators, more complicated than before; the interested reader will find it written out explicitly in [5].

   Notice that now we do not always have $\partial_x u_{n-1} = 0$, so this structure cannot be restricted to $u_{n-1} = 0$ as easily as before.

Example 5. (Second Hamiltonian structure for Lax equations, case $u_{n-1} = 0$.) We take B and L as in Example 3, but we can not quite take X as in Example 3, because we want to define ∂_x by (5.7) as before, and this expression may still have order $n-1$, not $n-2$ as we need. We proceed as follows. It is easy to check (from the second expression) that the coefficient of ξ^{n-1} in (5.7) is res $[X, L]$. So given a 1-form $x = (x_0, \ldots, x_{n-2})$, we set

(5.8) $X = \xi^{-1} x_0 + \ldots + \xi^{-(n-1)} x_{n-2} + \xi^{-n} x_{n-1}$

where x_{n-1} is determined by the condition res $[X, L] = 0$, or, explicitly,

(5.9) $\partial x_{n-1} = (1/n) \, \text{res} \, [\xi^{-1} x_0 + \ldots + \xi^{-(n-1)} x_{n-2}, L]$.

Since the right hand side here is in Im ∂ (see [11], Chap. 2, (3.3)), this equation does indeed determine an element $x_{n-1} \in B$ (uniquely if we agree that it is to have zero constant term). Also, it is clear from (5.9) that x_{n-1} is a linear combination (with coefficients involving the u_i) of derivatives of $x_0, \ldots, x_{n-2}$. It follows that if we define X by (5.8) and (5.9), and ∂_x by (5.7), then the map $x \mapsto \partial_x$ is still given by a matrix l of differential operators, this time so complicated that even Gel'fand and Dikii [5] do not care to write it out explicitly.

   The fact that the operators l in Examples 4 and 5 are Hamiltonian was conjectured by Adler [1] and first proved by Gel'fand and Dikii [5], essentially by direct calculation. We shall give a different proof in Sects. 8 and 11.

Example 6. (Compatibility of first and second Hamiltonian structures for Lax equations.) Let l_1, l_2 be the skew operators of Examples 2 and 4 (or 3 and 5). Then for any scalars α, β, the operator $\alpha l_1 + \beta l_2$ is Hamiltonian.

   This is a trivial consequence of the fact that l_2 is Hamiltonian. For if in the expression $(LX)_+ L - L(XL)_+$ (see (5.7)) we replace L by $L + \lambda$ (λ a scalar) the effect is just to add on a term $\lambda [L, X]_+$. That implies that if in the operator l_2 we replace u_0 by $u_0 + \lambda$, we get $l_2 + \lambda l_1$: this operator is therefore Hamiltonian (given that l_2 is). (That could be regarded as a very special case of (6.1) below, the Fréchet Jacobian here being the identity.) Finally, given that the operators $l_2 + \lambda l_1$ are Hamiltonian for all λ, it follows at once that l_1 is too.

Note on signs. Our definition of the second Hamiltonian structure has the opposite sign to that of Gel'fand and Dikii [5]. Changing the sign would improve our main results (8.6) and (11.5), but then a minus sign would appear in (7.13)(ii), which seems unacceptable.

Perhaps an explicit formula would be helpful at this point. Let us calculate the matrix l of Example 5 in the simplest case $n = 2$. So we have $L = \xi^2 + u$, $X = \xi^{-1} x_0 + \xi^{-2} x_1$; a short calculation yields

$$(LX)_+ L - L(XL)_+ = (\partial^2 x_0 - 2\partial x_1)\xi + \partial^3 x_0 - \partial^2 x_1 + 2u\partial x_0 + \partial u \cdot x_0.$$

To make this have order zero we must take $x_1 = \frac{1}{2}\partial x_0$, and then

$$\partial_x L = \frac{1}{2}\partial^3 x_0 + 2u\partial x_0 + \partial u \cdot x_0 = l x_0,$$

where l is the operator $\frac{1}{2}\partial^3 + u\partial + \partial u$. This is therefore the operator defining the second Hamiltonian structure for $n = 2$.

Before leaving the examples we comment quickly on the generalization to the case where L has matrix coefficients. Examples 2 and 3 present no problem, and in Example 4 the definition presents no problem. However, more work would be needed to prove that the skew operator arising in Example 4 is Hamiltonian. We conjecture that it is, and we see no obstacle to proving it either by our method or by the method of Gel'fand and Dikii [5]; but we have not checked the details. (To use our method, we should start off from an operator of the kind (4.1), but now each entry would be a matrix block. This method would not be any use in the case where L has order 1, but in that case it is easy to check directly that the relevant skew operator l is Hamiltonian, for example using the criterion of Gel'fand and Dorfman [6].)

If however we try to generalize Example 5 to the matrix case, we meet a different problem. Let L be a matrix operator of the form (3.3), satisfying conditions (3.4). Let us try to imitate the procedure of Example 5. To a 1-form x we shall now associate an operator

$$X = \xi^{-1} x_0 + \ldots + \xi^{-n} x_{n-1},$$

where the diagonal entries $x_{n-1,\alpha\alpha}$ have to be determined so as to make Eq. (5.7) consistent. As in the scalar case, this condition gives equations $\partial x_{n-1,\alpha\alpha} = \ldots$, but now the three dots are not necessarily in Im ∂. The reason is that in the matrix case it is only the trace of the residue of a commutator that is in Im ∂, not the individual entries. It follows that in the matrix case with $u_{n-1,\alpha\alpha} = 0$, the second Hamiltonian structure does not exist, at least in the (admittedly rather narrow) sense we have been discussing (one would have to let the matrix l involve 'integrations', a possibility we prefer not to contemplate here).

6. Restriction of Hamiltonian Structure

In this section we should like to discuss the functorial properties of Hamiltonian structures, that is, how they behave with respect to homomorphisms of differential algebras

$$\varphi\colon B(u_1, \ldots, u_r) \to B(v_1, \ldots, v_s).$$

Unfortunately, since derivations, like vector fields on manifolds, are not functorial, Hamiltonian structures will not be either. However, the case when φ is

injective is easy to analyse. Suppose that is the case (so that $r \leq s$: in the examples that will interest us $r = s$); we shall use φ to identify $B(u)$ with a subalgebra of $B(v)$. Let l be a skew matrix defining a Hamiltonian structure on $B(v)$. Then under certain circumstances l will induce a Hamiltonian structure on $B(u)$.

(6.1) **Proposition.** *Let D denote the Fréchet Jacobian of u with respect to v (defined below). Set $\hat{l} = DlD^*$. In general this will be a matrix of differential operators with coefficients in $B(v)$; suppose however that l and φ are such that all the coefficients lie in the subalgebra $B(u)$. Then*

(i) *$\hat{l}$ defines a Hamiltonian structure on $B(u)$*

(ii) *for $f \in B(u)$, the Hamiltonian vector field ∂_f on $B(u)$ determined by $\hat{l}$ is just the restriction to $B(u)$ of the Hamiltonian vector field $\partial_{\varphi(f)}$ on $B(v)$ determined by l. (In particular, all the derivations $\partial_{\varphi(f)}$ preserve $B(u)$.)*

We naturally say that the Hamiltonian structure defined by the operator $\hat{l}$ in (6.1) is obtained by restriction of the one defined by l. The simplest non-trivial example of this situation is the one discussed in the introduction: let $r = s = 1$, and let $\varphi: B(u) \to B(v)$ be defined by $\varphi(u) = \partial v - v^2$ (the Miura transformation). Let $l = -\frac{1}{2}\partial$. Here $D = \partial - 2v$, so an easy calculation gives

$$DlD^* = (\partial - 2v)(-\tfrac{1}{2}\partial)(-\partial - 2v) = \tfrac{1}{2}\partial^3 + u\partial + \partial u.$$

Proposition (6.1) therefore shows that this operator is Hamiltonian: as we have seen, it is the operator of Sect. 5, Example 5 in the case $n = 2$.

Proof of (6.1). If $f \in B(u)$ we shall write $\bar{f}$ instead of $\varphi(f)$ (thus $\bar{f}$ is just f regarded as a function of the variables v_i). First recall that the Fréchet Jacobian is the $r \times s$ matrix of differential operators with entries

$$D_{ij} = \sum_{k \geq 0} \frac{\partial u_i}{\partial v_j^{(k)}} \partial^k.$$

Its main properties are the following.

(a) *Let ∂_t be any evolutionary derivation of $B(v)$; then*

$$\partial_t u = D \partial_t v.$$

(b) *Let $f \in B(u)$, and let $\bar{f}$ as above be the corresponding element of $B(v)$. Then*

$$\frac{\delta \bar{f}}{\delta v} = D^* \overline{\frac{\delta f}{\delta u}}.$$

Property (a) follows at once from the definition of D, and justifies the term 'Jacobian'. Property (b) can be deduced from (a), or, better, from the fact that the 'total variation operator' δ (see Sect. 7) is functorial (commutes with φ). Property (b) holds for any homomorphism φ, not necessarily injective.)

Now let $f \in B(u)$. Then using (a) and (b) above, we have

$$\partial_{\bar{f}} u = D \partial_{\bar{f}} v = Dl \frac{\delta \bar{f}}{\delta v} = DlD^* \overline{\frac{\delta f}{\delta u}} = \hat{l} \frac{\delta f}{\delta u},$$

which proves (ii). The operator $\hat{f}$ is obviously skew; to see that it is Hamiltonian, note first that for any $f, g \in B(u)$ we have

$$\overline{\{f,g\}} = \overline{\partial_f\, g} = \partial_{\bar{f}}\, \bar{g} = \{\bar{f}, \bar{g}\},$$

so that

$$\partial_{\{f,g\}} = \partial_{\{\bar{f},\bar{g}\}} | B(u) = [\partial_{\bar{f}}, \partial_{\bar{g}}] | B(u) = [\partial_f, \partial_g].$$

That proves (i).

7. Hamiltonian Structure of Lax Equations

We start off from the usual algebra of differential polynomials $B = B(w_1, \dots, w_N)$. We let $\Omega = \Omega^1(B)$ denote the free B-module on the symbols $\delta w_i^{(j)}$, $j \geqq 0$, with 'universal derivation' (total variation, exterior derivative on the jet space) $\delta : B \to \Omega^1(B)$ defined by

$$\delta f = \sum_{i,j} \frac{\partial f}{\partial w_i^{(j)}} \delta w_i^{(j)}.$$

We extend ∂ to a derivation of Ω, commuting with δ; then the variational derivatives $\delta f / \delta w_i$ are characterized by

(7.1)
$$\delta f \equiv \sum_i \frac{\delta f}{\delta w_i} \delta w_i \quad \text{mod Im } \partial.$$

We denote by M the algebra $M_l(B)[\xi, \xi^{-1}]$ of $l \times l$ matrices of formal pseudo-differential operators with coefficients in B, and by $M(\Omega)$ the M-bimodule of $l \times l$ matrices of operators with coefficients in Ω ('bimodule' means that we have (obvious) right and left actions of M on $M(\Omega)$, and these actions commute). We extend ∂, δ to act coefficient-wise on matrices of operators; then we have a commutative diagram

$$
\begin{array}{ccc}
M & \xrightarrow{\;\delta\;} & M(\Omega) \\
\partial \downarrow & & \downarrow \partial \\
M & \xrightarrow{\;\delta\;} & M(\Omega)
\end{array}
$$

in which all the maps are derivations (in appropriate senses).

If $C \in M$, $D \in M(\Omega)$, we can form the 'commutator'

$$[C, D] = CD - DC \in M(\Omega).$$

(7.2) **Lemma.** *We have*

$$\text{tr res } [C, D] \in \partial \Omega.$$

(Here as usual res singles out the coefficient of ξ^{-1}.)

Now let $L \in M$ be a differential operator of the form (3.3), satisfying the conditions (3.4). The description of the centralizer $Z(L)$ of L in M given in Sect. 3

can be rephrased as follows. Let $X_\alpha^{[r]}$ denote the (unique) homogeneous element of $Z(L)$ with leading term $E_{\alpha\alpha}\zeta^r$; that is, if K is the operator of Sect. 3 that conjugates L into its leading term, we have

$$X_\alpha^{[r]} = KE_{\alpha\alpha}\zeta^r K^{-1}.$$

(Here and below, $E_{\alpha\beta}$ denotes the matrix with 1 in the place (α, β) and zeros elsewhere.) Then $Z(L)$ consists of the linear combinations of these elements $X_\alpha^{[r]}$; and an element of $Z(L)$ is homogeneous of degree r precisely when it is of the form $\sum_\alpha p_\alpha X_\alpha^{[r]}$, p_α scalar constants. We write X_α instead of $X_\alpha^{[1]}$. If r is positive, then $X_\alpha^{[r]}$ is just X_α^r; for $r \leq 0$ that is not strictly true, since X_α is not invertible. However, we have the following.

(7.3) **Lemma.** *For any integers r, s, we have*
  (i) $X_\alpha^{[r]} X_\alpha^{[s]} = X_\alpha^{[r+s]}$
  (ii) *if $\alpha \neq \beta$, then $X_\alpha^{[r]} X_\beta^{[s]} = 0$.*

Both parts of the lemma follow at once from the corresponding assertions for the operators $E_{\alpha\alpha}\zeta^r$, conjugate to $X_\alpha^{[r]}$.

In the next few lemmas we shall use the following notation: if R and S are elements of $M(\Omega)$, we write $R \equiv S$ to mean: $R - S$ is a sum of commutators of the form $[X_\alpha^{[r]}, D]$, $D \in M(\Omega)$. Lemma (7.2) shows that this use of the symbol $\equiv$ is compatible with our previous one, in the sense that if $R \equiv S$ then tr res $R \equiv$ tr res S (mod $\partial\Omega$).

(7.4) **Lemma.** *If $R \equiv S$, then for any α, r, we have $X_\alpha^{[r]} R \equiv X_\alpha^{[r]} S$ and $RX_\alpha^{[r]} \equiv SX_\alpha^{[r]}$.*

Proof. This follows at once from the fact that all the $X_\alpha^{[r]}$ commute with each other, and the relation $a[b,c] = [b,ac]$, valid if a and b commute.

(7.5) **Lemma.** *For any integer r, we have*

$$\delta X_\alpha^{[r]} \equiv r X_\alpha^{[r-1]} \delta X_\alpha.$$

Proof. Suppose first that $r \geq 2$. Since δ is a derivation, δX_α^r is a sum of r terms of the form $X_\alpha^i \delta X_\alpha \cdot X_\alpha^{r-i-1}$. But each of these clearly differs from $X_\alpha^{r-1} \delta X_\alpha$ by a commutator of the desired form.

Now let $r = 1$, so we have to prove that $\delta X_\alpha \equiv X_\alpha^{[0]} \delta X_\alpha$. Applying δ to the relation $X_\alpha = X_\alpha^{[0]} X_\alpha$, we get

$$\delta X_\alpha = X_\alpha^{[0]} \delta X_\alpha + \delta X_\alpha^{[0]} \cdot X_\alpha.$$

In view of (7.4), it is thus enough to show that $\delta X_\alpha^{[0]} \equiv 0$, which is the case $r = 0$ of our lemma. To prove that, we apply δ to the relation $X_\alpha^{[0]} = (X_\alpha^{[0]})^n$ to get

$$\delta X_\alpha^{[0]} \equiv n X_\alpha^{[0]} \delta X_\alpha^{[0]} \quad \text{for all } n \geq 2.$$

Hence $\delta X_\alpha^{[0]} \equiv 0$.

It remains to prove the lemma for $r < 0$: we omit this argument, since we shall use (7.5) only for $r \geq 1$.

(7.6) **Lemma.** *For any integers* r, s, *we have*

  (i) $r\, X_\alpha^{[r]} \delta X_\alpha^{[s]} \equiv s\, X_\alpha^{[s]} \delta X_\alpha^{[r]}$

  (ii) *if* $\alpha \neq \beta$, *then* $X_\alpha^{[r]} \delta X_\beta^{[s]} \equiv 0$.

Proof. From (7.5), it follows that the two sides in (i) are both congruent to the expression

$$r s\, X_\alpha^{[r+s-1]} \delta X_\alpha.$$

Part (ii) also follows at once from (7.5) and (7.3)(ii).

The next lemma is the crucial one that we have been aiming at. Our proof, like everything in this section, is modelled on the treatment of the scalar case $(l=1)$ given in [11].

(7.7) **Lemma.** *Let* P, $Q \in Z(L)$ *be homogeneous of degrees* r, s, *respectively. Then*

$$r P\, \delta Q \equiv s Q\, \delta P.$$

Proof. This follows from (7.6), since P and Q are just linear combinations of the elements $X_\alpha^{[r]}$ and $X_\alpha^{[s]}$, respectively.

(7.8) **Corollary.** *Let* $P \in Z(L)$ *be homogeneous of degree* r. *Then*

$$\delta(LP) \equiv (n+r)/n\; \delta L \cdot P.$$

Proof. Using (7.7) with $Q = L$, we find

$$\delta(LP) = \delta L \cdot P + L \cdot \delta P \equiv \delta L \cdot P + (r/n)\, \delta L \cdot P.$$

If $P \in Z(L)$ is homogeneous of degree r, we set

(7.9)
$$H_P = \begin{cases} (n/r)\, \mathrm{tr\ res}\, P, & r > 0 \\ 0, & r \leq 0. \end{cases}$$

If $P \in Z(L)$ is not homogeneous, we define H_P by adding (7.9) over its homogeneous components. Then (7.8) and (7.2) imply the following.

(7.10) **Corollary.** *Let* $P \in Z(L)$. *Then*

$$\delta H_{LP} \equiv \mathrm{tr\ res}(\delta L \cdot P) \quad \mathrm{mod\ Im}\; \partial.$$

This corollary can be used to calculate the variational derivatives $\delta H_{LP}/\delta w_i$ in terms of the coefficients of P_-. Let us do this in the case of the 'general' L, that is, the $u_{i,\alpha\beta}$ are differentially independent, so that we can work over the algebra $B(u_{i,\alpha\beta})$. For $H \in B(u_{i,\alpha\beta})$, let $\delta H/\delta u_i$ denote the matrix whose (α, β) entry is $\delta H/\delta u_{i,\beta\alpha}$ (note the transposition of α and β).

(7.11) **Corollary.** *Let* $P \in Z(L)$, L *general. Let* $H = H_{LP}$ *(defined as above). Then*

$$P_- = \xi^{-1} \frac{\delta H}{\delta u_0} + \xi^{-2} \frac{\delta H}{\delta u_1} + \ldots + \xi^{-n} \frac{\delta H}{\delta u_{n-1}} + \ldots.$$

(In the case of the coefficient of ξ^{-n}, since we are assuming $u_{n-1,xx}=0$, this has to be interpreted as meaning that the off-diagonal entries are those indicated.)

Proof. We have

$$\delta L = \sum_{i,x,\beta} \delta u_{i,x\beta} E_{x\beta} \xi^i,$$

so (7.11) follows at once from (7.1), (7.10) and the obvious fact that

(7.12) $\mathrm{tr\,res}(aE_{x\beta}\xi^i \cdot \xi^{-j-1} E_{\gamma\delta}b) = \begin{cases} ab & \text{if } i=j, \ \alpha=\delta, \ \beta=\gamma \\ 0 & \text{otherwise.} \end{cases}$

(7.13)   **Corollary.** (i) *For general L, the Lax equations*

$$\partial_t L = [L, P_-]$$

can be written in first Hamiltonian form (as in Sect. 5, Example 3) with Hamiltonian H_{LP}. (ii) *In the case of scalar coefficients ($l=1$), they can also be written in second Hamiltonian form (Sect. 5, Example 5) with Hamiltonian H_P.*

Perhaps (ii) deserves some explanation. In view of (7.11) it is clear that to have any chance of finding a Hamiltonian form for $\partial_t L = [L, P_-]$ with Hamiltonian H_P we must rewrite the equation in terms of $Q = L^{-1}P$. That is easy:

$$\partial_t L = [L, (QL)_-] = [L, (Q_-L)_-].$$

Since Q and L commute, we can also write this as

$$\partial_t L = L(Q_-L)_- - (LQ_-)_-L = (LQ_-)_+L - L(Q_-L)_+;$$

we have arrived at the strange-looking expression (5.7) (with Q_- for X). The last expression shows that $\partial_t L$ in fact depends only on the first n coefficients of Q_-. Since the first $n-1$ (right hand) coefficients are the $\delta H_P/\delta u_i$, and the nth coefficient obviously satisfies the consistency condition (5.9), part (ii) of (7.13) follows at once from the definition of the second Hamiltonian structure.

For want of a better place, we indicate here how (7.11) implies that if P has positive order the conserved density H_P for the general Lax equations is non-trivial (except when P is an integral power of L). For simplicity we give the argument in the scalar case ($l=1$). It is enough to show that if $P \in Z(L)$ is the homogeneous operator of the form $P = \xi^r + (\text{lower terms})$ and r is not divisible by n, then the conserved density $\mathrm{res}\,P$ is non-trivial. Suppose it were trivial. Then (7.11) shows that the first $n-1$ coefficients of $(L^{-1}P)_-$ would vanish; in particular, we should have $\mathrm{res}\,L^{-1}P = 0$. Repeating the argument, we deduce that the first $n-1$ coefficients of the operators $(L^{-q}P)_-$ vanish for all $q \geq 1$. But one of these operators has leading term ξ^{-i}, $1 \leq i \leq n-1$, a contradiction.

In the matrix case a similar argument (a little more elaborate, especially if $n=1$) shows also the linear independence of the conserved densities, as conjectured in the last paragraph of [15].

8. The Second Hamiltonian Structure

We now return to the study of the modified Lax equations constructed in Sect. 4. We begin by establishing the analogue of (7.11) for our specialized L

$$L = \omega\text{-circ}(\xi, v_1, \ldots, v_{n-1}).$$

For each $P \in Z(L)$, we have the 'Hamiltonian' H_P defined by (7.9): thus if P is homogeneous of degree $r > 0$, then $H_P = (1/r)\,\text{tr res }P$ (the n in (7.9) was the order of L). Extending the notation used in Sect. 2, we shall write

$$X = \omega^k\text{-circ}(x_0, \ldots, x_{n-1})^t$$

to mean: X is the ω^k-circulant whose first *column* is the vector indicated: thus the (i,j) entry of X is $\omega^{kj} x_{i-j}$. (As in Sect. 2, indices run from 0 to $n-1$ and are read mod n where necessary.)

(8.1)   **Proposition.** *Let* $P \in Z(L)$. *Then*

(i) *if* P *is an* ω^k-*circulant*, $k \neq 0$, *then* $H_P \in \text{Im } \hat{c}$
(ii) *if* P *is a circulant, then* [2]

$$\text{res } L^{-1} P = (1/n)\,\omega^{-1}\text{-circ}(0, \delta H_P/\delta v_1, \ldots, \delta H_P/\delta v_{n-1})^t.$$

Note that since every P is a sum of ω^k-circulants, (8.1) solves the 'second problem of specialization' (3.11) for this L: essentially just the conserved densities coming from circulants survive the specialization.

Proof of (8.1). Let $Q = L^{-1}P$, and let

$$\text{res } Q = \omega^{k-1}\text{-circ}(q_0, \ldots, q_{n-1})^t.$$

By (7.10) (with Q instead of P) we have

$$\begin{aligned}
\delta H_P &\equiv \text{tr res}(\delta L \cdot Q) \qquad (\text{mod Im } \hat{c}) \\
&\equiv \sum_{i,j} (\delta L)_{ij}(\text{res }Q)_{ji} \qquad (\text{by }(7.12)) \\
&\equiv \sum_{i,j} \omega^i \, \delta v_{j-i} \cdot \omega^{(k-1)i} q_{j-i} \\
&\equiv \left(\sum_i \omega^{ki}\right)\left(\sum_l \delta v_l \cdot q_l\right).
\end{aligned}$$

(In this calculation all sums are taken from 0 to $n-1$, indices $j-i$ are read mod n, and we set $v_0 = 0$.) Now, if $k \neq 0$, the sum of roots of unity is zero, so (7.1) shows that $\delta H_P/\delta v_i = 0$ for all i, which proves (i). If $k = 0$, then (7.1) gives

$$\frac{\delta H_P}{\delta v_i} = n q_i, \qquad i = 1, 2, \ldots, n-1,$$

which almost proves (ii); that $q_0 = 0$ will follow from our next calculation (8.2).

---

[2]   We exclude the case where P has a component that is of degree zero, that is, a multiple of the identity

Proposition (8.1) shows that a 'first' Hamiltonian structure for our modified equations, formed from circulant operators P, does not exist, since the relevant Hamiltonians H_{LP} do not survive the specialization. However, since the Hamiltonians H_P do survive, we may well have a 'second' Hamiltonian structure. We shall now verify that, and calculate explicitly the corresponding skew matrix l.

Let $x = (x_1, \ldots, x_{n-1})^t$, $x_i \in B$, be a '1-form'. Motivated by (8.1), we assign to x the operator

$$X = (1/n)\,\xi^{-1} \cdot \omega^{-1}\text{-circ}(x_0, x_1, \ldots, x_{n-1})^t$$

where x_0 is to be determined so that Eq. (5.7) is consistent.

(8.2) **Proposition.** *The equations*

$$\partial_x L = (LX)_+ L - L(XL)_+$$

are consistent if and only if x_0 is constant; in that case they take the form

$$\partial_x v_i = -(1/n)\,\partial x_{n-i}, \qquad i = 1, \ldots, n-1.$$

The first assertion completes the proof of (8.1): for we know that the Lax equation formed from a circulant P is consistent, and has the form in (8.2) with $X = Q_-$ (see the proof of (7.13)). It follows that the element q_0 in the proof of (8.1) is constant, hence zero if P has no degree zero component.

Proof of (8.2). Direct calculation. Let us set

$$V = \text{circ}(0, v_1, \ldots, v_{n-1}), \qquad \tilde{X} = (1/n)\,\text{circ}(x_0, \ldots, x_{n-1})^t$$

so that $L = \Omega(\xi + V)$, $X = \xi^{-1}\tilde{X}\Omega^{-1}$, where Ω is as in (2.3). We then find

$$(LX)_+ = \Omega\tilde{X}\Omega^{-1}, \qquad (XL)_+ = \tilde{X},$$

so that

$$(LX)_+ L = \Omega\tilde{X}\xi + \Omega\tilde{X}V, \qquad L(XL)_+ = \Omega\tilde{X}\xi + \Omega\partial\tilde{X} + \Omega V\tilde{X}.$$

Subtracting and using the fact that $\tilde{X}V = V\tilde{X}$, we get

$$\partial_x L = -\Omega\,\partial\tilde{X},$$

from which (8.2) follows at once.

Combining (8.1) and (8.2), we obtain the following.

(8.3) **Proposition.** *Let $P \in Z(L)$ be a circulant. Then the modified Lax equations*

$$\partial_t L = [P_+, L] = [L, P_-]$$

can be written in Hamiltonian form

$$\partial_t v = l\frac{\delta H_P}{\delta v},$$

where l is the matrix of operators with $-(1/n)\,\partial$ along the 'off-diagonal' and zeros elsewhere.

Note that l is indeed Hamiltonian (Sect. 5, Example 1).

Next we show that the conserved densities (Hamiltonians) of the original and modified equations are 'the same'. In the KdV case $(n=2)$ this is a very old observation, due to Miura (see [8]).

(8.4) **Proposition.** *Let $P \in Z(L)$ be the homogeneous circulant of the form $P = \mathrm{Id} \cdot \xi^r + \dots$ (see (4.2)); let $\tilde{P} = s(P) \in Z(\tilde{L})$. As usual, set*

$$H_P = (1/r)\,\mathrm{tr}\,\mathrm{res}\,P, \qquad H_{\tilde{P}} = (n/r)\,\mathrm{res}\,\tilde{P}.$$

Then $H_P \equiv H_{\tilde{P}} \pmod{\mathrm{Im}\,\partial}$.

Proof. Let $\mathrm{res}\,P = \mathrm{circ}(p_0, \dots, p_{n-1})$. Then a trivial calculation shows that the Lax equation $\partial_t L = [L, P_-]$ is equivalent to

$$\partial_t v_i = (1 - \omega^i)\,p_i, \qquad i = 1, 2, \dots, n-1.$$

It follows from (8.3) that $p_i \in \mathrm{Im}\,\partial$ for $i \neq 0$. Thus modulo $\mathrm{Im}\,\partial$, we have

$$(n/r)\,\mathrm{res}\,\tilde{P} = (n/r)\sum_0^{n-1} p_i \equiv (n/r)\,p_0 = (1/r)\,\mathrm{tr}\,\mathrm{res}\,P.$$

Finally, before stating our main theorem we need the following lemma, the proof of which we defer until Sect. 9. Recall from Sect. 4 that the coefficients u_i of $\tilde{L}$ are certain differential polynomials in the modified variables v_i.

(8.5) **Lemma.** *The variables $(u_0, \dots, u_{n-2})$ are differentially independent.*

That means that we have an inclusion of differential algebras

$$B(u_0, u_1, \dots, u_{n-2}) \subset B(v_1, v_2, \dots, v_{n-1})$$

so we are in a position to use Proposition 6.1.

(8.6) **Theorem.** *The Hamiltonian structure on $B(v)$ defined by the operator l of (8.3) restricts to give the second Hamiltonian structure on $B(u)$ (Sect. 5, Example 5).*

Proof. Let D be the Fréchet Jacobian of u with respect to v, and let $\tilde{l}$ be the skew operator defining the second structure on $B(u)$. By (6.1), we have to prove that for all $f \in B(u)$, we have

$$(D\,l\,D^* - \tilde{l})\frac{\delta f}{\delta u} = 0.$$

Now, (4.3)(ii), (8.3) and (8.4) show that this is true whenever $f = H_{\tilde{P}}$ is one of the Hamiltonians for the Lax equations. Hence the proof is completed by the next lemma, which will be proved in Sect. 10.

(8.7) **Lemma.** *Let S be a matrix of differential operators with coefficients in $B(v)$ that annihilates all the vectors $\delta H_{\tilde{P}}/\delta u$, $\tilde{P} \in Z(\tilde{L})$. Then $S = 0$.*

9. Proof of Lemma (8.5)

In view of (4.4), what has to be proved is clearly a differential analogue of (the easier part of) the standard lemma on symmetric functions. The problem is a little obscured by the fact that the 'roots' of $\tilde{L}$ are not independent, but add up to zero. Let us formulate the corresponding result without this restriction; it will be needed in any case to handle the Lax equations with $u_{n-1} \neq 0$ (see Sect. 11).

(9.1) **Proposition.** *Let* $x_1, \ldots, x_n$ *be differentially independent variables, and define the 'non-commutative elementary symmetric functions'* $\sigma_1, \ldots, \sigma_n \in B(x)$ *by*

(9.2) $\xi^n - \sigma_1 \xi^{n-1} + \ldots + (-1)^n \sigma_n = (\xi - x_n)(\xi - x_{n-1}) \ldots (\xi - x_1).$

Then $\sigma_1, \ldots, \sigma_n$ *are differentially independent.*

The usual (commutative) version of this is generally proved by induction on n (one considers the effect of putting, say, $x_1 = 0$), but we did not succeed in imitating that argument. Instead, we imitate the following one.

(9.3) **Proposition.** *Let* $x_1, \ldots, x_n$ *be elements of a field* F, *algebraically independent over some subfield, say* $\mathbb{C}$. *Let* λ *be an indeterminate (commuting with the elements of* F *), and define* $\sigma_i \in F$ *by*

$$\lambda^n - \sigma_1 \lambda^{n-1} + \ldots + (-1)^n \sigma_n = \prod_1^n (\lambda - x_i).$$

Then $\sigma_1, \ldots, \sigma_n$ *are algebraically independent (over* $\mathbb{C}$*).*

Proof. Consider the field extensions

$$\mathbb{C} \subset \mathbb{C}(\sigma_1, \ldots, \sigma_n) \subset \mathbb{C}(x_1, \ldots, x_n).$$

The composite extension has transcendence degree n, and the second extension is algebraic; hence the first extension has transcendence degree n. That implies that the σ_i are algebraically independent (for otherwise we could choose a proper subset of $\{\sigma_1, \ldots, \sigma_n\}$ such that all the remaining σ_i were algebraically dependent on this subset: the extension $\mathbb{C} \subset \mathbb{C}(\sigma_i)$ would then have transcendence degree less than n).

To adapt this argument to prove (9.1), we have to develop the theory of 'differential fields' (∂-fields for short), that is, fields F equipped with a derivation $\partial: F \to F$. Most of the necessary work is done in [12], so we shall sketch the theory only briefly. For simplicity, and also so as to be able to refer to [14], we confine ourselves to the case of finite transcendence degree.

Let $E \subset F$ be an extension of ∂-fields. An element $x \in F$ is said to be ∂-algebraic[3] over E if some (nontrivial) differential polynomial in x with coefficients in E vanishes. An element $x \in F$ is said to be ∂-*dependent* (with respect to E, considered fixed) on a subset $\{y_1, \ldots, y_r\} \subset F$ if x is ∂-algebraic over $E(y_1, \ldots, y_r)$. (This last symbol naturally denotes the smallest ∂-field containing $E, y_1, \ldots, y_r$.) We have the following basic properties of ∂-dependence.

---

[3]   In [12] the confusing term 'algebraically transcendental' is used

(i) *Every y_i is ∂-dependent on $\{y_1, \ldots, y_r\}$.*

(ii) *If x is ∂-dependent on $\{y_1, \ldots, y_r\}$ but not on $\{y_1, \ldots, y_{r-1}\}$, then y_r is ∂-dependent on $\{y_1, \ldots, y_{r-1}, x\}$.*

(iii) *If x is ∂-dependent on $\{y_1, \ldots, y_r\}$ and each y_i is ∂-dependent on $\{z_1, \ldots, z_s\}$, then x is ∂-dependent on $\{z_1, \ldots, z_s\}$.*

It follows from (i), (ii) and (iii) that most of the properties of algebraic dependence (see [14]) also hold for ∂-dependence. In particular one can define the *∂-transcendence degree*[4] of an extension $E \subset F$ to be the number of elements in any maximal ∂-independent subset of F. (*A subset of F is ∂-independent (with respect to E) if no element of it is ∂-dependent on the rest: this coincides with our usual notion of differential independence.*) We mention explicitly the following.

(9.4) **Proposition.** *An extension $E \subset F$ has ∂-transcendence degree 0 if and only if every element of F is ∂-algebraic over E.*

(9.5) **Proposition.** *An extension $E \subset E(x_1, \ldots, x_n)$ generated by n elements has ∂-transcendence degree at most n; it has ∂-transcendence degree n if and only if the x_i are ∂-independent.*

(9.6) **Proposition.** *If $E \subset F \subset G$ with the extensions $E \subset F$, $F \subset G$ having ∂-transcendence degrees p, q, respectively, then the extension $E \subset G$ has ∂-transcendence degree $p + q$.*

Finally, we need the next proposition, which would be a tautology in the usual case (without derivations).

(9.7) **Proposition.** *Let $E \subset F$ be ∂-fields, and let $x_1, \ldots, x_n \in F$. Define $\sigma_1, \ldots, \sigma_n$ by (9.2). Then each x_i is ∂-algebraic over $E(\sigma_1, \ldots, \sigma_n)$, so that the extension*

$$E(\sigma_1, \ldots, \sigma_n) \subset E(x_1, \ldots, x_n)$$

has ∂-transcendence degree 0.

(9.8) **Lemma.** *Let A, B, C be differential operators with constant leading coefficients and $A = BC$. Then the coefficients of any one of A, B, C are differential polynomials in the coefficients of the other two.*

The proof of (9.8) is trivial.

Proof of (9.7). We use descending induction on i. By the transitivity of ∂-dependence (property (iii) above) it will suffice to show that if we rewrite (9.2) as

(9.9) $\xi^n - \sigma_1 \xi^{n-1} + \ldots = (\xi^{n-1} + a_1 \xi^{n-i-1} + \ldots)(\xi - x)(\xi^{i-1} + b_1 \xi^{i-2} + \ldots)$

then $x = x_i$ is ∂-algebraic over the field $E(\sigma_j, a_k)$. We should like to put $\xi = x$ in (9.9), but because of the non-commutativity we can not do that in the usual way. However, we can expand the right hand side of (9.9) and move the powers of ξ over to the right to put it in the form $\sum d_i \xi^i$; equality of two such expressions means of course that all the coefficients are equal, so we can then substitute x

---

[4]   'Hypertranscendence degree' in [12]

for ξ. Doing that yields a relation of the form

(9.10) $x^n - \sigma_1 x^{n-1} + \ldots = (\textit{some differential polynomial in } (a_j, x, b_k))$

By (9.8), the right hand side of (9.10) can be rewritten as a differential poly-nomial in (a_j, x, σ_k), so we have a relation of the kind that we want. It is non-trivial because, for example, the term x^n does not occur on the right of (9.10). We omit the proof of that.

Proof of (9.1). This can now proceed exactly like the proof of (9.3) given above.

Proof of (8.5). This follows the same lines: the only difference is that the ∂-transcendence degrees are now $n - 1$, rather than n.

10. Proof of Lemma (8.7)

Since we shall not be referring to the modified equations in this section, we abandon the tildes, and let L denote the scalar operator $\xi^n + \ldots + u_0$. Similarly P will denote an element of $Z(L) \subset B(u)[\xi, \xi^{-1}]$.

To explain the idea of the proof of (8.7), let us first give it in the simplest case $n = 2$. In that case it is well known that the $\delta H_P / \delta u$, rewritten in terms of v, have the form

$$\frac{\delta H_P}{\delta u} = \beta\, v^{(r)} + (\text{lower terms}),$$

where β is a non-zero constant and 'lower' means: involving only derivatives of order less than r. Suppose

$$S = s_q \partial^q + \ldots + s_0, \qquad s_i \in B(v), \qquad s_q \neq 0,$$

is a differential operator that annihilates all these. Choose r greater than the orders of any derivatives of v involved in any of the coefficients of S. Then in

$$S(\beta\, v^{(r)} + \ldots) = 0$$

the only term involving $v^{(r+q)}$ is $\beta s_q v^{(r+q)}$. Hence $s_q = 0$, a contradiction.

To imitate this argument in the general case, we need an analogue of the fact that $\beta \neq 0$. That is provided by the work of Veselov [13].

Let $P = \xi^r + \ldots$ be the homogeneous element of $Z(L)$ of order r; we recall that 'homogeneous' refers to the natural grading such that ξ has degree 1 and $u_i^{(j)}$ has degree $n - i + j$. We write P_- in both 'left' and 'right' notations, singling out the linear terms:

$$P_- = \sum_{j=0}^{n-2} \sum_{i=0}^{\infty} \alpha_{ij} u_j^{(r-n+1+i+j)} \xi^{-i-1} + \ldots$$

$$= \sum_{j=0}^{n-2} \sum_{i=0}^{\infty} \xi^{-i-1} \beta_{ij} u_j^{(r-n+1+i+j)} + \ldots.$$

(Here the dots indicate non-linear terms involving only lower derivatives of u_j than those indicated.) We recall from (7.11) that the first $n - 1$ 'right hand'

coefficients of P_- are precisely the variational derivatives $\delta H_{LP}/\delta u_i$ that we are interested in. Let β denote the matrix

$$\beta = (\beta_{ij}), \qquad 0 \leq i, j \leq n-2.$$

(10.1)   **Proposition** [13]. *If r is prime to n, the matrix β is non-singular.*

For completeness we indicate how this is proved. First, it is clearly equivalent to prove that the corresponding matrix $\alpha = (\alpha_{ij})$ is non-singular, since α and β are related by a lower triangular matrix with 1's on the diagonal. The α_{ij} are determined as follows.

(10.2)   **Lemma** [13]. *Let $au_j^{(q)}$ be a linear term occurring in any of the 'left-hand' coefficients of P. Then a is equal to the coefficient of z^q in the power series*

$$[(1+z)^r - 1][(1+z)^n - 1]^{-1}.$$

Proof. The following seems to us simpler than the proof given in [13]. Let $P(\varepsilon)$ (and similarly $L(\varepsilon)$) denote the operator obtained by replacing each u_j by εu_j; set

$$\bar{P} = \frac{\partial}{\partial \varepsilon} P(\varepsilon)|_{\varepsilon = 0}.$$

Clearly, $\bar{P}$ is the 'linear part' of P that we want to compute. Finally, let $\bar{P}_j$ denote the result of setting all the variables except u_j equal to zero. Differentiating the relation

$$[L(\varepsilon), P(\varepsilon)] = 0,$$

we easily find

(10.3) $[\xi^n, \bar{P}_j] = [\xi^r, u_j \xi^j].$

Let us associate to each homogeneous operator A of the form

$$A = \sum_{q=0}^{\infty} a_q u_j^{(q)} \xi^{N-q}$$

the formal power series $\sum a_q z^q$. Then the power series associated to $[\xi^p, A]$ is

$$(\sum a_q z^q)[(1+z)^p - 1].$$

Hence (10.2) follows at once if we equate the formal power series corresponding to the two sides of (10.3).

Lemma 10.2 shows that α is a 'Hankel matrix'. If r is prime to n, the numerator and denominator of the power series in (10.2) have no common factor except z; (10.1) then follows from the theory of Hankel matrices (see [13] and [3], Sect. 5, exercice 3).

To prove (8.7), we have to write the $\delta H/\delta u_i$ in terms of the modified variables v_j. (That is essential, since it is not clear *a priori* that the operator $D l D^*$ can be expressed in terms of the u_j.) So we need the following.

(10.4) **Lemma.** *Let the linear part of u_i be given by*

$$u_i = \sum_j \theta_{ij} v_j^{(n-i-1)} + (lower\ terms).$$

Then the matrix (θ_{ij}) is non-singular.

Proof. Introduce new variables $y_0, \ldots, y_{n-2}$ by

$$y_i = \sum_j \omega^{ij} v_j,$$

so that we have

(10.5) $\xi^n + u_{n-2}\xi^{n-2} + \ldots + u_0 = \left(\xi - \sum_0^{n-2} y_i\right)(\xi + y_{n-2})\ldots(\xi + y_0).$

Since the (Vandermonde) matrix (ω^{ij}) is non-singular, it is enough to show that if

$$u_i = \sum_j \varphi_{ij} y_j^{(n-i-1)} + (\text{lower terms}),$$

then the matrix (φ_{ij}) is non-singular. But that is clear, because (φ_{ij}) is triangular with non-zero entries on the diagonal. (To see that, we imagine ourselves calculating u_i by expanding (10.5), picking either a ξ or a y from each factor. To get a term linear in y, we must pick only one y; and to get $n-i-1$ derivatives, we have to pick at least that number of ξ's before the y.)

Combining (10.1) and (10.4), we get the result that we really need.

(10.6) **Proposition.** *Set*

$$\frac{\delta H_{LP}}{\delta u_i} = \sum_{j=1}^{n-1} \gamma_{ij} v_j^{(r+i)} + (lower\ terms).$$

Then if r is prime to n, the matrix (γ_{ij}) is non-singular.

Note. Here, unfortunately, we have i running from 0 to $n-2$, j from 1 to $n-1$.

Proof of (8.7). Let $S = \sum_0^q s_k \partial^k$ be a matrix of differential operators that annihilates all the vectors $\delta H_{LP}/\delta u$. Since of course that means that each row of S separately annihilates all the $\delta H_{LP}/\delta u$, we may as well assume that S has just one row; thus each s_k is a row vector with entries $s_{k;j} \in B(v)$, $0 \leq j \leq n-2$, and we assume that $s_q \neq 0$. Choose r large enough (prime to n) so that no derivatives $v_i^{(l)}$ with $l \geq r$ occur in any of the $s_{k;j}$. Then in the expression

(10.7) $S\dfrac{\delta H_{LP}}{\delta u} = 0$

the only terms involving derivatives of order $p = q + r + n - 2$ are

$$\sum_j s_{q;n-2} \gamma_{n-2,j} v_j^{(p)}.$$

Hence $s_{q;n-2}\gamma_{n-2,j}=0$ for all j. Since γ is non-singular, some $\gamma_{n-2,j}\neq0$; it follows that the last entry $s_{q;n-2}$ of s_q vanishes.

Next we consider the (remaining) terms in (10.7) involving derivatives of order $p-1$. These are

$$\sum_j (s_{q;n-3}\gamma_{n-3,j}+s_{q-1;n-2}\gamma_{n-2,j})\,v_j^{(p-1)}.$$

Hence the expressions in the brackets here vanish for all j. Since γ is non-singular, some 2×2 submatrix formed from the last two rows is non-singular. It follows that the second last entry of s_q, and also the last entry of s_{q-1}, vanish. Continuing this painful argument, we deduce after $n-1$ steps that $s_q=0$, a contradiction.

11. The Case $u_{n-1}\neq0$

In this section we take

$$\tilde{L}=\xi^n+u_{n-1}\,\xi^{n-1}+\ldots+u_0,$$

and we work over the algebra $B=B(u_0,u_1,\ldots,u_{n-1})$. The algebra of operators $B[\xi,\xi^{-1}]$ has its usual grading (deg $\xi=1$, deg $u_i^{(j)}=n-i+j$). We want to discuss what difference the extra variable u_{n-1} makes to our theory.

The first step is to determine the centralizer $Z(\tilde{L})$. The answer is the same as before: for each integer r, $Z(\tilde{L})$ contains a unique homogeneous operator $\tilde{P}=\xi^r+\ldots$; and $Z(\tilde{L})$ consists of the linear combinations of these. There are at least two ways of seeing that. First, we could use the method of fractional powers [11], which works equally well whether or not $u_{n-1}=0$. Second, we could deduce the result for $u_{n-1}\neq0$ from the one for $u_{n-1}=0$ by the usual device for getting rid of the second coefficient: introduce a new symbol ζ such that

$$\zeta^{-1}\,\partial\zeta=-(1/n)u_{n-1}.$$

Then for all q, $\zeta^{-1}\,\partial^q\zeta$ is a differential polynomial in u_{n-1}, so that conjugating by ζ gives an automorphism of the algebra $B[\xi,\xi^{-1}]$. The operator $\zeta^{-1}\tilde{L}\zeta$ has vanishing second coefficient, hence the result.

For each $\tilde{P}\in Z(\tilde{L})$ we can define a Lax equation as usual:

$$\partial_t\tilde{L}=[\tilde{P}_+,\tilde{L}]=[\tilde{L},\tilde{P}_-].$$

But we shall always have $\partial_t u_{n-1}=0$.

Next we consider the modified equations. We introduce an extra 'modified' variable v_0, and start off from the operator

$$L=\omega\text{-circ}(\xi+v_0,v_1,\ldots,v_{n-1}).$$

The centralizer $Z(L)$ has the same description as before (to prove it, conjugate by $\text{Id}\cdot\zeta$, where $\zeta^{-1}\,\partial\zeta=-v_0$). In particular, we have circulants P as in (4.2) defining our modified Lax equations. In these equations we always have

$\partial_t v_0 = 0$. As before we set $\tilde{L} = s(L^n)$; the variables u_i, v_i are related by

$$\tilde{L} = \tilde{L}_{n-1} \dots \tilde{L}_1 \tilde{L}_0,$$

where

(11.1) $$\tilde{L}_i = \xi + v_0 + \omega^i v_1 + \dots + \omega^{(n-1)i} v_{n-1}.$$

Note that $u_{n-1} = n v_0$.

Now we come to the Hamiltonian structure. The calculation in the proof of (8.2) shows that the 'second' Hamiltonian structure on $B(v) = B(v_0, \dots, v_{n-1})$ is as follows: let $x = (x_0, \dots, x_{n-1})$ be a 1-form; then the corresponding vector field is given by

$$\partial_x v_0 = -(1/n) \partial x_0$$
$$\partial_x v_i = -(1/n) \partial x_{n-i}, \qquad i = 1, \dots, n-1.$$

The corresponding skew matrix l is therefore just the direct sum of our previous l with an extra $-(1/n)\partial$ in the top left corner.

(11.2) **Theorem.** *This Hamiltonian structure restricts to give the second Hamiltonian structure on $B(u_0, \dots, u_{n-1})$ (Sect. 5, Example 4).*

The proof starts off like that of (8.6). The modified Lax equations can be written in Hamiltonian form with the above l and Hamiltonians H_p defined as before; it follows that the operator $D \, l \, D^* - \tilde{l}$ (which we want to prove to be zero) annihilates all the vectors $\delta H_{\tilde{p}}/\delta u$. However, the analogue of (8.7) is now false. To see that, note that since in the modified Lax equations $\partial_t v_0 = 0$, we must always have $\delta H_p / \delta v_0 = 0$. Hence if we were to change l by substituting anything else in the top left corner, the modified Lax equations would not know the difference; thus the Hamiltonians $H_{\tilde{p}}$, $\tilde{P} \in Z(\tilde{L})$ are clearly not enough to detect what we put in this corner. The remedy is clear: we have to throw in a few more Hamiltonians to tie down this corner entry. As our 'extra' Hamiltonians we take

$$H_r = (-1)^r \tfrac{1}{2}(u_{n-1}^{(r)})^2 = (-1)^r \tfrac{1}{2} n^2 (v_0^{(r)})^2.$$

(11.3) **Lemma.** *Let S be a matrix of differential operators (with coefficients in $B(v)$) that annihilates all the vectors $\delta H_r/\delta u$. Then the last column of S vanishes.*

Proof. Since obviously

$$\frac{\delta H_r}{\delta u} = (0, \dots, 0, u_{n-1}^{(2r)})^t,$$

each entry in the last column of S annihilates all the functions $v_0^{(2r)}$. It follows easily that these entries vanish (see the argument at the beginning of Sect. 10).

(11.4) **Lemma.** *Let $S = \sum_0^q s_k \partial^k$ be a matrix of operators that annihilates all the vectors $\delta H/\delta u$ where either $H = H_{\tilde{p}}$ or $H = H_r$. Then $S = 0$.*

Proof. By (11.3), the last column of S vanishes. For each $\tilde{P} \in Z(\tilde{L})$, define the matrix $\gamma = (\gamma_{ij})$ as in (10.6), but with i, j now running from 0 to $n-1$. It is easy to

see that the $(n-1) \times (n-1)$ matrix obtained by deleting the last row and first column of γ is precisely the matrix γ that we had in Sect. 10; it is therefore non-singular (for infinitely many $\tilde{P}$). Now we can use the same argument as in the proof of (8.7) to show that the remaining $n-1$ columns of the leading coefficient s_q vanish. (Since the last column of S is zero, the last row of γ will play no part in this argument.)

Proof of (11.2). It remains only to show that the operator $DlD^* - \tilde{l}$ annihilates the vectors $\delta H_r/\delta u$. We do that by direct calculation. Since

$$\frac{\delta H_r}{\delta v} = (n^2 \, v_0^{(2r)}, 0, \ldots, 0)^t,$$

the Hamiltonian vector field on $B(v)$ corresponding to H_r is

$$\partial_H v_0 = -n \, v_0^{(2r+1)}$$
$$\partial_H v_i = 0, \quad i > 0.$$

Hence for all i we have

$$\partial_H \tilde{L}_i = -n \, v_0^{(2r+1)} = -u_{n-1}^{(2r+1)},$$

the $\tilde{L}_i$ being as in (11.1). The restriction of ∂_H to $B(u)$ is therefore given by

$$\partial_H \tilde{L} = -\sum_{i=0}^{n-1} \tilde{L}_{n-1} \ldots \tilde{L}_{i+1} u_{n-1}^{(2r+1)} \tilde{L}_{i-1} \ldots \tilde{L}_0.$$

On the other hand, let us calculate the second Hamiltonian vector field $\tilde{\partial}_H$ on $B(u)$ determined by H_r, using the definition in Sect. 5, Example 4. We have

$$X = \xi^{-n} u_{n-1}^{(2r)}; \quad \text{so} \quad (\tilde{L}X)_+ = (X\tilde{L})_+ = u_{n-1}^{(2r)},$$

hence

$$\tilde{\partial}_H \tilde{L} = [u_{n-1}^{(2r)}, \tilde{L}].$$

Writing $\tilde{L}$ as the product of the $\tilde{L}_i$ and using the relations

$$\tilde{L}_i u_{n-1}^{(2r)} = u_{n-1}^{(2r)} \tilde{L}_i + u_{n-1}^{(2r+1)},$$

we see at once that this is the same expression as before.

To conclude on a more dignified note, let us point out that Theorem (11.2) takes an even simpler form when expressed in terms of the actual 'roots' of $\tilde{L}$, rather than our variables v_i. That is, let us set

$$\xi^n + u_{n-1} \xi^{n-1} + \ldots + u_0 = (\xi - x_{n-1}) \ldots (\xi - x_1)(\xi - x_0).$$

Thus we have $x_i = -\sum \omega^{ij} v_j$, giving an isomorphism $B(x) \approx B(v)$. A short calculation shows that if we transfer our Hamiltonian structure on $B(v)$ to $B(x)$ via this isomorphism, the resulting structure is given by the matrix $l = -\text{Id} \cdot \partial$. Thus we have the following.

(11.5) **Theorem.** *The Hamiltonian structure on $B(x_0, \ldots, x_{n-1})$ defined by the matrix $l = -\operatorname{Id}.\partial$ restricts to give the second Hamiltonian structure on $B(u_0, \ldots, u_{n-1})$.*

There must surely be a more direct proof of this simple assertion than the one we have just given.

References

1. Adler, M.: On a trace functional for formal pseudodifferential operators and the symplectic structure of the Korteweg-de Vries type equations. Inventiones Math. **50**, 219–248 (1979)
2. Aitken, A.C.: Determinants and matrices, Edinburgh and London: Oliver and Boyd, 1939
3. Bourbaki, N.: Algèbre, ch. 4. Paris: Hermann, 1950
4. Chen, H.-H.: Relation between Bäcklund transformations and inverse scattering problems. In: 'Bäcklund transformations' (R.M. Miura, ed.), Lecture notes in mathematics 515, pp. 241–252. Berlin-Heidelberg-New York: 1976
5. Gel'fand, I.M., Dikii, L.A.: A family of Hamiltonian structures connected with integrable non-linear differential equations. Preprint, Inst. Appl. Math. Acad. Sci. USSR, 1978, no. 136 (Russian)
6. Gel'fand, I.M., Dorfman, I.Ya.: Hamiltonian operators and algebraic structures related to them. Funct. Anal. Appl. **13**, 13–30 (1979) (Russian), 248–262 (English)
7. Kirillov, A.A.: Elements of the theory of representations. Berlin-Heidelberg-New York: Springer-Verlag 1976
8. Kruskal, M.D.: Non-linear wave equations. In: Lecture notes in physics 38. Berlin-Heidelberg-New York: Springer-Verlag 1975
9. Lebedev, D.R., Manin, Yu.I.: The Gel'fand-Dikii Hamiltonian operator and the co-adjoint representation of the Volterra group. Funct. Anal. Appl. **13**, 40–46 (1979) (Russian), 268–273 (English)
10. Magri, F.: A simple model of the integrable Hamiltonian equation. J. Math. Phys. **19**, 1156–1162 (1978)
11. Manin, Yu.I.: Algebraic aspects of non-linear differential equations. Itogi Nauki i Tekhniki, Ser. Sovremennye Problemy Matematiki **11**, 5–152 (1978) (Russian). English translation in J. Sov. Math. **11**, 1–122 (1979)
12. Raudenbush, H.W., Jr.: Differential fields and ideals of differential forms. Ann. Math. **34**, 509–517 (1933)
13. Veselov, A.P.: On the Hamiltonian formalism for the Novikov-Krichever equations of commutativity of two operators. Funct. Anal. Appl. **13**, 1–7 (1979) (Russian), 1–6 (English)
14. Van der Waerden, B.L.: Modern Algebra, vol. 1, second edition. New York: Ungar, 1949
15. Wilson, G.: Commuting flows and conservation laws for Lax equations. Math. Proc. Camb. Phil. Soc. **86**, 131–143 (1979)
16. Adler, M., Moser, J.: On a class of polynomials connected with the Korteweg-de Vries equation. Comm. Math. Phys. **61**, 1–30 (1978)
17. Reyman, A.G., Semenov-Tian-Shanskii, M.A.: A family of Hamiltonian structures, a hierarchy of Hamiltonians and reduction for first order matrix differential operators. Funct. Anal. Appl. **14**, 77–78 (1980) (Russian)
18. Sokolov, V.V., Shabat, A.B.: (L, A)-pairs and a substitution of Riccati type. Funct. Anal. Appl. **14**, 79–80 (1980) (Russian)
19. Gel'fand, I.M., Dorfman, I.Ya.: The Schouten bracket and Hamiltonian operators. Funct. Anal. Appl. **14**, 71–74 (1980) (Russian)

Note Added in Proof

The additional reference [19] below contains a proof that the second structure is Hamiltonian: it is a more sophisticated version of the proof in the unpublished preprint [5]. The proof is valid also in the case of a matrix operator L with $u_{n-1,\alpha\alpha} \neq 0$, which confirms the opinion expressed at the end of Sect. 5 above that this generalization would not present any extra problems.