

# The Twistor Theory of Equations of KdV Type: I

L.J. Mason<sup>1</sup>\*, M.A. Singer<sup>2</sup>\*\*

<sup>1</sup> The Mathematical Institute, 24–29 St Giles, Oxford OX1 3LB, U.K. E-mail: lmason@maths.oxford.ac.uk

<sup>2</sup> Department of Mathematics and Statistics, James Clerk Maxwell Building, King's Buildings, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, U.K. E-mail: msinger@spam.adelaide.edu.au Fax (Singer) + 618 232 5670

Received: 29 October 1993

**Abstract:** This article is the first of two concerned with the development of the theory of equations of KdV type from the point of view of twistor theory and the self-dual Yang–Mills equations. A hierarchy on the self-dual Yang–Mills equations is introduced and it is shown that a certain reduction of this hierarchy is equivalent to the  $n$ -generalized KdV-hierarchy. It also emerges that each flow of the  $n$ -KdV hierarchy is a reduction of the self-dual Yang–Mills equations with gauge group  $SL_n$ . It is further shown that solutions of the self-dual Yang–Mills hierarchy and their reductions arise via a generalized Ward transform from holomorphic vector bundles over a twistor space. Explicit examples of such bundles are given and the Ward transform is implemented to yield a large class of explicit solutions of the  $n$ -KdV equations. It is also shown that the construction of Segal and Wilson of solutions of the  $n$ -KdV equations from loop groups is contained in our approach as an ansatz for the construction of a class of holomorphic bundles on twistor space.

A summary of the results of the second part of this work appears in the Introduction.

## 1. Introduction

In the subject “Integrable systems and self-duality” – which has seen much recent activity [Wa, MS, M, MW] – one can identify two clear goals. First, to relate as many integrable systems as possible to the self-duality equations; and secondly, to understand the many techniques for generating solutions of integrable systems in terms of the twistor description of the self-duality equations.

This paper and its sequel are devoted to integrable systems of Korteweg de Vries (KdV) type. As for the first of the above goals, we relate the  $n$ -generalized KdV hierarchy to a self-dual Yang–Mills (SDYM) hierarchy and give the corresponding twistor description. As for the second goal, we describe how the twistor description bears on many of the methods associated with these hierarchies: the solutions

---

\* SERC Advanced Fellow

\*\* Most of this work was done while Darby Fellow of Mathematics at Lincoln College, Oxford

constructed by Segal and Wilson from loop groups, the Krichever construction of solutions from Riemann surfaces, and the  $\tau$ -function. The reader is referred to [SW] for a convenient account of this standard material.

*1.1. Outline of Results.* After a rapid review in Sect. 2 of the basic definitions and facts about the  $n$ -KdV hierarchy, we define in Sect. 3 a SDYM hierarchy for each pair of integers  $(m, K)$  and Lie group  $G$ . This is a system of non-linear partial differential equations on a  $G$ -connection and a system of Higgs fields on a trivial bundle over  $\mathbb{C}^{(m+1)K}$ . When  $(m, K) = (1, 2)$  the self-dual Yang–Mills equations are recovered and when  $K = 1$ , the system reduces to the Bogomolny hierarchy (with  $m$  levels) of [MS]. Our first main result is Theorem 3.1: a certain reduction (by a group of  $(n - 1)$  translations of  $\mathbb{C}^{(m+1)(n-1)}$ ) of the  $SL_n(\mathbb{C})$ -SDYM hierarchy of order  $K = n - 1$  is essentially equivalent to the  $n$ -KdV hierarchy. We also show that any of the integrable equations in 2 variables that is obtainable from the  $n$ -KdV hierarchy is a reduction of the (standard) SDYM equations.

In Sect. 4, we show that the SDYM hierarchy (and the reduction relevant to  $n$ -KdV) has a twistor description via a generalization of the Ward transform. In outline, for each  $(m, K)$ , there is a twistor space  $Z$  of complex dimension  $(K + 1)$ , which has the structure of the vector bundle  $p : Z = \mathbb{C}^K \otimes \mathcal{O}(m) \rightarrow \mathbb{C}P^1$  where  $\mathcal{O}(m) \rightarrow \mathbb{C}P^1$  is the complex line bundle on  $\mathbb{C}P^1$  of Chern class  $m$ ; the associated “space-time”  $X$ , being the space of holomorphic sections of  $p$ , is  $\mathbb{C}^{(m+1)K}$ . The Ward transform gives a 1:1-correspondence between (generic) holomorphic  $SL_n(\mathbb{C})$ -bundles over regions in  $Z$  and local holomorphic solutions of the first  $(m + 1)$  levels of the  $SL_n$ -self-dual Yang–Mills hierarchy of order  $K$  over  $X$ . Applying this to the appropriate reduction of the SDYM hierarchy, we obtain our second main theorem (Theorem 4.2): the notation of the present paragraph being retained, take  $K = n - 1$ . Then there is a 1:1-correspondence between (generic) *invariant* holomorphic  $SL_n(\mathbb{C})$ -bundles over  $Z$  and holomorphic solutions of the first  $m(n - 1)$  flows of the  $n$ -KdV hierarchy. The term “invariant” is defined precisely in Sect. 4.

In Sect. 5, we explain the intimate relation between the solutions of  $n$ -KdV constructed from loop groups by Segal and Wilson and the twistor description of Sect. 4. We also give an explicit treatment of some interesting examples of the construction and a characterization of the class of solutions that arise from loop groups in this way.

It is worth stressing, however, that the twistor construction is strictly more general than that of Segal and Wilson. What is proved in Sect. 5 is that the *restriction* of the Ward transform to a class of bundles associated to the loop group coincides with the construction of Segal and Wilson.

*1.2. Summary of Part II.* In a future publication, the twistor description developed here will be used to investigate other aspects of the theory of equations of KdV type. We include here a summary of this work.

*1.2.1. The  $\tau$ -function.* The  $\tau$ -function (cf. Sect. 2.1.2) is a potential for the solution  $u$  of the KdV equations,

$$u = 2\partial_x^2 \log \tau. \quad (1.1)$$

It plays a basic role in Sato theory and in [SW] where it measures the action of  $\Gamma_+$  on the canonical section of the determinant line bundle over the Grassmannian  $Gr$  of Hilbert space.

We give a description of the  $\tau$ -function that arises out of the twistor geometry. The main idea is to use the determinant bundle over the space of  $\bar{\partial}$ -operators defined by Quillen rather than the determinant line bundle over Gr.

Recall that the space of  $\bar{\partial}$ -operators on a vector bundle  $E$  over  $\mathbb{C}P^1$  (or indeed any Riemann surface) can be identified with the complex Hermitian affine space  $\mathcal{A}$  of unitary connections on  $E$ . Thus  $\mathcal{A}$  has a natural symplectic structure  $\Omega$ . Quillen showed how to define a complex line-bundle Det over  $\mathcal{A}$  and also how to equip it with a natural connection whose curvature is  $\Omega$ . Defined in this way, Det also has a canonical section  $\sigma$  whose value at  $a \in \mathcal{A}$  is to be thought of as the determinant of the  $\bar{\partial}$ -operator corresponding to  $a$ .

Now this theory has relevance to the twistor description of  $n$ -KdV, because the bundle  $E \rightarrow Z$  may be characterized by its  $\bar{\partial}$ -operator relative to a fixed smooth trivialization. On pulling back to the correspondence space  $Y = X \times \mathbb{C}P^1$  we obtain a family of  $\bar{\partial}$ -operators on a vector bundle  $E$  over  $\mathbb{C}P^1$  parametrized by  $X$ . According to the general theory alluded to in the previous paragraph, this family defines a determinant bundle over  $X$ . The action of translations of  $X$  is symplectic, so one can lift it to Det using the standard geometric quantization formula. The  $\tau$ -function measures the action of the translations on  $\sigma$ . It is defined by

$$\sigma(x) = \tau(x)V(x - x_0)\sigma(x_0),$$

where  $V(x)$  is the action of the translation along  $x$  on Det and  $x_0$  is a basepoint in  $X$ . We show that  $\tau$  defined in this way satisfies (1.1).

*1.2.2. Connections with Quantum Field Theory.* The  $\tau$ -function has been expressed as a vacuum expectation value of certain operators in a free Fermionic quantum field theory [JM]. We show that this quantum field theory is naturally identified with the quantum field theory of holomorphic sections of the twistor vector bundle restricted to each sphere in  $\mathbb{C}P^1$ .

*1.2.3. Spectral Curves and the Krichever Construction.* The Krichever construction is a method for constructing solutions of  $n$ -KdV that are invariant under one of the higher flows of the hierarchy. Given such an invariant solution one constructs, in a natural way, a spectral curve  $\Sigma$ . Then one defines a natural map from “space-time” to the Jacobian of  $\Sigma$ ,

$$X \rightarrow \text{Jac}(\Sigma). \tag{1.2}$$

The importance of the Krichever construction lies in the fact that (1.2) is affine linear, so the flows of the  $n$ -KdV hierarchy go over to linear flows on the Jacobian.

The Krichever construction should be thought of as the transformation of  $n$ -KdV to action-angle variables: the action variables are the moduli of  $\Sigma$ , and the angle variables are the linear coordinates on  $\text{Jac}(\Sigma)$ .

We show how the Krichever construction arises out of the twistor construction. The key point is that if the holomorphic bundle  $E \rightarrow Z$  corresponds to an invariant solution of  $n$ -KdV, then one can construct, in a natural way, a global section  $\Phi$  of  $\text{End}(E)$  that is invariant.

The spectral curve  $\Sigma$  is then given by the equation  $\det(\Phi - \nu) = 0$  and is thus a branched cover of  $\mathbb{C}P^1$ . Now pull back the twistor space  $Z$  to  $\Sigma$ , to get a branched covering  $\tilde{Z}$  of  $Z$ . The dual of the bundle  $E$  on  $Z$  is determined as the pushdown from  $\tilde{Z}$  of the dual of the line bundle  $L = \ker(\Phi - \nu)$ . Thus  $E^*$  is determined by the line bundle  $L$  on  $\tilde{Z}$ . (This is the idea of “abelianization” of Nigel Hitchin.)

To describe the map (1.2), note that each point  $x \in X$  determines a copy  $\Sigma_x$  of  $\Sigma$  inside  $\tilde{Z}$ . Then we claim that the assignment  $x \mapsto L|_{\Sigma_x}$  induces (1.2): viewed in this way, the linearity of this map is immediate.

## 2. Equations of KdV Type

The standard KdV equation for the height  $u(x, t)$  of water waves in a shallow channel

$$\partial_t u = \partial_x^3 u + 6u\partial_x u,$$

is the first in a family of equations of “KdV type.” This section gives an account of this family of equations. We begin with the Gelfand–Dikii description in terms of ordinary differential operators. Then we describe how the KdV equations fit into the framework of Drinfeld–Sokolov, and give a very brief account of the more general integrable systems constructed by these authors. The reader is referred to [GD, SW] for the Gelfand–Dikii theory and to [DS, W] for Drinfeld–Sokolov. These are convenient sources: this is by no means a complete list of references on this subject.

### 2.1. Gelfand–Dikii Theory

**2.1.1. Definition of the Hierarchies.** Consider the space  $M$  of ordinary differential operators of the form

$$L = D^n + u_{n-2}D^{n-2} + \dots + u_0 \quad (D = d/dx), \tag{2.1}$$

where  $u_0, \dots, u_{n-2}$  are functions of  $x$ . A flow of the  $n$ -KdV hierarchy is by definition an evolution equation of the form

$$\frac{dL}{dt} = [P, L], \tag{2.2}$$

where  $P$  is any differential operator of the form  $D^m + v_{m-2}D^{m-2} + \dots + v_0$ . It is known that for each integer  $m \geq 2$ , there exists an operator,  $P_m$  say, of this form, such that  $[P, L]$  is a differential operator of order  $n - 2$ , and hence such that (2.2) makes sense.  $P_m$  is determined essentially uniquely by  $L$ : the only freedom in  $P_m$  is that of adding on linear combinations, with constant coefficients, of the  $P_k$  with  $k \leq m - 2$ . We shall always assume that the  $P_m$  are chosen so that  $P_{rn} = L^r$  for each  $r = 1, 2, \dots$ . The corresponding evolution equations are trivial: it is sometimes convenient to retain them, and sometimes better to throw them away.

As an illustration, let us find  $P = D^2 + v$  when  $n = 3$ . Since

$$[P, L] = (2u'_1 - 3'v)D^2 + (u''_1 + 2u''_0 - 3v'')D + (u''_0 - u_1v' - v''') \tag{2.3}$$

one must have  $v = 2u_1/3$  and then (2.2) yields:

$$\begin{aligned} \dot{u}_1 &= 2u'_0 - u''_1, \\ \dot{u}_0 &= u''_0 - 2u'''_1/3 - 2u_1u'_1/3. \end{aligned}$$

$u_0$  can be eliminated by differentiating the first with respect to  $t$  and twice with respect to  $x$ , and the second with respect to  $x$ . Then  $u_1$  satisfies the Boussinesq equation

$$3\ddot{u}_1 = -u_1'''' - 4(u_1u_1')'$$

It is convenient to work with all evolution equations of the form (2.2) simultaneously; to do so, introduce a sequence  $t_1 = x, t_2, \dots$  of flow parameters or “times,” denoted collectively by  $\mathbf{t}$ , and consider the  $n$ -KdV hierarchy of evolution equations:

$$\partial_r = [P_r, L] \quad r = 2, 3, \dots \tag{2.4}$$

It is known (and will re-emerge below) that

$$[\partial_r - P_r, \partial_s - P_s] = 0 \text{ for all } r \text{ and } s. \tag{2.5}$$

This fact implies that it is indeed consistent to consider all flows simultaneously as in (2.4). It is sometimes convenient to refer to the set of evolution equations (2.5) also as the  $n$ -KdV hierarchy: notice that (2.2) arises when  $r = n$  and  $s = m$  (recall that everything in sight is independent of  $t_n$ ). We shall also often refer to the collection of operators  $\partial_r - P_r$  as the  $n$ -KdV hierarchy; no confusion appears to result from this practice.

2.1.2. *Baker and  $\tau$ .* The Baker function  $\psi$  and the  $\tau$ -function are important parts of the general theory. The former is a (formal) solution of the equations

$$\partial_r \psi = P_r \psi \tag{2.6}$$

of the form  $\psi = g\tilde{\psi}$ , where

$$g(\mathbf{t}, z) = \exp\left(\sum_1^\infty t_r z^r\right) \quad \text{and} \quad \tilde{\psi} = 1 + \sum_1^\infty a_r z^{-r}. \tag{2.7}$$

Then  $\psi$  is unique up to multiplication by a formal power series  $1 + c_1 z^{-1} + \dots$  with constant coefficients.

In (2.7), the  $a_r$  are functions only of  $\mathbf{t}$  and the sums are best treated formally; only for a restricted (but interesting) class of solutions will  $\tilde{\psi}$  exist as a holomorphic function in a neighbourhood of  $z = \infty$ . Its asymptotic expansion is well-defined, however, and since one is usually only interested in the expansion to finite order, convergence is not of great importance at present.

We shall always assume that the  $a_r$  are chosen to be independent of  $t_{rn}$  for each  $r$ ; then (2.6) with  $r = n$  is equivalent to the equations

$$L\psi = z^n \psi \text{ or } L^{(z)}\tilde{\psi} = z^n \tilde{\psi}. \tag{2.8}$$

In the latter,  $L^{(z)}$  is the operator obtained by replacing  $D$  by  $D + z$  in  $L$ . From either of these equations one sees that  $\tilde{\psi}$  determines  $L$  in the sense that if it is known that there exists an operator of the form (2.1) for which (2.8) holds, then the  $u$ 's can be found from the  $a$ 's. For example,

$$u_{n-2} = -na_1'; \tag{2.9}$$

$$u_{n-3} = -n(n-1)a_1''/2 - na_2' + na_1a_1'; \tag{2.10}$$

and more generally,  $a_1, \dots, a_{n-1}$  determine  $u_{n-2}, \dots, u_0$ .

The  $\tau$ -function is a “generating function” for the power-series part  $\tilde{\psi}$  of the Baker function: there is an equality of formal series

$$\tilde{\psi}(\mathbf{t}, z) = \tau^{-1}(\mathbf{t})\tau(x - z^{-1}, t_2 - (2z)^{-1}, t_3 - (3z^3)^{-1}, \dots). \tag{2.11}$$

For example,

$$a_1 = -\tau^{-1} \partial_x \tau; \tag{2.12}$$

$$a_2 = -(2\tau)^{-1} (\partial_2 \tau - \partial_{xx}^2 \tau). \tag{2.13}$$

Combining with (2.9) and (2.10), we obtain expressions for the  $u$ 's in terms of  $\tau$ :

$$u_{n-2} = n \partial_{xx}^2 \log \tau; \tag{2.14}$$

$$2u_{n-3} = n \partial_2 \partial_x \log \tau - n \partial_x ((\partial_x^2 \tau) / \tau) + n(n-1) \partial_x^3 \log \tau + 2n (\partial_x^2 \log \tau) (\partial_x \log \tau). \tag{2.15}$$

**2.2. Drinfeld–Sokolov Theory.** Elegant as the previous description of  $n$ -KdV is, we shall find an alternative, due to Drinfeld and Sokolov [DS], in terms of first-order matrix-valued operators, better suited to our purpose. The relation between the two descriptions corresponds roughly to the familiar relation between scalar ODEs of  $n$ -th order and *systems* of  $n$  ODEs of first order. Drinfeld and Sokolov do much more in [DS]. They associate to each Kac–Moody algebra (plus certain other data) an integrable system; applied to the loop algebra of  $SL_n$ , their construction yields  $n$ -KdV. The association of integrable systems to Kac–Moody algebras has been generalized still further, cf. [KW, GHM]. We shall not give a complete discussion of the theory of [DS]. We shall however give brief indications of the modifications that are needed to treat the systems that arise from other Kac–Moody algebras.

**2.2.1. Definitions.** The space of dependent variables  $M$  is the space of  $N$ -gauge equivalence classes of connections  $L_1$  on the line, of the form

$$L_1 = \partial_x + a - A, \tag{2.16}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & & 1 \\ \lambda & 0 & 0 & \dots & 0 \end{bmatrix}, \tag{2.17}$$

$a$  is lower triangular and trace-free, and  $N$  is the group of strictly (i.e. with 1s on the diagonal) lower triangular matrices depending only on  $x$ . We always consider  $L_1$  to act on  $n$ -component column-vectors  $\vec{v}$  on the line.

In every  $N$ -orbit there is a unique element with

$$a = \begin{bmatrix} 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ u_0 & \dots & u_{n-2} & 0 \end{bmatrix} \tag{2.18}$$

Consider the equation  $L_1 \vec{v} = 0$  in this gauge. If the top component of  $\vec{v}$  is  $\psi$ , then this is equivalent to (2.8) with  $\lambda = z^n$ . This observation is the essence of the identification between the present description of  $M$  and the one given in the previous section.

In every  $N$ -orbit there is also a unique element with  $a$  diagonal. This choice identifies  $M$  with the space of dependent variables of the so-called “modified”  $n$ -KdV hierarchy. From the present point of view it is unnatural to make any distinction between these hierarchies.

We shall be interested in flows on  $M$  defined by the condition

$$[L_1, L_r] = 0, \tag{2.19}$$

where

$$L_r = \partial_r - b_r, r = 2, 3, \dots, \tag{2.20}$$

can be regarded as another component of the connection and  $b_r$  is trace-free and has entries that are polynomial in  $\lambda$ . Such flows correspond to evolution equations like (2.2).

**2.2.2. The Formal Dressing Transformation.** The main technical tool that we use in our study of the  $n$ -KdV hierarchy is the formal dressing transformation. This is an  $SL_n(\mathbb{C}[\lambda^{-1}])$ -valued function of (in the first instance)  $x$ , i.e. an  $n \times n$  matrix, each of whose entries is a formal power series in  $\lambda^{-1}$ , with coefficients that are functions of  $x$ , and unit determinant. The key fact is that given any  $L_1$  in  $M$ , there exists a *formal dressing transformation*  $g \in SL_n(\mathbb{C}[\lambda^{-1}])$  such that

$$L_1 = g(\partial_x - A)g^{-1}, \quad g(\lambda = \infty) \in N. \tag{2.21}$$

Moreover, if  $g_1$  and  $g_2$  both satisfy (2.21), then

$$g_1 = g_2 s, \text{ where } s = 1 + \sum_{j=1}^{\infty} s_j A^{-j} \tag{2.22}$$

and the  $s_j$  are constants. See [DS, Sect. 3] for a derivation of this. They prove a slightly different result; but a further dressing of their form with  $1 + \sum_{i < 0} f_i A^i$  will give (2.21).

The dressing transformation is closely related to the Baker function  $\psi$ . Begin by noting that Eq. (2.8)

$$L\psi(x, z) = \lambda\psi(x, z)$$

holds for each of the  $n$ -th roots,  $z_1, \dots, z_n$  say, of  $\lambda$ :

$$L(\psi(x, z_1), \dots, \psi(x, z_n)) = \lambda(\psi(x, z_1), \dots, \psi(x, z_n)).$$

Multiplying on the right by  $A(\lambda)^{-1}$ ,

$$A(\lambda) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_n \\ \dots & \dots & \dots & \dots \\ z_1^{n-1} & z_2^{n-1} & \dots & z_n^{n-1} \end{bmatrix}, A(\lambda)^{-1} = n^{-1} \begin{bmatrix} 1 & z_1^{-1} & \dots & z_1^{1-n} \\ \dots & \dots & \dots & \dots \\ 1 & z_n^{-1} & \dots & z_n^{1-n} \end{bmatrix} \tag{2.23}$$



In this gauge, since  $L_1$  contains no non-negative powers of  $\lambda$ , (2.21) gives the formula

$$L_1 = \partial_x - (gAg^{-1})_+,$$

where the subscript  $+$  on any function of  $\lambda$  and  $\lambda^{-1}$  denotes the part which involves the non-negative powers of  $\lambda$ . If, as before, the expansion of  $g$  in powers of  $\lambda^{-1}$  is  $g = 1 + g_1\lambda^{-1} + O(\lambda^{-2})$  then we find

$$L_1 = \partial_x + [e_1, g_1] - A, \tag{2.31}$$

where  $e_1$  is the coefficient of  $\lambda$  in  $A$ . It follows that  $a$  is “L-shaped” (its only non-zero entries appear in the first column and the last row).

The dependent variables  $u_0, \dots, u_{n-2}$  can be recovered from  $g_1$  by using (2.28, 2.29), and the relations between the  $u$ 's and the  $a$ 's. The simplest formula of this kind is

$$a_1 = \text{tr}(e_1 g_1).$$

From this and (2.9), we find

$$u_{n-2} = -n\partial_x \text{tr}(e_1 g_1). \tag{2.32}$$

We can use an equation analogous to (2.21) to define  $L_r$ :

$$L_r = g(\partial_r - A^r)g^{-1} = \partial_r - (gA^r g^{-1})_+. \tag{2.33}$$

Notice that the equality of these two expressions for  $L_r$  implies a specific dependence of  $g$  on  $t_r$ , and is equivalent to the  $r$ -th evolution equation (2.6) satisfied by  $\psi$ .

From (2.33), we can read off a number of important properties of the  $L_r$ . For example, we see that

$$b_r = (gA^r g^{-1})_+ = A^r + O(\lambda^{r-1}) \tag{2.34}$$

in (2.20). Moreover, the  $L_r$  commute among themselves:

$$[L_r, L_s] = g[\partial_r - A^r, \partial_s - A^s]g^{-1} = 0. \tag{2.35}$$

Also, if  $r$  is a multiple of  $n$ , then the corresponding flow is trivial (for  $A^n = \lambda I$ ). Finally, the uniqueness statement about the dressing transformation (2.22) yields the essential uniqueness of the  $n$ -KdV hierarchy. To give a precise statement, let  $L_r$  be as in (2.34) and suppose that we have an operator  $L' = \partial_t - b'$  such that the equation  $[L_1, L'] = 0$  makes sense. The “essential uniqueness” of the  $n$ -KdV hierarchy is the statement that  $L'$  is a constant linear combination of a finite number of the  $L_r$  and follows from Eq. (2.34) and the fact that any differential operator that commutes with  $\partial_x - A$  has the form  $\partial_t$  minus a constant linear combination of the  $A^s$ .

It should be stressed that although the radius of convergence about  $\lambda = \infty$  of  $g$  is in general equal to 0, the  $b_r$  are perfectly well defined, since the definition of  $b_r$  involves only the first  $r$  terms in the formal expansion of  $g$ . In fact, if  $N > 0$  and  $g_N$  is the  $N$ -th partial sum of  $g$ , we refer to  $g_N$  as a truncation of  $g$ . Then

$$L_1 = g_N(\partial_x - A)g_N^{-1} + O(\lambda^{-N});$$

if one is concerned with only the first  $k < N$  flows of the  $n$ -KdV hierarchy, then  $g_N$  can be used instead of  $g$  throughout.

**2.2.3. The General Drinfeld–Sokolov Systems.** Drinfeld and Sokolov associate a completely integrable system to the following data:

1. a Kac–Moody algebra  $\mathcal{G}$ ;
2. a representation of  $\mathcal{G}$  as a subalgebra of a Loop algebra (this is referred to in [DS] as a choice of a “standard gradation” and these in turn correspond to the choice of a vertex of the Dynkin diagram of  $\mathcal{G}$ ).

In this scheme, the  $n$ -KdV hierarchy corresponds to the loop algebra of  $SL_n$  in its standard representation (the vertex associated to this gradation is then the “special vertex” of the Dynkin diagram of the loop algebra of  $SL_n$ ).

The given representation enables one to expand each element of  $\mathcal{G}$  in powers of  $\lambda$  and hence defines a grading  $\mathcal{G} = \bigoplus_{i=-\infty}^{\infty} \mathcal{G}_i$ . In this set-up,

$$L_1 = \partial_x + a - A,$$

where now  $a$  is to lie in the intersection of  $\mathcal{G}_0$  with the Borel subalgebra (i.e. the subalgebra generated by the Cartan subalgebra and the negative roots),  $A$  is taken to be the sum of the positive roots of  $\mathcal{G}$  in this representation and the gauge freedom is taken to be generated by the intersection of  $\mathcal{G}_0$  with the nilpotent subalgebra generated by the negative roots.

The matrix  $A$  again has the property that it is linear in  $\lambda$  (the chosen root has grade 1 and thus is multiplied by  $\lambda$  whereas the others have grade 0 and so are multiplied by  $\lambda^0$ ). The elements of the Lie algebra that commute with  $A$  form an abelian subalgebra  $\mathcal{L}$ . The flows are again constructed by dressing, where the role of the  $A^i$  is now played by the generators of the intersection of  $\mathcal{L}$  with  $\mathcal{G}^+ = \bigoplus_{i \geq 0} \mathcal{G}_i$ . (In the above a good system of such generators was provided by  $A^i, i \geq 1$ .)

This framework has been generalized still further by [KW] and [GHM].

### 3. The Self-Dual Yang–Mills Hierarchy

In the previous section, we showed how for each choice of time  $t_j$ , the condition  $[L_1, L_j] = 0$  determines a system of partial differential equations for  $u_0, \dots, u_{n-2}$  as functions of  $(x = t_1, t_j)$ . We will see that each such equation can be realized as a reduction of the self-dual Yang–Mills equations. However, it turns out that to deal with the hierarchy as a whole, one requires a corresponding *self-dual Yang–Mills hierarchy*. This is not surprising, in that there are only four independent variables for the self-dual Yang–Mills equations, whereas there are arbitrarily many for the  $n$ -KdV hierarchy!

The formulation of the  $n$ -KdV hierarchy as a reduction of the self-dual Yang–Mills hierarchy is also a key step in giving a twistor description of the  $n$ -KdV equations. For the twistor description, we shall have to fix the number of flows with which we are working at any given time. The definition of the  $n$ -KdV hierarchy makes it natural to discard the trivial flows (those corresponding to the times  $t_n, t_{2n}, \dots$ ) and to group the non-trivial flows into *levels* each with  $n - 1$  flows. The first  $n - 1$  flows will be referred to as the first level, and subsequent levels will be labeled consecutively. A convenient way to truncate the hierarchy is to consider an arbitrary but fixed number  $m$  of levels. These considerations provide some motivation for the following

**Definition.** We work on  $\mathbb{R}^{K(m+1)}$  with coordinates  $x_{Aj} (A = 1, \dots, K, j = 0, \dots, m)$  for a pair of positive integers  $(m, K)$ . Let  $(P_{Ai}, Q_{Ai-1})$  for  $A = 1, \dots, K, i = 1, \dots, m$  be a system of Lie ( $G$ )-valued functions on  $\mathbb{R}^{K(m+1)}$  for some Lie group  $G$ . The

first  $m$  levels of the SDYM hierarchy of order  $K$  and gauge group  $G$  are defined to be the system of non-linear partial differential equations generated by insisting that the differential operators

$$X_{Ai} = \partial_{Ai} - P_{Ai} - \lambda(\partial_{Ai-1} - Q_{Ai-1}) \quad (1 \leq i \leq m) \tag{3.1}$$

all commute. (Here  $\partial_{Ai} = \partial/\partial x_{Ai}$ .)

The SDYM hierarchy is invariant under the action of elements  $g$  of the gauge group of  $G$ -valued functions on  $\mathbb{R}^{K(m+1)}$ :

$$(P_{Ai}, Q_{Aj}) \rightarrow (gP_{Ai}g^{-1} - g\partial_{Ai}g^{-1}, gQ_{Aj}g^{-1} - g\partial_{Aj}g^{-1}).$$

We shall always regard 2 solutions of the SDYM hierarchy as equivalent if they are related by a gauge transformation.

Geometrically we should think of the operators  $X_{Ai}$  as acting on sections of a bundle  $E$  over the domain in  $\mathbb{R}^{K(m+1)}$  on which the solution is defined. The operators can be thought of as being constructed out of a  $G$ -connection on  $E$ ,

$$\nabla = \sum_{A=1, j=0}^{K, m} dx_{Aj} \left( \frac{\partial}{\partial x_{Aj}} - \frac{j}{m} P_{Aj} - \frac{m-j}{m} Q_{Aj} \right) \tag{3.2}$$

and a collection  $\phi_{Aj} = P_{Aj} - Q_{Aj}, j = 1, \dots, m - 1$  of ‘‘Higgs fields,’’ i.e. sections of the adjoint bundle.

Notice that in the limit  $m \rightarrow \infty$ , the connection  $\nabla$  becomes  $\sum_j dx_{Aj}(\partial/\partial x_{Aj} - Q_{Aj})$ . One consequence of the field equations is that this connection is flat.

*Remarks*

1) For  $m = 1$ , there are no Higgs fields and these are simply equations on a connection alone. The SDYM equations arise for  $K = 2$  and  $m = 1$ . The (complexified) hyper-complex equations on a vector bundle with connection arise when  $K$  is even and  $m = 1$ . For general  $K, m = 1$ , the equations are the condition that the connection be compatible with a paraconformal structure [BE], i.e. flat along all the  $K$ -planes spanned by the vector-field parts of the  $X_{A1}$ . See [Wa84] for a treatment of gauge fields in this context.

2) The Bogomolny hierarchy of [MS] up to level  $m$  is, in the present notation, the SDYM hierarchy with  $K = 1$ . Indeed, one can think of the system of order  $K > 1$  as being a system of  $K$  ‘‘interleaved’’ Bogomolny hierarchies. For each fixed value of  $A, (P_{Ai}, Q_{Ai-1}), i = 1, \dots, m$  satisfy the equations of the Bogomolny hierarchy up to level  $m; \{X_{Ai}, i = 1, \dots, m\}$  all commute. Alternatively one can think of this system as a hierarchy over the hypercomplex equations.

3) Given a solution of the SDYM hierarchy, for any pair  $(Ai, Bj)$  we can consider the potentials  $(P_{Ai}, Q_{Ai-1}, P_{Bj}, Q_{Bj-1})$  as functions of the four variables  $(x_{Ai}, x_{Ai-1}, x_{Bj}, x_{Bj-1})$  alone. Then the fact that  $[X_{Ai}, X_{Bj}] = 0$  means that  $(P_{Ai}, Q_{Ai-1}, P_{Bj}, Q_{Bj-1})$  is a solution of the self-dual Yang–Mills equations as a function of  $(x_{Ai}, x_{Ai-1}, x_{Bj}, x_{Bj-1})$ . If  $A = B$  and  $i = j \pm 1$  then there are only three independent coordinates among  $(x_{Ai}, x_{Ai-1}, x_{Bj}, x_{Bj-1})$  and  $(P_{Ai}, Q_{Ai-1}, P_{Bj}, Q_{Bj-1})$  solve the Bogomolny equations (the SDYM equations with one symmetry).

4) The full symmetry group for these equations for  $m > 1$  is  $SL(K, \mathbb{C}) \times SL(2, \mathbb{C})$  together with translations. To see this, we consider  $\lambda$  as a coordinate on the Riemann sphere,  $\mathbb{C}P^1$  and introduce homogeneous coordinates  $\pi_{A'} = (\pi_0', \pi_1')$  so that  $\lambda =$

$\pi_{1'}/\pi_{0'}$ . We can then write the coordinates for space-time as  $x_A^{A'_1 \dots A'_m}$ , symmetric over all the primed indices, related to the coordinates above by

$$x_A^{A'_1 \dots A'_m} \pi_{A'_1} \dots \pi_{A'_m} = (\pi_{0'})^m \sum_{i=0}^m x_{Ai} \lambda^i .$$

The dependent variables  $(P_{Ai}, Q_{Ai-1})$  can be assembled into an object

$$\Phi_{AA'_1 \dots A'_{m-1} A'_m} = \Phi_{A(A'_1 \dots A'_{m-1}) A'_m}$$

(symmetric over the first  $(m - 1)$  primed indices only). In this notation, (3.1) becomes

$$\pi^{A'_m} (\partial_{AA'_1 \dots A'_{m-1} A'_m} - \Phi_{AA'_1 \dots A'_m}) ,$$

where  $\pi^{A'} = (-\pi_{1'}, \pi_{0'})$  has had its index raised by the skew form  $\varepsilon^{A'B'}$  on  $\mathbb{C}^2$ . The  $X_{Ai}$  are retrieved by setting  $i$  of the  $A'_1 \dots A'_{m-1}$  to  $1'$  and the rest to  $0'$ . The connection  $\nabla$  is just given by  $\partial_{AA'_1 \dots A'_m} - \Phi_{A(A'_1 \dots A'_m)}$ , and the Higgs fields by  $\Phi_{AA'_1 \dots A'_{m-1} A'_m} \varepsilon^{A'_{m-1} A'_m}$ .

It is thus clear that the system is invariant under  $SL(2)$  acting on the primed indices, and  $SL(K)$  acting on the unprimed indices. When  $m = 1$  the system is in fact  $SL_{K+2}$  invariant and space-time can be thought of as an open dense subset of  $Gr_2(\mathbb{C}^{K+2})$ .

The symmetry will be broken in the next section when we impose translational invariance on the system.

**3.1. Reduction to  $n$ -KdV.** We shall now show that the SDYM hierarchy that we have just introduced reduces to the  $n$ -KdV hierarchy when:  $G = SL_n(\mathbb{C}); K = n - 1$ ; symmetry is imposed along the  $\partial_{A0}$ ; and certain of the Higgs fields have a specified normal form.

The lift of each translation is generated by a *Lie derivative* operator  $\mathcal{L}_A$  acting on sections of  $E$  satisfying the Leibnitz rule

$$\mathcal{L}_A(fs) = \partial_{A0}fs + f\mathcal{L}_As ,$$

whenever  $f$  is a smooth function and  $s$  is a smooth section of  $E$ . Since they give the infinitesimal action of an abelian group, the operators commute among themselves:

$$[\mathcal{L}_A, \mathcal{L}_B] = 0 .$$

An *invariant* solution of the SDYM hierarchy is by definition one for which each of the operators of (3.1) also commutes with each of the  $\mathcal{L}_A$ :

$$[\mathcal{L}_A, X_{Bi}] = 0 .$$

Two invariant solutions are regarded as equivalent if they are related by a “restricted” or “invariant” gauge transformation  $g$ , i.e. one which satisfies  $\mathcal{L}_A g = 0$  for each  $A$ .

It is often helpful to study a bundle with a symmetry of this type by use of an *invariant gauge* in which each of the  $\mathcal{L}_A$  acts by coordinate differentiation. In an invariant gauge, the data  $(P, Q)$  an invariant solution of the SDYM hierarchy is indeed independent of the coordinates  $x_{A0}$  for  $A = 1, \dots, K$ , and each of the  $Q_{A0}$  becomes a Higgs field, in the sense that it transforms under the adjoint representation

of the restricted gauge group. To be completely explicit, the operators  $X_{A_i}$  defining an invariant solution of the SDYM hierarchy, acting on invariant sections of  $E$ , and written in an invariant gauge, take the following form:

$$X_{A_1} = \partial_{A_1} - P_{A_1} + \lambda Q_{A_0} ; \tag{3.3}$$

$$X_{A_i} = \partial_{A_i} - P_{A_i} - \lambda(\partial_{A_{i-1}} - Q_{A_{i-1}}) \quad (i > 1). \tag{3.4}$$

For brevity, we shall often refer to the system of differential equations generated by the condition that these operators all commute as the *reduced* SDYM hierarchy.

We can now state and prove the reduction theorem.

**Theorem 3.1.** *There is a natural (1:1)-correspondence between the following:*

- (i) *solutions of the  $n$ -KdV hierarchy;*
- (ii) *(gauge-equivalence classes of) solutions  $(P, Q)$  of the  $SL_n$ -SDYM hierarchy of order  $K = n - 1$  which are invariant under the group of translations generated by the vector fields  $\partial/\partial x_{A_0}$  and which satisfy:*

- (a) *the  $Q_{A_0}$  are linearly independent and one of them,  $Q_{K_0}$ , say, is nilpotent of maximal rank in an invariant gauge;*
- (b) *a genericity condition on  $P_{11}$  to be given precisely below.*

*Remarks.* The field equations imply that the  $Q_{A_0}$  commute with each other and are constant in an invariant gauge. Thus  $K = n - 1$  is forced by condition (a).

We prove this theorem under the assumption that  $m = \infty$ . If  $m < \infty$ , the same method of proof shows, going in one direction, that a solution of the first  $m$  levels of the  $n$ -KdV hierarchy yields an invariant solution of the first  $m$  levels of the SDYM hierarchy defined in (ii). In the other direction, we obtain a solution of the first  $m$  levels of the  $n$ -KdV hierarchy from any solution of the first  $(m + 1)$  levels of the SDYM hierarchy satisfying the conditions in (ii). Since we shall subsequently be able to take  $m$  as large as we like, this discrepancy is of minor importance.

*Proof of Theorem 3.1.* We begin by showing how to go from a solution of  $n$ -KdV to a solution of the self-dual Yang–Mills hierarchy that satisfies the conditions of (ii), Theorem 3.1. We suppose given  $L_1$  as in (2.16) and let  $g$  be the formal dressing transformation of Sect. 2.2.2. Thus we may suppose that for each  $r \neq n, 2n, \dots$ ,

$$L_r = \partial_r - (gA^r g^{-1})_+ . \tag{3.5}$$

With the gauge chosen so that the formal expansion of  $g$  in powers of  $\lambda^{-1}$  is  $g = 1 + g_1 \lambda^{-1} + g_2 \lambda^{-2} + \dots$ , (2.33) gives, for  $r < n$ ,

$$L_r = \partial_r - ((1 + g_1 \lambda^{-1})A^r (1 - g_1 \lambda^{-1}))_+ \tag{3.6}$$

$$= \partial_r - f_r - [g_1, e_r] - \lambda e_r , \tag{3.7}$$

where we define  $e_r$  and  $f_r$  from the equation

$$A^r = \lambda e_r + f_r . \tag{3.8}$$

Thus the first  $n - 1$  operators defining the  $n$ -KdV hierarchy are of the form of the operators (3.3) of the reduced SDYM hierarchy, with  $x_{A_1} = t_A$  and

$$X_{A_1} = \partial_{A_1} - P_{A_1} + \lambda Q_{A_0} = \partial_{A_1} - (f_A + [g_1, e_A]) - \lambda e_A .$$

For the higher operators, we compute:

$$L_r - \lambda L_{r-n} = \partial_r - (gA^r g^{-1})_+ - \lambda(\partial_{r-n} - (gA^{r-n} g^{-1})_+) \tag{3.9}$$

$$= \partial_r - \lambda \partial_{r-n} - (\lambda gA^{r-n} g^{-1})_+ + \lambda(gA^{r-n} g^{-1})_+ . \tag{3.10}$$

However, if  $f(\lambda) = \sum_i f_i \lambda^i$ , then  $\lambda(f(\lambda))_+ - (\lambda f(\lambda))_+ = -f_{-1}$  so we see that

$$L_r - \lambda L_{r-n} = \partial_r - \phi_r - \lambda \partial_{r-n} , \tag{3.11}$$

where  $\phi_r$  is the coefficient of  $\lambda^{-1}$  in the power-series expansion of  $gA^{r-n} g^{-1}$ . We note that  $\phi_r$  is trace-free because this is true of  $gA^r g^{-1}$ .

Equation (3.11) matches up with the higher operators (3.4),  $i > 1$ , of the reduced SDYM hierarchy with

$$x_{Aj} = t_{jn+A}, \quad P_{Aj} = \phi_{nj+A} \text{ and } Q_{Aj} = 0 \quad (j > 0)$$

so that

$$X_{Aj} = L_{nj+A} - \lambda L_{nj+A-1} = \partial_{Aj} - \phi_{jn+A} - \lambda \partial_{Aj-1} .$$

Notice that there is no analogue in the SDYM hierarchy of the trivial flows  $t_n, t_{2n}, \dots$

Note finally that the fact that the  $L_j$  all commute implies that all the  $X_{Aj}$  commute. In other words, these operators give a solution of the reduced SDYM hierarchy as in (ii).

To go in the other direction, we begin by trivializing the bundle  $E$  in such a way that  $Q_{Aj} = 0$  for all  $A$  and for all  $j > 0$ . This is possible because the  $\lambda^2$  part of  $[X_{Ai}, X_{Bj}] = 0 (i, j > 0)$  implies that the connection  $\sum_{Ai} dx_{Ai} (\partial/\partial x_{Ai} - Q_{Ai})$  is flat.

We may also choose this trivialization to be invariant, so that for each  $A$ ,  $\mathcal{L}_A$  acts as  $\partial/\partial x_{A0}$ . We cannot, however, assume this trivialization to be chosen so that  $Q_{A0} = 0$ . This is precisely because  $Q_{A0}$  transforms under the adjoint representation of the restricted gauge group (cf. the discussion immediately before the statement of Theorem 3.1). Instead, the  $\lambda^2$  part of the other commutator equations imply that in this gauge, each  $Q_{A0}$  is constant and that these matrices commute among themselves.

Assumption (a) of the theorem now says that there is a constant gauge transformation of our invariant gauge after which we have

$$Q_{A0} = -e_A \quad (0 \leqq A \leqq K - 1) . \tag{3.12}$$

(Recall the definition of  $e_r$  in (3.8).)

*Remark.* We could get away with less than this. If we suppose that the  $Q_{A0}$  generate an algebra conjugate to that generated by the  $e_r$ , then (3.12) will hold after a linear change of coordinates. This would follow, for example, from the condition that they be linearly independent with one of them conjugate to  $e_{n-1}$ . This is because the centralizer of  $e_{n-1}$  is just the algebra of the  $e_r$ 's. The condition that they should be linearly independent forces  $K = n - 1$ .

With these steps complete, our differential operators take the following form:

$$X_{A1} = \partial_{A1} - P_{A1} - \lambda e_A \quad (1 \leqq A \leqq K = n - 1) \tag{3.13}$$

and

$$X_{Aj} = \partial_{Aj} - P_{Aj} - \lambda \partial_{Aj-1} \quad (j > 0, 1 \leqq A \leqq K = n - 1) ; \tag{3.14}$$

all the information of the  $\lambda^2$ -part of the field equations has now been taken into account.

Our next task is to show that  $X_{11}$  is of the form of  $L_1$  in (2.16); that is, that  $P_{11}$  is forced by the equations of the SDYM hierarchy to be the sum of a lower triangular matrix and  $f_1$  (whose definition appears in (3.8)). But the  $\lambda^1$ -part of  $[X_{11}, X_{A1}] = 0$  yields

$$[P_{11}, e_A] = [P_{A1}, e_1];$$

in other words,  $[P_{11}, e_A]$  is “L-shaped”: its only non-zero entries are in the first column and the last row. It is a direct calculation that this forces

$$P_{11} = \begin{bmatrix} q_1 & p_1 & 0 & 0 & \dots & 0 \\ q_2 & p_2 & p_1 & 0 & \dots & 0 \\ q_3 & p_3 & p_2 & p_1 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ q_{n-1} & p_{n-1} & p_{n-2} & & \dots & p_1 \\ q_n & q_{n+1} & q_{n+2} & & \dots & q_{2n-1} \end{bmatrix} \tag{3.15}$$

for certain  $p$ 's and  $q$ 's. Now the  $\lambda$ -term of the other commutators involving  $X_{11}$  yield that each derivative of  $P_{11}$  is also L-shaped; so the  $p$ 's are forced to be constants.

This is the point at which the assumption (b) of genericity enters. We assume that  $p_1 \neq 0$ . Then by multiplying  $x_{11}$  by a constant, we may suppose that  $p_1 = 1$  so that  $X_{11} = L_1$  is a representative of a point in the orbit space  $M$ . By the uniqueness of the  $n$ -KdV hierarchy (Sect. 2.2.2), it is now sufficient to prove that the field equations  $[X_{Ai}, X_{Bj}] = 0$  imply evolution equations of the form  $[L_1, L_r] = 0$ , where  $L_r$  is of the form (2.20). But that is trivial; put

$$L_r = X_{Aj} + \lambda X_{Aj-1} + \lambda^2 X_{Aj-2} + \dots,$$

where again  $r = jn + A$ .

*Remark.* This proof is a higher-dimensional version of the corresponding result in the Appendix of [MS].

As a corollary we obtain:

**Theorem 3.2.** *For any positive integers  $r, s$  not divisible by  $n$ , the equation  $[\partial_r - P_r, \partial_s - P_s] = 0$  (cf. (2.5)), considered as an evolution equation in the 2 variables  $t_r$  and  $t_s$ , is a reduction of the SDYM equations.*

*Proof.* Use Theorem 3.1 to embed the given equation (which is equivalent to (2.35)) in the SDYM hierarchy. Now apply Remark 3 after the definition of the SDYM hierarchy.

It is clear that other choices of gauge group (together with appropriate modifications of conditions (a) and (b) of the Theorem) will give rise to the Drinfeld–Sokolov hierarchies described in Sect. 2.2.3.

It is also clear that for a fixed choice of gauge group, one can make different ansätze for the algebra generated by the  $Q_{A0}$  and that these will yield other systems of differential equations. For example, if  $G = SL_n$  and the  $Q_{A0}$  are chosen to be generic and diagonal, one will obtain the standard Zakharov–Shabat framework

as described in Sect.1 of [DS]. For  $n = 2$  this gives the non-linear Schrödinger equation. More elaborate variations are presumably those of [GHM] and [KW].

3.2. *Summary; Dressing the SDYM Hierarchy.* An enlightening and useful re-interpretation of some of the calculations of the present section is as follows. Consider the “vacuum” reduced SDYM hierarchy in invariant gauge:

$$U_{A1} = \partial_{A1} - \lambda \partial_{A0} - A^A, \tag{3.16}$$

$$U_{Ai} = \partial_{Ai} - \lambda \partial_{Ai-1}, \tag{3.17}$$

$$\text{action given by } \partial_{A0}. \tag{3.18}$$

Then if  $g$  is a dressing transformation for  $n$ -KdV, with  $g(x, \infty) = 1$  (as usual), the operators

$$X_{Ai} = g \cdot U_{Ai} \cdot g^{-1} \tag{3.19}$$

give the reduced SDYM hierarchy corresponding to the  $n$ -KdV hierarchy determined by  $g$ . If we write

$$X_{Ai} = \partial_{Ai} - P_{Ai} - \lambda(\partial_{Ai-1} - Q_{Ai-1}), \tag{3.20}$$

then we obtain explicit formulae

$$P_{A1} - \lambda Q_{A0} = (gA^A g^{-1})_+ \text{ so } Q_{A0} = -e_A \tag{3.21}$$

and, for  $i > 1$ ,

$$P_{Ai} - \lambda Q_{Ai-1} = (g\lambda^{i-1}A^A g^{-1})_+ - (g\lambda^{i-2}A^A g^{-1})_+ \text{ so } Q_{Ai-1} = 0. \tag{3.22}$$

Conversely if  $g$  satisfies (3.19, 3.21, 3.22) then it is a dressing transformation for  $n$ -KdV.

Notice that  $g$  satisfies the differential equations

$$X_{A1}(g) = -gA^A, \tag{3.23}$$

$$X_{Ai}(g) = 0, i > 0. \tag{3.24}$$

We can avoid the distinction between  $i = 1$  and  $i > 1$  by changing to a certain *non-invariant* gauge. Put

$$g' = g \exp(-\sum x_{A0} \lambda^{-1} A^A). \tag{3.25}$$

Then from (3.23) and (3.24),

$$X_{Ai}(g') = 0 \text{ for } i = 1, 2, \dots. \tag{3.26}$$

We shall refer to  $g'$  as the *modified* dressing transformation. Notice that like  $g, g' \in SL_n(\mathbb{C}[\lambda^{-1}])$ , but

$$g'(x_{Ai}, \infty) = \exp(-\sum x_{A0} e_A). \tag{3.27}$$

The modified dressing transformation is so called because the operators  $X_{Ai}$  of (3.19) are obtained by dressing the “vacuum” reduced SDYM hierarchy

$$V_{Ai} = \partial_{Ai} - \lambda \partial_{Ai-1} \quad i = 1, 2, \dots, \tag{3.28}$$

with

$$\text{action given by } \partial_{A0} - \lambda^{-1} A^A, \tag{3.29}$$

thus:

$$X_{Ai} = g' \cdot V_{Ai} \cdot g'^{-1}, \partial_{A0} = g' \cdot (\partial_{A0} - \lambda^{-1} A^A) \cdot g'^{-1}. \tag{3.30}$$

Various of these formulae will be useful later on. The following observation will prove to be of surprising importance in Sect. 5. Suppose we have a matrix  $g'$ , such that the  $X_{Ai}$  defined from (3.30) comprise the linear system of the reduced SDYM hierarchy in invariant gauge; that is, such that the only powers of  $\lambda$  which appear in  $X_{Ai}$  are the zeroth and first. Then this reduced SDYM hierarchy *corresponds to the solution of  $n$ -KdV with modified dressing transformation*  $= g'$ . This observation is immediate from the previous discussion.

#### 4. The Twistor Correspondence

In this section, we first describe the twistor correspondence between holomorphic solutions of the first  $m$  levels of the self-dual Yang–Mills hierarchy of order  $K$  and holomorphic  $G$ -bundles over an open subset of an appropriate twistor space. This twistor space will be a rank- $K$  vector bundle over the Riemann sphere.

We then describe the specialization to holomorphic solutions of  $n$ -KdV.

*4.1. The Underlying Geometry.* The twistor space,  $Z$ , is the rank  $K$  vector bundle over the Riemann sphere  $\mathbb{C}P^1$  obtained by tensoring the trivial bundle with  $\mathcal{O}(m)$ .

$$Z = \mathcal{O}(m) \oplus \dots \oplus \mathcal{O}(m) \quad (K \text{ terms in the sum}). \tag{4.1}$$

The associated “space-time” (i.e., the space of independent variables of the solution of the SDYM hierarchy) is defined to be the space of sections of this bundle which, as a manifold, is

$$X = \Gamma(\mathbb{C}P^1, Z) \simeq C^{(m+1)K}.$$

We will denote the holomorphically embedded  $\mathbb{C}P^1$  in  $Z$  corresponding to the point  $x \in X$  by  $L_x$ .

Introduce homogeneous coordinates  $(u, v)$  on  $\mathbb{C}P^1$  and homogeneous coordinates  $\mu_A, A = 1, \dots, K$  on the fibres of  $Z$  so that  $(u, v, \mu_A) \sim (\alpha u, \alpha v, \alpha^m \mu_A), \alpha \in \mathbb{C}^*$ . The general section of  $Z$  is then given by

$$\mu_A = \sum_{j=0}^m x_{Aj} u^j v^{m-j}.$$

A point  $z \in Z$  is represented by the subset of  $X$  consisting of those  $x \in X$  such that  $L_x$  contains  $z$ . This is the affine codimension- $K$  hyperplane  $\Sigma_z$  in  $X$  given by holding  $(\mu_A, u, v)$  fixed in the above equation and letting the  $x$ -variables vary.

The correspondence between  $X$  and  $Z$  is conveniently studied in terms of the “correspondence space”

$$Y = \{(x, z) \in X \times Z : z \in L_x\}. \tag{4.2}$$

The restrictions to  $Y$  of the projections on the two factors of  $X \times Z$  define a double fibration

$$\begin{array}{ccc} & Y & \\ \phi_1 \swarrow & & \searrow \phi_2 \\ X & & Z \end{array} \tag{4.3}$$

where for each  $x$  in  $X$ ,  $\wp_1^{-1}(x)$  is a copy of  $L_x = \mathbb{C}P^1$  while for each  $z \in Z$ ,  $\wp_2^{-1}(z)$  is  $\Sigma_z$ , the space of  $x$  such that  $L_x$  passes through  $z$ . To understand these maps more explicitly, it will be convenient to introduce coordinates  $x_{Aj}$  on  $X$  so that coordinates on  $Y$  are

$$(x_{Aj}, u : v) \tag{4.4}$$

with

$$\wp_1(x_{Aj}, u : v) = x_{Aj} \tag{4.5}$$

and

$$\wp_2(x_{Aj}, u : v) = (\mu_A : u : v), \quad \text{where} \quad \mu_A(x, u : v) = \sum_{j=0}^m x_{Aj} u^j v^{m-j}. \tag{4.6}$$

**4.2. The Ward Transform for the Self-Dual Yang–Mills Hierarchy.** In this section we prove:

**Theorem 4.1.** *There is a natural one-one correspondence (the Ward transform) between (biholomorphic equivalence classes of) holomorphic  $G$ -bundles  $E$  such that  $E|_{L_x}$  is trivial for all  $x$  and (gauge equivalence classes of) holomorphic solutions of the first  $m$  levels of the  $G$ -SDYM hierarchy of order  $K$ .*

*Remark.* The apparently strong condition that  $E$  be trivial for each  $L_x$  actually holds generically for an open dense subset  $U$  of  $X$  provided that  $E$  is topologically trivial – this is a consequence of Birkhoff factorization, [PS, Ch. 8]. For the generic bundle, one will obtain a solution of the SDYM hierarchy on  $U$  with rational singularities on  $X - U$ .

*Proof.* The first step in the Ward transform is to pull  $E$  back to  $Y$  and to define a bundle  $E'$  on  $X$  by  $E'|_x = \Gamma(\wp_1^{-1}(x), \wp_2^*E) = \Gamma(L_x, E)$ . For brevity, we write  $\tilde{E}$  for  $\wp_2^*(E)$ .

From its definition,  $\tilde{E}$  possesses a canonical flat holomorphic relative connection

$$\tilde{\nabla} : \tilde{E} \rightarrow \tilde{E} \otimes \Omega_{\wp_2}^1. \tag{4.7}$$

Here  $\Omega_{\wp_2}^1$  denotes the bundle of  $\wp_2$ -relative holomorphic one-forms (i.e. the bundle dual to the bundle of holomorphic vector fields which are tangent to the fibres of  $\wp_2$ ). In a frame for  $\tilde{E}$  that is pulled back from one for  $E$  on  $Z$ ,  $\tilde{\nabla}$  is just the restriction  $d_{\wp_2}$  of the standard exterior derivative on  $Y$  to  $\Omega_{\wp_2}^1$ , (so that only the contraction of vector fields tangent to the fibres of  $\wp_2$  into  $d_{\wp_2}f$  is defined).

The second step of the Ward transform is to show how the relative connection  $\tilde{\nabla}$  induces data on  $E'$ . To do this, one notes that  $\tilde{E}$  is holomorphically trivial. Thus there is a gauge transformation that identifies  $\tilde{E}$  with  $Y \times \mathbb{C}^n$ . In such a gauge, our relative connection (4.7) takes the form

$$\tilde{\nabla} = d_{\wp_2} - \tilde{a}, \tag{4.8}$$

where  $\tilde{a}$  is an  $n \times n$  matrix of holomorphic  $\wp_2$ -relative one-forms global on  $Y$ . It is  $\tilde{a}$  that defines data on  $E'$ ; and the field equations come from the relative flatness of  $\tilde{\nabla}$ :

$$\tilde{\nabla}^2 = 0 \text{ (as a section of } \text{End}(\tilde{E}) \otimes \Omega_{\wp_2}^2 \text{)}. \tag{4.9}$$

To analyze  $\tilde{a}$ , note that in the coordinates  $(x_{Aj}, u : v)$  of (4.4),  $d_{\wp_2}$  may be identified with the operator

$$F(x, u : v) \mapsto \{(v\partial_{Aj} - u\partial_{Aj-1})F(x, u : v)\}_{\{1 \leq A \leq K, 1 \leq j \leq m\}}. \tag{4.10}$$

Accordingly, each entry in  $\tilde{a}$  can be identified with a collection of  $mK$  holomorphic Lie(G)-valued functions which are homogeneous of degree 1 in  $(u : v)$ . Thus each entry of  $\tilde{a}$  is actually *linear* in  $(u : v)$  because it is globally defined over the Riemann sphere of  $(u : v)$ . When this is taken into account, (4.8) may be identified with the collection of operators

$$\{X_{Aj}\} = \{v(\partial_{Aj} - P_{Aj}) - u(\partial_{Aj-1} - Q_{Aj-1})\}_{\{1 \leq A \leq K, 1 \leq j \leq m\}}, \tag{4.11}$$

where the  $P$ 's and  $Q$ 's are matrix functions on  $X$ . Since the condition (4.9) of relative flatness is gauge invariant,  $[X_{Ai}, X_{Bj}] = 0$  for all values of  $(u : v)$ ; these are the field equations. If we use the affine coordinate  $\lambda = u/v$ , we see that this collection of operators is precisely the linear system of the SDYM hierarchy.

Note that the globality that is used in this construction is in the direction of the fibres of  $Y$  over  $X$ . The Ward transform can be restricted to give a (1:1)-correspondence between holomorphic solutions of the SDYM hierarchy on any convex Stein open set  $U \subset X$  and holomorphic bundles on  $\wp_2(\wp_1^{-1}(U))$  in  $Z$ .

**4.3. Specialization to  $n$ -KdV.** To obtain the twistor description of  $n$ -KdV, we must encode the  $K$  symmetries and the associated Higgs fields ( $Q_{A0}$  etc.; cf. Theorem 3.1) into the structure of the Ward bundle on  $Z$ . For consistency with the notation of the present section, it is necessary to change slightly the notation for the symmetries adopted in Sect. 3. Let  $E$  be our holomorphic bundle over twistor space, let  $E' \rightarrow X$  be the Ward transform of  $E$  and let  $\tilde{E} = \wp_2^*(E) = \wp_1^*(E')$  be the pull-back of  $E$  over the correspondence space  $Y$ . We assume that the solution of the SDYM hierarchy on  $E'$  that arises from the Ward transform (Theorem 4.1) is invariant in the sense of Theorem 3.1, the action being generated by Lie derivative operators  $\mathcal{L}'_A$ .

There is a natural lift  $\tilde{\mathcal{L}}_A$  to  $\Gamma(\tilde{E})$  of the action of  $\mathcal{L}'_A$  on  $\Gamma(E')$ . This is because translations preserve the canonical product structure  $Y = X \times \mathbb{C}P^1$  and  $\tilde{E} = \wp_1^*(E')$ . To be quite explicit, if  $f$  is a smooth function of  $(u : v)$  and  $s$  is a smooth local section of  $E'$ , the formula

$$\tilde{\mathcal{L}}_A(f \otimes s) = f \otimes \mathcal{L}'_A s$$

serves to define  $\tilde{\mathcal{L}}_A$ .

To see that  $\tilde{\mathcal{L}}_A$  induces an action on  $\Gamma(E)$ , it is sufficient to check that if  $s$  is a local section of  $\tilde{E}$  such that  $\tilde{\nabla}s = 0$ , then also  $\tilde{\nabla}\tilde{\mathcal{L}}_A s = 0$ , since then  $\tilde{\mathcal{L}}_A s$  will define a section of  $E$ . But this is an immediate consequence of the assumed invariance of our solution of the SDYM hierarchy, which, on  $Y$ , can be expressed as

$$[\tilde{\mathcal{L}}_A, \tilde{\nabla}] = 0.$$

If  $\tilde{\nabla}s = 0$ , this gives

$$\tilde{\nabla}\tilde{\mathcal{L}}_A s = \tilde{\mathcal{L}}_A \tilde{\nabla}s - [\tilde{\mathcal{L}}_A, \tilde{\nabla}]s = 0.$$

Let us denote by  $\mathcal{L}_A$  the induced action on sections of  $E$ .

To bring the operator  $\mathcal{L}_A$  into sharper focus, we may start with the observation that this procedure, applied to the action of  $\partial_{A0}$  on functions on  $Y$  pulled back from  $X$ , yields the action of the vector field

$$\xi_A = v^m \partial / \partial \mu_A$$

on functions on  $Z$ . This corresponds to the fact that the translation  $X_{A0} \rightarrow x_{A0} + a_{A0}$  is induced by the biholomorphic map  $\mu_A \rightarrow \mu_A + v^m a_A$  of  $Z$ . Accordingly, we have the Leibnitz rule

$$\mathcal{L}_A(fs) = \xi_A(f)s + f \mathcal{L}_A s \tag{4.12}$$

(when  $f$  is a local function and  $s$  is a local section of  $E$ ). It follows from (4.12) that at the zeros of  $\xi_A$ , the  $\mathcal{L}_A$  act linearly on the fibre  $E_z$ . In fact, since the zeros of  $\xi_A$  at  $\lambda = \infty$  have order  $m$ ,  $\mathcal{L}_A$  acts linearly on  $E_z$  to order  $m$  in  $\lambda^{-1}$ .

To compute this linear action, recall the formulae of Sect. 3.2:

$$\begin{aligned} X_{A1} &= \partial_{A1} - (g\lambda^A g^{-1})_+ - \lambda \partial_{A0}, \\ X_{Ai} &= \partial_{Ai} - (g\lambda^{i-1} A^A g^{-1})_+ \\ &\quad - \lambda(\partial_{Ai-1} - (g\lambda^{i-2} A^A g^{-1})_+) \quad (1 < i \leq m), \\ \tilde{\mathcal{L}}_A &= \partial_{A0}. \end{aligned}$$

By multiplying  $X_{Ai}$  by  $\lambda^{-i}$  and summing, we find

$$\tilde{\mathcal{L}}_A = -\sum_{i=1}^m \lambda^{-i} X_{Ai} + \lambda^{-m} (\partial_{Am} - (g\lambda^{m-1} A^A g^{-1})_+).$$

But, as we have already observed, a local section  $s$  of  $E$  corresponds to a local section of  $\tilde{E}$  that is annihilated by each of the  $X_{Ai}$ , so the previous formula yields:

$$\mathcal{L}_A(s) = \xi_A(s) - \lambda^{-m} (g\lambda^{m-1} A^A g^{-1})_+ s \tag{4.13}$$

$$= \xi_A(s) - g(\lambda^{-1} A^A) g^{-1} s + O(\lambda^{-m-1}). \tag{4.14}$$

Thus at the zeros ( $\lambda^{-1} = 0$ ) of  $\xi_A$ , the action of  $\mathcal{L}_A$  on the fibre  $E_z$  is conjugate to  $\lambda^{-1} A^A + O(\lambda^{-m})$ .

This argument is clearly reversible, so that we have:

**Theorem 4.2.** *The Ward transform defines a one-one correspondence between solutions of the  $n$ -KdV hierarchy on  $\mathbb{C}^{(n-1)^m}$  and (biholomorphic equivalence classes of) holomorphic  $SL_n(\mathbb{C})$ -bundles  $E$  over  $Z$  that are trivial over each  $L_x$  such that:*

- (i) *the  $\mathbb{C}^{n-1}$ -action on  $Z$  generated by the  $\xi_A$  lifts to an action on  $E$ , generated by the operators  $\mathcal{L}_A$ ;*
- (ii) *at a fixed point  $z$  of the action of  $\xi_A$ , the action of the  $\mathcal{L}_A$  on  $E_z$  generates an algebra conjugate to that generated by the  $A^A$ .*

### 5. The Segal–Wilson Ansatz

In this section, we describe an ansatz for the construction of holomorphic vector bundles over twistor space which satisfy the hypotheses of Theorem 4.2. The input is a holomorphic map  $\gamma(\lambda)$ ,

$$\gamma : \bar{D}_\infty \rightarrow \text{GL}_n(\mathbb{C}), \gamma(\infty) = 1, \tag{5.1}$$

where  $D_\infty$  is some open neighbourhood of  $\lambda = \infty$  in  $\mathbb{CP}_1$ . The output of the ansatz is a holomorphic vector bundle  $E_\gamma$  over  $Z$  which satisfies the conditions of Theorem 4.2, and so a solution of  $n$ -KdV. We refer to the construction as the Segal–Wilson ansatz because we shall see at the end of this section that the solution of  $n$ -KdV that corresponds to  $E_\gamma$  is identical to that associated to  $\gamma$  by the construction in [SW, Sect. 5]. The rest of this section is devoted to setting up the ansatz and to performing the Ward transform explicitly in the case that  $\gamma$  extends to a rational map from  $\mathbb{CP}_1$  to  $\text{GL}_n(\mathbb{C})$ .

*5.1. Description of the Ansatz.* Cover twistor space  $Z$  by open sets  $U_0 = \{\lambda \neq \infty\}$  and  $U_\infty =$  the part of  $Z$  that lies above  $D_\infty$ , where  $D_\infty$  is a small disc in  $\mathbb{CP}_1$ , centred on  $\lambda = \infty$ . (Recall from Sect. 4 that  $Z$  has the structure of a bundle over  $\mathbb{CP}_1$ .)

Let  $F_0$  and  $F_\infty$  be holomorphic frames over  $U_0$  and  $U_\infty$  respectively and define  $E_\gamma$  via the patching relation

$$F_\infty = G \cdot \gamma F_0, \tag{5.2}$$

where  $\gamma$  is holomorphic as in (5.1) and  $G$  is the matrix function

$$G(\mu_A, \lambda) = \exp(\lambda^{-1} \sum \mu_A \lambda^A). \tag{5.3}$$

The  $\mathbb{C}^{n-1}$ -action on  $Z$  is lifted to  $E_\gamma$  by decreeing that  $F_0$  be an invariant frame, so that  $\mathcal{L}_A$  acts as  $\xi_A = \partial/\partial\mu_A$  relative to the frame  $F_0$  and

$$\mathcal{L}_A \text{ acts by } \xi_A - \lambda^{-1} A^A \text{ relative to } F_\infty. \tag{5.4}$$

(The formula for the action in  $F_\infty$  follows from (5.2).)

Bundles of the form  $E_\gamma$  are characterized by the existence of a frame  $F_\infty$  with this property:

**Proposition 5.1.** *A bundle  $E$  which satisfies the conditions of Theorem 4.2 is an  $E_\gamma$  iff there exists a frame  $F_\infty$  for  $E$  which is holomorphic on some neighbourhood of  $\lambda = \infty$  and satisfies (5.4).*

*Proof.* Given the bundle  $E$  and the frame  $F_\infty$ , we construct an appropriate  $F_0$  as follows. Let  $L$  be the line in  $Z$  which corresponds to  $x_{Ai} = 0$ . Since  $E|L$  is trivial, choose a frame  $F_L$  for  $E|L$  and extend to a frame  $F_0$  over  $U_0$  by solving

$$\mathcal{L}_A F_0 = 0, \quad F_0|U_0 \cap L = F_L.$$

Then on  $U_0 \cap U_\infty$  we have 2 frames  $F_0$  and  $G(\mu, \lambda)F_\infty$  that are invariant: relative to each,  $\mathcal{L}_A$  acts as  $\xi_A$ . Thus these frames are related by a function of  $\lambda$  only,  $\gamma(\lambda)$ , say. This yields Eq. (5.2), so  $E = E_\gamma$ . With this construction,  $\gamma(\lambda)$  is just the value of the patching matrix between  $F_\infty$  and  $F_L$  on  $U_\infty \cap L$ .

We remark that (5.2) does not quite define an  $\text{SL}_n(\mathbb{C})$ -bundle. However, since  $\gamma(\infty) = 1$ , a simple rescaling of  $\gamma$  could be used to make the transition function have

determinant = 1. It turns out to be more convenient to work with the normalization  $\gamma(\infty) = 1$  than with the condition  $\det(G \cdot \gamma) = 1$ .

5.2. *Implementation of the Ward Transform.* As in Sect. 4.2, start by pulling  $E_\gamma$  back to  $Y$  to get  $\tilde{E}_\gamma$ ; this has the explicit description

$$\tilde{F}_\infty = G(\mathbf{x}, \lambda) \cdot \gamma(\lambda) \tilde{F}_0, \tag{5.5}$$

where  $\tilde{F}_0$  and  $\tilde{F}_\infty$  are the pull-backs of  $F_0$  and  $F_\infty$  and  $G(\mathbf{x}, \lambda)$  is the pull-back of  $G(\mu, \lambda)$ :

$$G(\mathbf{x}, \lambda) = \exp\left(\sum x_{Ai} \lambda^{i-1} A^i\right) = \exp\left(\sum x_{A0} \lambda^{-1} A^i\right) G_0. \tag{5.6}$$

In this frame, the components of the relative connection are just  $V_{Ai} = \partial_{Ai} - \lambda \partial_{Ai-1}$ , while the action is given by  $\tilde{\mathcal{L}}_A = \partial_{A0} - \lambda^{-1} A^A$ .

The second step is to transform to a global holomorphic frame. This is achieved by Birkhoff factorization of  $G \cdot \gamma$ :

$$G(\mathbf{x}, \lambda) \gamma(\lambda) = g'_\infty(\mathbf{x}, \lambda)^{-1} g_0(\mathbf{x}, \lambda), \tag{5.7}$$

where  $g'_\infty : X \times D_\infty \rightarrow \text{GL}_n(\mathbb{C})$  and  $g_0 : X \times \{\lambda \neq \infty\} \rightarrow \text{GL}_n(\mathbb{C})$  are holomorphic maps. We can and shall assume further that

$$g'_\infty(\mathbf{x}, \lambda) = g_\infty(\mathbf{x}, \lambda) \exp\left(-\sum x_{A0} \lambda^{-1} A^A\right), \tag{5.8}$$

where

$$g_\infty \text{ is independent of } x_{A0}, g_\infty(\mathbf{x}, \infty) = 1. \tag{5.9}$$

Then a global holomorphic frame  $F$  on  $Y$  is given by

$$F = g'_\infty \cdot \tilde{F}_\infty = g_0 \cdot \tilde{F}_0 \tag{5.10}$$

and relative to this frame, the components of the relative connection are

$$X_{Ai} = g'_\infty \cdot V_{Ai} \cdot g'^{-1}_\infty = g_0 \cdot V_{Ai} \cdot g_0^{-1}. \tag{5.11}$$

Moreover, our choice of dependence of  $g'_\infty$  upon  $x_{A0}$  means that  $F$  is an *invariant* frame.

The Ward transform guarantees that the  $X_{Ai}$  generate a solution of the reduced SDYM hierarchy; comparison with the formulae in Sect. 3.2 and use of the observation made at the end of that section shows that  $g'_\infty$  is the modified dressing transformation for the corresponding  $n$ -KdV solution. Thus we have here an example of a solution of  $n$ -KdV for which the dressing transformation converges to a holomorphic function in some neighbourhood of  $\lambda = \infty$ .

As in Proposition 5.1, the converse is also true:

**Proposition 5.2.** *A (holomorphic) solution of  $n$ -KdV has a convergent dressing transformation  $g$  iff it is the Ward transform of an  $E_\gamma$ .*

*Proof.* This follows from the formulae in Sect. 3.2 and the discussion above.

5.3. *The Trivial Solution.* The simplest example of the Segal–Wilson ansatz arises by taking  $\gamma(\lambda) = 1$  and corresponds to the trivial solution of  $n$ -KdV. Note that the bundle  $E_1$  is a non-trivial holomorphic vector bundle over  $Z$ . In this case, the Birkhoff factorization of  $G_0 \cdot \gamma$  is given simply by  $g_0 = G_0$ ,  $g_\infty = 1$ , so (5.11) yields the vacuum reduced SDYM hierarchy in invariant gauge:

$$X_{A1} = \partial_{A1} - \lambda \partial_{A0} - A^A \text{ and } X_{Ai} = \partial_{Ai} - \lambda \partial_{Ai-1} \text{ for } i \geq 2$$

(cf. (3.16) and (3.17)). This indeed corresponds to the trivial solution of  $n$ -KdV.

**5.4. Rational Soliton Solutions.** Let us consider the Segal–Wilson ansatz with  $\gamma$  a rational matrix function of  $\lambda$  satisfying  $\gamma(\infty) = 1$ . For such  $\gamma$  it is possible to perform the Birkhoff factorization (5.7) explicitly. We deal first with the generic case. Thus we assume that

$$\det \gamma(\lambda) = \prod_1^k \frac{\lambda - \lambda_r}{\lambda - \tilde{\lambda}_r} \tag{5.12}$$

for  $2k$  points  $\lambda_1, \dots, \lambda_k, \tilde{\lambda}_1, \dots, \tilde{\lambda}_k$  such that

1. the  $\lambda_r$  and the  $\tilde{\lambda}_s$  are distinct points of  $\mathbb{C}$ ;
2.  $\text{res}_{\tilde{\lambda}_r}(\gamma)$  is of rank 1 with image spanned by  $\tilde{v}_r$ ;
3.  $\text{Im}(\gamma(\lambda_r))$  is of codimension 1 and is annihilated by the functional  $\alpha_r$ ;
4. the  $k \times k$  matrix  $M$  with entries  $\alpha_s(\tilde{v}_r)/(\tilde{\lambda}_r - \lambda_s)$  is invertible.

If condition 4 is satisfied, then the data consisting of the  $k$  quadruples  $\{\lambda_s, [\alpha_s], \tilde{\lambda}_s, [\tilde{v}_s]\} \in \mathbb{C} \times \mathbb{C}\mathbb{P}^{n-1} \times \mathbb{C} \times \mathbb{C}\mathbb{P}^{n-1}$  uniquely determines a loop satisfying conditions 1–3 and Eq. (5.12). Indeed we claim that  $\gamma(\lambda)$  has a simple partial-fraction expansion

$$\gamma(\lambda) = 1 + \sum_1^k \frac{\tilde{v}_s \otimes \tilde{\beta}_s}{\lambda - \tilde{\lambda}_s}, \tag{5.13}$$

where the  $\tilde{\beta}_r$  are linear functionals to be determined. (Throughout the discussion, we use  $u, v$ , etc. for vectors in  $\mathbb{C}^n$  and  $\alpha, \beta$ , etc. for functionals on  $\mathbb{C}^n$ .) The RHS of (5.13) has the correct residue at each of the poles  $\tilde{\lambda}_r$ . Substituting condition 3 into (5.13), we obtain

$$0 = \alpha_r(\gamma(\lambda_r)u) = \alpha_r(u) + \sum \frac{\alpha(\tilde{v}_s)\tilde{\beta}_s(u)}{\lambda_r - \tilde{\lambda}_s} \tag{5.14}$$

for any  $u$  in  $\mathbb{C}^n$ ; and this can be solved uniquely for  $\tilde{\beta}_s$  given condition 4.

We now turn to the explicit Birkhoff factorization of  $G_0(\mathbf{x}, \lambda)\gamma(\lambda)$ . We assume that  $g_\infty$  in (5.8) has the form

$$g_\infty(\mathbf{x}, \lambda) = 1 + \sum \frac{v_s \otimes \beta_s}{\lambda - \lambda_s}, \tag{5.15}$$

where the  $\mathbf{x}$ -dependent vectors  $v_s$  and functionals  $\beta_s$  are determined by the condition that  $g_\infty \cdot G_0 \cdot \gamma$  be holomorphic in the finite complex  $\lambda$ -plane. Now the residue of  $g_\infty$  at  $\lambda_r$  is of rank 1; this will be cancelled by  $G_0(\mathbf{x}, \lambda_r)\gamma(\lambda_r)$  precisely if  $G_0(\mathbf{x}, \lambda_r)\beta_r$  annihilates  $\text{Im}(\gamma(\lambda_r))$ , i.e.

$$\beta_r = G_0^{-1}(\mathbf{x}, \lambda_r)\alpha_r \tag{5.16}$$

(cf. condition 3 above). On the other hand, at the point  $\tilde{\lambda}_r$ , the residue  $\tilde{v}_r \otimes \tilde{\alpha}_r$  of  $\gamma$  will be cancelled iff  $g_\infty(\mathbf{x}, \tilde{\lambda}_r)$  annihilates  $G_0(\mathbf{x}, \tilde{\lambda}_r)\tilde{v}_r$ . This condition is equivalent to the equation

$$-G_0(\mathbf{x}, \tilde{\lambda}_r)\tilde{v}_r = \sum_s M_{rs}v_s, \tag{5.17}$$

where

$$M_{rs}(\mathbf{x}) = \frac{\alpha_s(G_0^{-1}(\mathbf{x}, \lambda_s)G_0(\mathbf{x}, \tilde{\lambda}_r)\tilde{v}_r)}{\tilde{\lambda}_r - \lambda_s}. \tag{5.18}$$

Note that  $M_{rs}(0)$  = the matrix  $M$  whose invertibility was assumed in Condition 4 above. Thus  $M(\mathbf{x})$  will be invertible for sufficiently small – in fact, for almost all – values of  $\mathbf{x}$ . Equations (5.15–5.18) thus determine  $g_\infty$ ; since this can be interpreted as the dressing transformation, we can now, in principle, compute  $(u_0, \dots, u_{n-2})$  of  $n$ -KdV.

Formula (2.32) offers a short-cut to an expression for  $u = u_{n-2}$  in terms of  $M(\mathbf{x})$ . Recall that (2.32) says that

$$u = -n\partial_x \text{tr}(g_1 e_1),$$

where  $x = x_{1,1}, A = \lambda e_1 + f_1$  (as usual), and  $g_1$  is the coefficient of  $\lambda^{-1}$  in the expansion of  $g_\infty$  about  $\lambda = \infty$ . In the present case, we obtain from (5.15–5.18):

$$g_1 = \sum_1^k v_s \otimes \beta_s = -\sum_{s,t} (M^{-1})_{st} G_0(\tilde{\lambda}_t)\tilde{v}_t \otimes G_0^{-t}(\lambda_s)\alpha_s$$

so

$$u = n\partial_x \left\{ \sum_{s,t} \alpha_s [G_0^{-1}(\lambda_s)e_1 M_{st}^{-1} G_0(\tilde{\lambda}_t)\tilde{v}_t] \right\}.$$

From (5.18), however,

$$\partial_x M_{rs} = \alpha_s [G_0^{-1}(\lambda_s)G_0(\tilde{\lambda}_r)e_1 \tilde{v}_r],$$

and finally

$$u = n\partial_x \text{tr}(M^{-1}\partial_x M) = n\partial_x^2 \log \det M. \tag{5.19}$$

Thus we have identified the  $\tau$ -function of Sect. 2.1.2, in this case, with  $\det M$ .

It is possible to relax the conditions 1–4 above at the cost of greater complication in the formulae for  $v_s$  and  $\beta_s$  in (5.15). We give a brief treatment of one other interesting case, corresponding to the example on pp. 20–22 of [SW].

We replace (5.12) by

$$\det \gamma(\lambda) = \lambda^{-k}(\lambda - \lambda_1) \dots (\lambda - \lambda_k), \tag{5.20}$$

where  $\lambda_1, \dots, \lambda_k$  are distinct points of  $\mathbb{C}$ , none equal to 0, and assume further that

$$\gamma(\lambda) = A^{-k} \cdot P(\lambda), \gamma(\infty) = 1, \tag{5.21}$$

where  $P(\lambda)$  is polynomial in  $\lambda$ . Then  $P$  is completely determined by (5.20), (5.21), and the annihilator  $\alpha_r$  of  $\text{Im}(\gamma(\lambda_r)) (r = 1, \dots, k)$ , at least if the latter are appropriately generic. For the Birkhoff factorization of  $G \cdot \gamma$ , we assume  $g_\infty$  to be given by (5.15), and parallel reasoning (the zeros of  $\det \gamma$  are still distinct) once again leads to Eq. (5.16) for  $\beta_s$  in terms of  $\alpha_s$ . To deal with the multiple pole at 0 in  $\det \gamma$ , first write  $k = in + A$ , where  $i \geq 0$  and  $0 \leq A \leq n - 1$ . Then  $A^{-k} = \lambda^{-i} A^{-A} = \lambda^{-i} e_{n-A} + \lambda^{-i-1} f_{n-A}$ . As  $P$  is polynomial,  $g_\infty$  will cancel the pole at  $\lambda = 0$  iff

$$g_\infty \text{ has a zero of order } i \text{ at } 0 \text{ and } g_\infty^{(i+1)}(\mathbf{x}, 0) f_{n-A} = 0.$$

This condition yields the following system of equations for the  $v_s$ :

$$1 - \sum \frac{v_s \otimes \beta_s}{\lambda_s} = 0 ;$$

$$\sum_s \frac{v_s \otimes \beta_s}{\lambda_s^r} = 0 \quad (r = 1, \dots, i) ;$$

$$\left[ \sum_s \frac{v_s \otimes \beta_s}{\lambda_s^{i+1}} \right] f_{n-A} = 0 .$$

With (5.16), these determine  $g_\infty$  uniquely. These considerations lead to an identification of the  $\tau$ -function as the determinant of a  $k \times k$  matrix, as in the previous case; explicit formulae appear on p. 21 of [SW].

5.5. *The Ward Transform and the Construction of Segal–Wilson.* In this section we shall establish the equivalence of the construction of Segal–Wilson and the restriction of the Ward transform to bundles of the form  $E_\gamma$ .

We begin with a brief account of the Segal–Wilson construction. This begins with the Grassmannian  $\text{Gr}$  of the Hilbert space  $H = L^2(S^1)$  and the abelian group

$$\Gamma_+ = \left\{ \exp \left( \sum_1^\infty t_i z^i \right) \right\} ,$$

where the sum is assumed convergent in  $\{|z| \leq 1\}$ ;  $\Gamma_+$  acts by multiplication on  $\text{Gr}$ . To a subspace  $W \subset H$  of virtual dimension zero in  $\text{Gr}$  is associated the *Baker function*  $\psi_W$  which is characterized by:

- (a) for all  $\mathbf{t}$ ,  $\psi(\mathbf{t}, z)$  lies in  $W$ ;
- (b)  $\psi$  has an expansion of the form  $\psi(\mathbf{t}, z) = \exp(\sum t_i z^i)(1 + a_1 z^{-1} + \dots)$ .

(Compare with the formal Baker function in Sect. 2.1.2. As there, we write  $\mathbf{t}$  for the collection of all the “time” variables  $t_i$ .) A more geometric version of condition (b) is: for all  $\mathbf{t}$ ,  $g(\mathbf{t}, z)^{-1} \psi(\mathbf{t}, z)$  lies in  $1 + H_-$ . (Here  $H_\pm$  are the positive and negative parts of the standard polarization of  $H$ .) Then  $\psi$  exists for almost all values of  $\mathbf{t}$  because  $W$  is of virtual dimension zero, and for almost all  $\mathbf{t}$ ,  $W \cap g(\mathbf{t})H_- = 0$ .

The significance of  $\psi$  is that it satisfies a family of linear differential equations, derived as follows. With  $\partial_r$  standing for differentiation with respect to  $t_r$  and  $t_1 = x$ , note that

$$\partial_r \psi = g(\mathbf{t}, z)(z^r + a_1 z^{r-1} + O(z^{r-2}))$$

and

$$\partial_x^r \psi = g(\mathbf{t}, z)(z^r + a_1 z^{r-1} + O(z^{r-2})) .$$

Thus

$$\partial_r \psi - D^r \psi = g(\mathbf{t}, z) \times O(z^{r-2}) ,$$

where  $D = \partial_x$ . One can proceed to eliminate the non-negative powers of  $z$  on the RHS by subtraction of appropriate multiples of  $D^i \psi$  for  $i = r - 2, r - 1, \dots, 0$ . The result is a polynomial differential operator  $P_r = D^r + q_{1r} D^{r-2} + \dots$  such that

$$\partial_r \psi - P_r \psi = g(\mathbf{t}, z) \times O(z^{-1}) .$$

Now one argues that the LHS lies in  $W$  by definition of  $\psi$  (the  $t$ -derivatives also lie in  $\psi$  because  $W$  is a closed subspace of  $H$ ) whereas the RHS lies in  $g(\mathbf{t}, z)H_-$ . For almost all  $\mathbf{t}$ ,  $W \cap g(\mathbf{t})H_- = 0$  so *both sides of this equation must vanish*. Thus  $\psi$  satisfies the equation

$$\partial_r \psi - P_r \psi = 0$$

for each  $r = 2, 3, \dots$ ; the compatibility conditions for these equations constitute the equations of the KP-hierarchy.

The  $n$ -KdV hierarchy arises from this construction when the subspace  $W$  satisfies the condition  $z^n W \subset W$ . Indeed, for such  $W$ ,

$$\partial_n \psi - z^n \psi = g(\mathbf{t}, z) \times O(z^{-1})$$

and by the same argument using the transversality of  $W$  and  $gH_-$ , it follows that

$$\partial_n \psi = P_n \psi = z^n \psi$$

in addition to the evolution equations found before. Putting  $P_n = D^n + u_{n-2}D^{n-2} + \dots + u_0$ , and using the essential uniqueness of the  $n$ -KdV hierarchy, we see that we have indeed recovered the Gelfand–Dikii form of that hierarchy (cf. Sect. 2.1). Note again the important principle that the Baker function determines the solution.

To make contact with the twistor construction, some more of the general theory of  $\text{Gr}$  is needed. The key point is that the submanifold  $\text{Gr}^{(n)}$  of subspaces  $W$  which satisfy  $z^n W \subset W$  is intimately related to the loop groups  $LGL_n(\mathbb{C})$  and  $LU_n$  and that, roughly speaking, the transversality argument that we used above corresponds to Birkhoff factorization in  $LGL_n(\mathbb{C})$ .

This theory begins with the identification of  $H$  with  $H^{(n)} = L^2(S^1, \mathbb{C}^n)$ . Indeed, given a function of  $z$ , we obtain a row-vector-valued function of  $\lambda$  by means of

$$f(z) \mapsto (F_1(\lambda), \dots, F_n(\lambda)) = (f(z_1), \dots, f(z_n))A(\lambda)^{-1}, \tag{5.22}$$

where  $z_1, \dots, z_n$  are the  $n$ th roots of  $\lambda$  and

$$A(\lambda)^{-1} = n^{-1} \begin{bmatrix} 1 & z_1^{-1} & \dots & z_1^{1-n} \\ \dots & \dots & \dots & \dots \\ 1 & z_n^{-1} & \dots & z_n^{1-n} \end{bmatrix}. \tag{5.23}$$

The inverse is given more simply by

$$[F_1(\lambda), \dots, F_n(\lambda)] \mapsto f(z) = F_1(z^n) + zF_2(z^n) + \dots + z^{n-1}F_n(z^n).$$

(These formulae have a geometric interpretation in terms of the  $n$ -fold covering map from the  $z$ -plane to the  $\lambda$ -plane. Cf. also Sect. 2.2.2. Because we made our connections act on column-vector-valued functions throughout Sects. 2–4, we are forced into the above identification in which we think of  $H^{(n)}$  as row-vector-valued functions on  $S^1$ . In [SW], the opposite convention is adopted. This is relevant when comparing the formulae at the end of Sect. 5 of their paper with those of the present section.)

If now  $\gamma$  is in  $LGL_n(\mathbb{C})$ , then  $W_\gamma = H_+^{(n)}\gamma^{-1}$  lies in  $\text{Gr}(H^{(n)})$ . Moreover,  $\lambda W \subset W$  so  $W$  lies in  $\text{Gr}^{(1)}(H^{(n)})$ . It is easy to check that under the correspondence between  $H$  and  $H^{(n)}$ ,  $\text{Gr}^{(1)}(H^{(n)})$  corresponds to  $\text{Gr}^{(n)}$ ; so the assignment

$$\gamma \mapsto W_\gamma \tag{5.24}$$

defines a map  $LGL_n(\mathbb{C}) \rightarrow \text{Gr}^{(n)}$ . In fact, this map is *onto* and  $W_\gamma = W_\delta$  iff  $\gamma^{-1}\delta \in L^+GL_n(\mathbb{C})$ . Thus (5.24) induces a map

$$LGL_n(\mathbb{C})/L^+GL_n(\mathbb{C}) \rightarrow \text{Gr}^{(n)}$$

(where on the LHS,  $L^+GL_n(\mathbb{C})$  acts according to  $(g_+, \gamma) \mapsto \gamma g_+^{-1}$ ). In natural topologies on the two sets involved, this map is a diffeomorphism of infinite-dimensional manifolds. Moreover, the connected component of the identity on the LHS is mapped onto the submanifold of virtual-dimension-zero points of  $\text{Gr}^{(n)}$ . We note that Birkhoff factorization identifies a dense open subset of this identity-component with  $L^-GL_n(\mathbb{C})$  (and as we indicated before, this identification with  $\text{Gr}^{(n)}$  can be used to *prove* the Birkhoff factorization theorem [PS, Ch. 8]).

To bring this general theory to bear upon the  $n$ -KdV hierarchy, we must first identify the action on  $LGL_n(\mathbb{C})$  that is induced by the multiplication action of  $\Gamma_+$  on  $\text{Gr}$ . From (5.22) and (5.23) (cf. also (2.24) and (2.25)), this is given by

$$(F_1, \dots, F_n) \mapsto (F_1, \dots, F_n)A(\lambda) \text{diag}(g(z_1), \dots, g(z_n))A(\lambda)^{-1} \tag{5.25}$$

$$= (F_1, \dots, F_n) \exp(\sum t_r A^r) . \tag{5.26}$$

Now fix  $W = W_\gamma$  in  $\text{Gr}^{(n)}$ . One calculates, as in Sect. 2.2.2, that the Baker function  $\psi$  of  $W$  goes over to the “Baker *matrix*”  $\Psi$  of  $\gamma$  which has the following characterization:

- (a)' for all  $\mathbf{t}$ ,  $\Psi(\mathbf{t}, \lambda)^{-1}$  lies in the orbit  $\gamma \cdot L^+GL_n$  in  $LGL_n$ ;
- (b)'  $\Psi$  has a factorization of the form  $\Psi(\mathbf{t}, \lambda) = g_\infty(\mathbf{t}, \lambda)G_0(\mathbf{t}, \lambda)$  with  $g_\infty(\mathbf{t}, \infty) = 1$ .

(Here  $G_0 = \exp(\sum t_r A^r)$ .)

In terms of these quantities, the construction of Segal and Wilson associates to  $\gamma$  the  $n$ -KdV solution with dressing transformation =  $g_\infty$ .

One can see that  $g_\infty$  is a dressing transformation in 2 ways. The first is to note that  $\psi$  is the Baker function of an  $n$ -KdV solution and that  $g_\infty$  is related to  $\psi$  exactly as were the formal dressing transformation  $g$  and the formal Baker function  $\psi$  in Sect. 2.2.2. The second is to use the two expressions for  $\Psi$ ,

$$\Psi(\mathbf{t}, \lambda) = g_0(\mathbf{t}, \lambda)\gamma(\lambda)^{-1} = g_\infty(\mathbf{t}, \lambda)G_0(\mathbf{t}, \lambda) , \tag{5.27}$$

which follow from (a)' and (b)', to derive evolution equations of the type  $L_r \Psi = 0$ , where  $L_r$  is as in (2.33) with  $g = g_\infty$ .

The solution of the  $n$ -KdV equation that comes from the Ward transform of  $E_\gamma$  also has dressing transformation =  $g_\infty$ , for eliminating  $\Psi$  from (5.27) gives

$$g_\infty(\mathbf{t})^{-1} \cdot g_0(\mathbf{t}) = G_0(\mathbf{t}) \cdot \gamma .$$

To recover (5.7), just pre-multiply both sides by  $\exp(\sum x_{a0} \lambda^{-1} A^a)$ , replacing  $g_\infty$  by  $g'_\infty$  and  $G_0(\mathbf{t})$  by  $G(\mathbf{x})$ .

Hence the Ward transform, restricted to bundles generated by the Segal–Wilson ansatz, coincides with the Segal–Wilson construction.

*Acknowledgement.* The authors are grateful to Nicholas Woodhouse for a number of helpful observations and discussions.

## References

- [BE] Bailey, T.N., Eastwood, M.G.: Complex paraconformal manifolds – their differential geometry and twistor theory. *Forum. Math.* **3**, 61–103 (1991)
- [DS] Drinfeld, V.G., Sokolov, V.V.: Equations of Korteweg de Vries type and simple Lie algebras. *Sov. Math. Dokl.* **23**, 457–462 (1981). See also the long version: Lie algebras and equations of the Korteweg de Vries type. *J. Sov. Math.* **30**, 1975 (1985)
- [Du] Dubrovin, B.A.: Theta functions and nonlinear equations. *Russ. Math. Surveys.* **36**, 2, 11–92 (1981)
- [GD] Gelfand I.M., Dikii, L.A.: Fractional powers of operators and Hamiltonian systems. *Funct. Anal. Appl.* **10**, (4) (1976)
- [GHM] De Groot, M.F., Hollowood, T.J., Miramontes, J.L.: Generalized Drinfeld–Sokolov hierarchies. *Commun. Math. Phys.* **145**, 57 (1992)
- [JM] Jimbo M., Miwa T.: Solitons and infinite dimensional Lie algebras. *Publ. RIMS, Kyoto Univ.*, **19**, 943–1001 (1983)
- [KW] Kac V.G., Wakimoto.: Exceptional hierarchies of soliton equations. *Proceedings of Symposia in pure Mathematics*, Vol. **49**, 191 (1989)
- [M] Mason L.J.: Twistor theory, self-duality and integrability. In: *Proceedings of NATO A.R.W., Exeter 1992*, Clarkson, P., (ed.) Kluwer, 1993 See also  $\mathcal{H}$ -space, a universal integrable system? in *Twistor News-letter* **30**, (1990)
- [MS] Mason, L.J., Sparling, G.A.J.: Twistor correspondences for the soliton hierarchies. *J. Geom. Phys.* **8**, 243–271 (1992). See also Nonlinear Schrödinger and Korteweg de Vries are reductions of self-dual Yang–Mills. *Phys. Lett. A* **137**, 29–33 (1989)
- [MW] Mason, L.J., Woodhouse, N.M.J.: Self-duality and the Painleve transcendents. To appear in *Nonlinearity*, 1993
- [PS] Pressley, A., Segal, G.B.: *Loop Groups*. Oxford Science Publications, 1986, OUP
- [SW] Segal, G.B., Wilson, G.: Loop Groups and equations of KdV type. *Publ. Math. IHES* **61**, 5–65 (1985)
- [Wa77] Ward, R.S.: On self-dual gauge fields. *Phys. Lett. A* **61**, 81–82 (1977)
- [Wa84] Ward, R.S.: Completely solvable gauge fields in dimensions greater than four. *Nucl. Phys. B* **236**, 381–396 (1984)
- [Wa] Ward, R.S.: Integrable and solvable systems and relations among them, *Phil. Trans. Roy. Soc. Lond. A* **315**, 451–457 (1985). See also: Multidimensional integrable systems. In: *Field Theory, Quantum Gravity and Strings*. de Vega, H.J., Sanchez, N. (eds.) *Lecture Notes in Physics*, Vol. **246** (Berlin, Heidelberg, New York: Springer:) 1986, and ‘Integrable systems in twistor theory’. In: *Twistors in Mathematics and Physics*. Bailey, T.N., Baston, R.J. (eds.) *L.M.S. Lecture Notes* **156**, C.U.P.
- [W] Wilson, G.: Habillage et fonctions  $\tau$ . *C. R. Acad. Sc. Paris* **299**, Sér. I, 587–590 (1984)

Communicated by M. Jimbo