Systems & Control Letters 9 (1987) 361–367
North-Holland

A remark on bilinear systems and moduli spaces of instantons

Eduardo D. SONTAG *
Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA

Received 12 May 1987
Revised 15 July 1987

Abstract: Explicit equations are given for the moduli space of framed instantons as a quasi-affine variety, based on the representation theory of noncommutative power series, or equivalently, the minimal realization theory of bilinear systems.

Keywords: Bilinear systems, Invariant theory, Yang–Mills theory.

1. Introduction

In a recent paper [3], Helmke showed how results of Donaldson [1] in Yang–Mills theory are closely related to system theoretic notions, in particular to what are sometimes called 'multirate systems'. He then went on to provide a number of results on the topology of the space of framed instantons and of a certain space in which they can be naturally embedded. In this note we simply remark that it is also possible to view in a natural way these same objects as bilinear systems, or equivalently, via minimal representations of matrix power series. An advantage of this alternative interpretation is that the machinery of Hankel matrices can then be applied to understand the structure of the corresponding moduli space. In particular, we obtain one natural representation of this quotient space as a quasi-affine variety. For motivation and for a discussion of the origin of the problem being considered, see [3] and the references given there.

2. Framed instantons

We shall use the conclusion of [3], Theorem 2.2, as our starting point. Thus we shall be concerned with the set $\mathcal{M}(n,m)$ (or, more precisely, $M(\mathrm{SU}(m),n)$ in [3]) obtained as follows for all positive integers n and m.

First let $\mathcal{B}(n,m)$ be the set of all quadruples

$$(A_1,\ A_2,\ B,\ C) \in \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times n} \times \mathbb{C}^{n\times m} \times \mathbb{C}^{m\times n}$$

that satisfy the following three conditions:

(1) $A_2A_1 = A_1A_2 + BC$,

(2) rank $\mathcal{R}(A_1,\ A_2,\ B,\ n-1) = n$, and

(3) rank $\mathcal{R}(A_1',\ A_2',\ C', n-1) = n$.

Here prime indicates transpose; for each k and all matrices X_1, X_2 and Y of sizes $n\times n$, $n\times n$, and $n\times m$ respectively, the expression $\mathcal{R}(X_1,\ X_2,\ Y,\ k)$ stands for the block matrix

$$\left[Y,\ X_1Y,\ X_2Y,\ X_1^2Y,\ X_1X_2Y,\ldots,\ X_\alpha Y,\ldots,\ X_2^kY\right].$$

* Research supported in part by US Air Force Grant 0247.

Each block has size $n \times m$. There are $2^{k+1}-1$ such blocks indexed by the possible sequences α of 1's and 2's of length at most k (including the empty sequence λ) ordered lexicographically, and

$$X_\alpha := X_{\alpha_1} X_{\alpha_2} \cdots X_{\alpha_l},$$

for each $\alpha = (\alpha_1, \ldots, \alpha_l)$. Note that the columns of $\mathscr{R}(X_1, X_2, Y, n-1)$ span the smallest subspace of $\mathbb{C}^n$ which contains the columns of Y and is both X_1- and X_2-invariant; thus property 2 is equivalent to the rank of $\mathscr{R}(A_1, A_2, B, j)$ being n for any $j \geq n$, and similarly for property 3.

Now consider the following natural algebraic action of $\mathrm{GL}(n, \mathbb{C})$ on the variety $\mathscr{B}(n, m)$,

$$(A_1, A_2, B, C) \mapsto (T^{-1}A_1T, T^{-1}A_2T, T^{-1}B, CT).$$

Then $\mathscr{M}(n, m)$ is defined topologically as the quotient space of $\mathscr{B}(n, m)$ under this action. We shall show how, as a quotient space of a variety under an algebraic group, hence in particular also as a topological space, $\mathscr{M}(n, m)$ is isomorphic to a subset of $\mathbb{C}^k$ (k an integer depending on n, m) defined by certain polynomial equalities and inequalities, that is, a quasi-affine variety, and more interestingly, how this follows from the facts in the representation theory of noncommutative power series. The study of the latter was initiated by Schützenberger and Fliess, who developed the foundational results. Expositions can be found in [2] and [4], [5]; here we follow and refine the material in [7], Section 19, and [8].

A system-theoretic interpretation of the set $\mathscr{B}(n, m)$ is, modifying Fliess [6], via continuous-time (or discrete-time) *bilinear systems*

$$\dot{x}(t) = (A_1 + u(t)A_2)x(t) + Bv(t), \quad x(0) = 0,$$
$$y(t) = Cx(t),$$

where the complex state x is n-dimensional, and there are $m+1$ independent inputs u, $v = (v_1, \ldots, v_m)$ and an m-dimensional output. In the case of [6], there is no independent v-control, but the initial state is not necessarily zero. Thus we are really looking at the somewhat more general class of *state-affine* systems introduced in [7], Section 20.

Conditions 2 and 3 in the definition of $\mathscr{B}(n, m)$ correspond to *span reachability* and *observability* of the system. The latter is the usual concept, and the former means that there is no proper subspace of the state space $\mathbb{C}^n$ which contains the set reachable from the origin.

Condition 1 appears not very natural in this context, but becomes easy to handle in an input/output sense. The input/output behavior of such systems is described uniquely by the set of impuls-response coefficient matrices, each $m \times m$:

$$H_\alpha = CA_{\alpha_1}A_{\alpha_2} \cdots A_{\alpha_l}B \tag{2.1}$$

where α is a sequence of 1's and 2's, that is, an element of the free monoid (of 'words' in the 'letters' 1,2) $W = \{1,2\}^*$. Note that condition 1 in the definition of $\mathscr{B}(n, m)$ implies that

$$H_{\alpha 21 \beta} = H_{\alpha 12 \beta} + H_\alpha H_\beta \quad \text{for all words } \alpha, \beta \in W. \tag{2.2}$$

Conversely, under the span rechability and observability properties 2 and 3, property (2.2) implies property 1, as we remark later. The impulse response coefficients are invariant under the change-of-basis action of $\mathrm{GL}(n,\mathbb{C})$ used in obtaining the quotient $\mathscr{M}(n,m)$, and the canonical realization theory of state affine systems tells us that they are complete invariants, in a precise sense to be reviewed below. In particular, by partial realization theory, we can embed the space $\mathscr{M}(n,m)$ in a truncated space of impulse-response coefficient matrices. Moreover, property (2.2) will in fact tell us that we can restrict attention to a much smaller space of rational power series in two commutative variables, corresponding to what in image processing are called 'separable' 2D recursive filters. All these interpretations are of course used only for motivation; the results to follow are purely algebraic.

3. Impulse-response matrices

From now on, fix n, m and let $B := \mathscr{B}(n,m)$ and $M := \mathscr{M}(n,m)$. Let $N := m^2(2^{2n+1}-1)$. We view $\mathbb{C}^N$ as the set of all sequences $\mathscr{H}$ of $m \times m$ matrices

$$(H_\lambda,\ H_1,\ H_2, \ldots,\ H_\alpha, \ldots,\ H_{2^{2n}})$$

indexed lexicographically by the words in $\mathscr{W}$ of length at most $2n$. To any such $\mathscr{H}$ and positive integers i, j with $+i+j \le 2n$, we associate the *(generalized) Hankel* or *behavior matrix* Hankel($\mathscr{H}, i, j$). The latter is defined as follows. It is a block matrix, with blocks of size $m \times m$; the block rows and columns are indexed respectively by the words in $\mathscr{W}$ of length at most i and j respectively, again ordered lexicographically. In block position (α, β), we place the entry $H_{\alpha\beta}$, where the concatenation $\alpha\beta$ of

$$\alpha = (\alpha_1, \ldots,\ \alpha_k) \quad \text{and} \quad \beta = (\beta_1, \ldots,\ \beta_l)$$

is the sequence

$$(\alpha_1, \ldots,\ \alpha_k,\ \beta_1, \ldots,\ \beta_l)$$

(in the case in which $\alpha = \lambda$, the empty word, this is just β, and similarly if $\beta = \lambda$). Thus Hankel($\mathscr{H}, i, j$) has the pattern

$$\begin{pmatrix} H_\lambda & H_1 & H_2 & H_{11} & H_{12} & \cdots & H_{2^j} \\ H_1 & H_{11} & H_{12} & H_{111} & H_{112} & \cdots & H_{12^j} \\ H_2 & H_{21} & H_{22} & H_{211} & H_{212} & \cdots & H_{2^{j+1}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{2^i} & H_{2^i1} & H_{2^{i+1}} & H_{2^i11} & H_{2^i12} & \cdots & H_{2^{i+j}} \end{pmatrix}.$$

Given an element $\mathscr{H}$ of $\mathbb{C}^N$, we shall say that $\mathscr{H}$ *is of degree n* iff the following condition holds:

$$\text{rank Hankel}(\mathscr{H},\ n-1,\ n-1) = \text{rank Hankel}(\mathscr{H},\ n,\ n) = n. \tag{3.1}$$

Note that, since rank $= n$ means that an $n \times n$ minor must be nonzero and all minors of size $n+1$ must vanish, the set of all $\mathscr{H}$ of degree n can be seen as a Zariski-open subset of the algebraic set consisting of all $\mathscr{H}$ for which Hankel($\mathscr{H}, n, n$) has rank at most n. Thus the set of $\mathscr{H}$ of degree n is a locally closed subset of $\mathbb{C}^N$, that is, a *quasi-affine subvariety*. Finally, we consider the subset $\mathscr{N}$ of $\mathbb{C}^N$ consisting of all the sequences $\mathscr{H}$ of degree n for which the condition (2.2) holds for all words $\alpha 12 \beta$ of length at most $2n$. This set is again a quasi-affine variety, because it is obtained by adding the quadratic equations (2.2) to the above minor conditions.

Given an element $\Sigma = (A_1,\ A_2,\ B,\ C) \in \mathscr{B}$, the formula (2.1) allows us to define an element $\gamma(\Sigma) \in \mathbb{C}^N$. Since for each i and j it holds that

$$\text{Hankel}(\gamma(\Sigma),\ i,\ j) = \mathscr{R}(A_1',\ A_2',\ C',\ i)'\mathscr{R}(A_1,\ A_2,\ B,\ j),$$

$\gamma(\Sigma)$ has degree n. Further, $A_2A_1 = A_1A_2 + BC$ implies that

$$CA_\alpha A_2 A_1 A_\beta B = CA_\alpha A_1 A_2 A_\beta B + CA_\alpha BCA_\beta B$$

for all words α, β, so that $\gamma(\Sigma)$ is in fact in $\mathscr{N}$.

Conversely, given a sequence $\mathscr{H} \in \mathscr{N}$, we now show how to build an element $\Sigma \in \mathscr{B}$ that maps into it. First we consider the n-dimensional vector subspace $\mathscr{X}$ of $\mathbb{C}^{m(2^n-1)}$ spanned by the columns of Hankel($\mathscr{H}$, $n-1$, n). Because of the assumption that $\mathscr{H}$ has degree n, $\mathscr{X}$ is spanned also by the columns of the submatrix Hankel($\mathscr{H}$, $n-1$, $n-1$).

Let A_μ, $\mu = 1,2$, be the linear operator defined on the generators of $\mathscr{X}$ as follows. In general, let $v_{\beta,i}$ (respectively, $\tilde{v}_{\beta,i}$) be the i-th column $(1 \le i \le m)$ of the block of columns of Hankel($\mathscr{H}$, $n-1$, n)

(respectively, Hankel($\mathscr{H}$, n, n)) indexed by β. Then, if β has length at most $n-1$ (so that $v_{\beta,i}$ is a column of Hankel($\mathscr{H}$, $n-1$, $n-1$)), let

$$A_\mu v_{\beta,i} := v_{\mu\beta,i}.$$

Since these generators do not necessarily form a basis, it must be checked that any relation among them is preserved by A_μ. Assume that there are complex numbers $c_{\tau,i}$ such that

$$\sum c_{\tau,i} v_{\tau,i} = 0, \tag{3.2}$$

where the sum is taken over $i = 1, \ldots, m$ and over all words τ of length $\leq n$. We wish to show that then it must also hold that

$$\sum c_{\tau,i} v_{\mu\tau,i} = 0. \tag{3.3}$$

letting η be the column vector whose entries are the $c_{\tau,i}$'s in an appropriate order, equation (3.2) can be restated as

$$\text{Hankel}(\mathscr{H},\ n-1,\ n-1)\eta = 0. \tag{3.4}$$

Write Hankel($\mathscr{H}$, n, $n-1$) in partitioned form

$$\text{Hankel}(\mathscr{H},\ n,\ n-1) = \begin{pmatrix} \text{Hankel}(\mathscr{H},\ n-1,\ n-1) \\ Y \end{pmatrix}.$$

Because of condition (3.1), the rows of Y are linearly dependent on those of Hankel($\mathscr{H}$, $n-1$, $n-1$), so there is a matrix X such that

$$Y = X\ \text{Hankel}(\mathscr{H},\ n-1,\ n-1). \tag{3.5}$$

It follows from equations (3.4) and (3.5) that $Y\eta = 0$, so that also Hankel($\mathscr{H}$, n, $n-1$)$\eta = 0$. Thus (3.2) holds also for the longer columns,

$$\sum c_{\tau,i} \tilde{v}_{\tau,i} = 0. \tag{3.6}$$

In terms of the entries of the $\tilde{v}_{\tau,i}$'s, (3.6) means that

$$\sum c_{\tau,i} H_{\alpha\tau,i} = 0 \tag{3.7}$$

for all α of length $\leq n$, where $H_{\beta,i}$ denotes the i-th column of the matrix H_β. Now consider the desired relation (3.3). In terms of block entries, it is required that

$$\sum c_{\tau,i} H_{\alpha\mu\tau,i} = 0 \tag{3.8}$$

for all α of length at most $n-1$ (rather than n). Since for such α, the word $\alpha\mu$ has length at most n, (3.8) is a particular case of the known equalities (3.7). We conclude that the mappings A_μ are indeed well-defined.

We now complete the construction of Σ by defining B and C. Let $B : \mathbb{C}^m \to \mathscr{X}, e_i \mapsto v_{\gamma,i}$ (e_i = i-th canonical basis element) and let $C : \mathscr{X} \to \mathbb{C}^m$ be the projection on the first m components. Choosing any basis for $\mathscr{X}$, there results a quadruple

$$\Sigma = (A_1,\ A_2,\ B,\ C)$$

that satisfies properties 2 and 3. Assume for a moment that we proved that $\gamma(\Sigma) = \mathscr{H}$. We now show that then property 1 is satisfied too. More generally, we wish to establish that $A_2A_1 = A_1A_2 + BC$ follows if properties 2 and 3 hold and property (2.2) holds for all α, β of length at most $n-1$. Letting

$$L := [A_2,\ A_1] = A_2A_1 - A_1A_2, \qquad N := BC,$$

it is enough to establish that

$$CA_\alpha LA_\beta B = CA_\alpha NA_\beta B \tag{3.9}$$

for all such α, β implies that $L = N$.

But the equation (3.9) implies that

$$\mathscr{R}(A_1', A_2', C', n-1)' L \mathscr{R}(A_1, A_2, B, n-1) = \mathscr{R}(A_1', A_2', C', n-1)' N \mathscr{R}(A_1, A_2, B, n-1).$$

It follows from the full rank assumptions 2 and 3 that indeed $L = N$. Thus $\Sigma \in \mathscr{B}$, as desired.

We now establish the claim that $\gamma(\Sigma) = \mathscr{H}$. Since

$$\text{rank Hankel}(\mathscr{H}, n-1, n) = \text{rank Hankel}(\mathscr{H}, n-1, n-1),$$

it follows that the j-th column of the block with index β, length of $\beta = n$, is in the span of those with shorter indexes. Thus for each such β, j there are complex numbers $c_{\tau,i}^{\beta,j}$ such that

$$v_{\beta,j} = \sum c_{\tau,i}^{\beta,j} v_{\tau,i}, \tag{3.10}$$

the sum over all the indices $i = 1, \ldots, m$ and all the words τ of length less than n. We now *define* elements $v_{\sigma,j}$ for σ of length larger than n in the following way. Let σ be any such word, and factor it as $\theta\beta$, with β of length n. Now use the formula

$$v_{\theta\beta,j} = \sum c_{\tau,i}^{\beta,j} v_{\theta\tau,i}, \tag{3.11}$$

inductively on the length of θ. We claim that $A_\mu v_{\sigma,j} = v_{\mu\sigma,j}$ for each $\mu = 1,2$ and each σ, j. This is true by definition of the linear maps A_μ when the length of σ is at most n. For other σ it follows by induction on θ when applying A_μ to both sides of (3.11). When restricted to each block of m rows, equation (3.10) says that

$$H_{\theta\beta,j} = \sum c_{\tau,i}^{\beta,j} H_{\theta\tau,i} \tag{3.12}$$

($H_{\alpha,i}$ denotes the i-th column of H_α) for each β of length n and each θ of length at most n. By construction we know that

$$Cv_{\sigma,j} = H_{\sigma,j}$$

for all σ of length at most n. We claim that this is true for σ of length up to $2n$. Decomposing $\sigma = \theta\beta$ as above, the claim follows by induction on the length of θ:

$$Cv_{\theta\beta,j} = CA_\theta v_{\beta,j} = \sum c_{\tau,i}^{\beta,j} CA_\theta v_{\tau,i} = \sum c_{\tau,i}^{\beta,j} Cv_{\theta\tau,i} = \sum c_{\tau,i}^{\beta,j} H_{\theta\tau,i} = H_{\theta\beta,j},$$

the last two steps by induction and by (3.12) respectively. This establishes the claim. From this it follows that

$$(CA_\alpha B)_i = CA_\alpha v_{\lambda,i} = Cv_{\alpha,i} = H_{\alpha,i}$$

and hence that $\lambda(\Sigma) = \mathscr{H}$, as desired.

From standard realization theory (see e.g., [7], Section 19) the map γ is one-to-one on $\mathscr{B}$ up to the above action of $\mathrm{GL}(n, \mathbb{C})$ (elements of $\mathscr{B}$ are 'minimal' or 'canonical' realizations of $\mathscr{H}$). We have therefore proved that γ induces a bijection between $\mathscr{M}$ and $\mathscr{N}$. We now establish that $(\mathscr{N}, \gamma)$ is a quotient in the category of varieties as well.

Consider any variety $\mathscr{C}$ and a morphism $f : \mathscr{B} \to \mathscr{C}$ constant on orbits. There is a unique map

$$g : \mathscr{N} \to \mathscr{C}$$

so that $g \circ \gamma = f$, and the problem is to show that this map is a morphism. To show this, it is sufficient to cover $\mathscr{N}$ by Zariski-open sets in such a way that the restriction of g to each such open is a morphism. For each possible n-minor of Hankel($\mathcal{H}$, $n-1$, $n-1$), consider the open set where this is nonzero. In each such set, the Σ constructed above can be obtained by rational functions of the entries of the matrices $\mathcal{H}_\alpha$; this is basically Cramer's rule, see [8], realization algorithm 2.8. Let h be the morphism on the corresponding open set which assigns Σ to the given $\mathcal{H}$. Then,

$$g = f \circ h$$

on this open set, which is a morphism as desired. We summarize the above discussion with:

Theorem 3.1. *The mapping* $\gamma : \mathcal{B} \to \mathcal{N}$ *is a quotient of* $\mathcal{B}$ *under the action of* $\mathrm{GL}(n, \mathbb{C})$.

4. Some simplifications

The space $\mathbb{C}^N$ in which the quasi-affine variety $\mathcal{N}$ was defined is of unnecessarily high dimension. Indeed, it is also possible to build a quotient as follows. Let $K := m^2\binom{n+1}{2}$. We view elements of $\mathbb{C}^K$ as matrix polynomials

$$L(x, y) = \sum_{i+j \leq n} L_{i,j} x^i y^j,$$

by listing all coefficient matrices in any fixed order. Now, given any $\mathcal{H}$ satisfying (2.2), it is possible to 'straighten' each word β into a commutative product $1^i 2^j$, with i (respectively, j) being the number of 1's (respectively, 2's) appearing in β, modulo words of smaller length. More precisely, there exist for all words β of length at most $2n$, polynomials P_β in K variables such that, for each $\mathcal{H} \in \mathcal{N}$,

$$H_\beta = P_\beta(\mathcal{H}(x, y)),$$

where $\mathcal{H}(x, y)$ denotes the matrix polynomial with $L_{i,j} = H_{1^i 2^j}$.

It follows that the quasi-affine subvariety $\mathcal{N}'$ of $\mathbb{C}^K$ consisting of all $\{\mathcal{H}(x, y), \mathcal{H} \in \mathcal{N}\}$ is isomorphic to $\mathcal{N}$.

For any $\mathcal{H}(x, y) \in \mathcal{N}'$, consider the corresponding $\mathcal{H}$ and pick any Σ with $\gamma(\Sigma) = \mathcal{H}$. We may now define

$$H_\alpha := CA_\alpha B$$

for *all* words $\alpha \in \mathcal{W}$. The corresponding sequence of all possible such H_α is a *rational* multiset or noncommutative power series, in the sense of automata theory (see [2,4,5,7,8]). Its restriction (Hadamard product) to the set of indices of the type $1^i 2^j$ is hence also rational as a noncommutative power series. Thus the matrix power series in *commuting* variables

$$L(x, y) := \sum_{i,j=0}^{\infty} H_{1^i 2^j} x^i y^j$$

is *recognizable* ([4], page I.2.26). Recall that a recognizable matrix power series can be represented as an $(m \times m)$ matrix of rational functions

$$\frac{p(x, y)}{q_1(x) q_2(y)},$$

and that these appear in image processing; see for instance [9], Section IV. It follows that the quotient space $\mathcal{M}$ can be naturally represented in terms of a set of recognizable matrix power series. It might be of some interest to understand this representation in some detail.

References

[1] S.K. Donaldson, Instantons and geometric invariant theory, *Commun. Math. Phys.* **93** (1984) 453–460.
[2] S. Eilenberg, *Automata, Languages, and Machines, Volume* A (Academic Press, New York, 1974).
[3] U. Helmke, Linear dynamical systems and instantons in Yang–Mills theory, *IMA J. Math. Control and Information* **3** (1986) 151–166.
[4] M. Fliess, Sur certaines familles de séries formelles, Thèse de Doctorat d'Etat, Université Paris VII (1972).
[5] M. Fliess, Matrices de Hankel, *J. Math. Pures Appl.* **53** (1974) 197–224.
[6] M. Fliess, Sur la realization des systèmes dynamiques bilineaires, *C.R. Acad. Sci. Paris Sér. A.* **277** (1973) 243–247.
[7] E.D. Sontag, *Polynomial Response Maps* (Springer, Berlin–New York, 1979).
[8] E.D. Sontag, Realization theory of discrete-time nonlinear systems: Part I – The bounded case, *IEEE Trans. Circuits and Systems* **26** (1979) 342–356.
[9] E.D. Sontag, On first-order equations for multidimensional filters, *IEEE Trans. Acoustics Speech Signal Process.* **26** (1978) 480–482.