

## On the solvability of linear matrix equations

**Citation for published version (APA):**

Hautus, M. L. J. (1982). *On the solvability of linear matrix equations*. (Eindhoven University of Technology : Dept of Mathematics : memorandum; Vol. 8207). Technische Hogeschool Eindhoven.

**Document status and date:**

Published: 01/01/1982

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

696962

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics and Computing Science

Memorandum 1982-07

ON THE SOLVABILITY OF LINEAR  
MATRIX EQUATIONS

by

M.L.J. Hautus

April 1982

ON THE SOLVABILITY OF LINEAR

MATRIX EQUATIONS

by

M.L.J. Hautus

Dept. of Mathematics & Computing Science  
University of Technology  
Eindhoven, the Netherlands.

ABSTRACT

Linear matrix equations were studied by Sylvester, Stéphanos, Datuashvili and Roth. In this paper, solvability conditions given by these authors are generalized in various directions: to nonsquare equations, nonpolynomial type equations, in particular equations given by an integral, and finally to equations over an arbitrary commutative ring with unit element.

## 1. Introduction

The object of this paper is to give necessary and sufficient conditions for the matrix equation

$$(1.1) \quad \sum_{i=1}^k A_i X B_i = C$$

to have a solution  $X$ . Distinction is made between universal and individual solvability. Equation (1.1) is called universally solvable if it has a solution for every  $C$ . Universal solvability thus is a condition on the matrices  $A_i$  and  $B_i$ . Equation (1.1) is called (individually) solvable if it has a solution for the particular  $C$  given.

Equations of the form (1.1) were considered in literature (see e.g. [4, Ch VIII], [8, Ch VIII], [6]). In principle, it is possible to rewrite (1.1) using tensor products and to give solvability conditions in terms of the coefficient matrices thus obtained (see [6], [8]). Our objective, however, is to find conditions expressed more directly in terms of the matrices  $A_i$  and  $B_i$ . It seems unlikely that such a condition can be found for the general case of equation (1.1). But for special cases, a number of (more or less known) results can be given. Sometimes these conditions are formulated in terms of the spectrum (i.e. the set of eigenvalues) of the map

$$\mathcal{L} : X \mapsto \sum_{i=1}^k A_i X B_i.$$

For this to be possible it is necessary that  $\mathcal{L}$  map a certain matrix space (say,  $\mathbb{R}^{n \times m}$ , the space of real  $n \times m$  matrices) into itself. In this par-

ticular case, computation of the spectrum  $\sigma(\mathcal{L})$  of  $\mathcal{L}$  is equivalent to determination of universal solvability conditions for (1.1). In fact, (1.1) is universally solvable iff  $0 \notin \sigma(\mathcal{L})$ . Conversely,  $\lambda \in \sigma(\mathcal{L})$  iff the equation

$$\mathcal{L}(X) - \lambda X = C$$

is universally solvable. When in the rest of this introduction referring to the literature, we will not explicitly distinguish between universal solvability conditions and spectrum computations.

Let us briefly describe some of the most important results on the solvability of equations of the type (1.1). In 1884, Sylvester showed that the equation

$$(1.2) \quad AX - XB = C$$

is universally solvable iff  $\sigma(A) \cap \sigma(B) = \emptyset$  (see [8, Theorem 46.2]). This equation will henceforth be referred to as Sylvester's equation. The result was extended in 1900 by C. Stéphanos (see [8, Theorem 43.8]) to equations of the form (1.1), where  $A_i = p_i(A)$ ,  $B_i = q_i(B)$ . Here  $A$  and  $B$  are matrices and  $p_i$  and  $q_i$  are polynomials. Stéphanos expresses his condition in terms of the polynomial

$$(1.3) \quad p(z,s) := \sum p_i(z)q_i(s)$$

associated to equation (1.1). Specifically, he shows that (1.1) is universally solvable iff  $p(\lambda,\mu) \neq 0$  for  $\lambda \in \sigma(A), \mu \in \sigma(B)$ . This result is easily seen to be an extension of Sylvester's result.

A further generalization was obtained by G.S. Datuashvili in 1966 (see [3]). Datuashvili allows  $A_i$  to be arbitrary and maintains the condition

that  $B_i$  be of the form  $B_i = q_i(B)$ :

(1.4) THEOREM. (Datuashvili). Let  $A_i \in \mathbb{R}^{m \times m}$ ,  $B \in \mathbb{R}^{p \times p}$  and let  $q_i(s)$  be a polynomial for  $i = 1, \dots, k$ . The equation

$$(1.5) \quad \sum_{i=1}^k A_i X q_i(B) = C$$

is universally solvable iff the associated polynomial matrix

$$(1.6) \quad A(s) := \sum_{i=1}^k A_i q_i(s)$$

is nonsingular for  $s \in \sigma(B)$ .

Again, it is easily seen that the result generalizes Stéphanos' result. Datuashvili's proof can briefly be described as follows: First he assumes without loss of generality that  $B$  is upper triangular. Writing the map  $X \mapsto \sum A_i X q_i(B)$  as a tensor product map he notices that the coefficient matrix will be upper block triangular, so that its invertibility properties can be inferred from the entries on the block diagonal.

In section 2 we give an alternative proof, which does not use tensor products but is based on the substitution of matrices into polynomial matrices (see [4, Ch IV, §3]). The proof given has the advantage that it yields an explicit formula for the solution. Furthermore, it can be generalized in various ways. In Theorem 2.4, Datuashvili's result is generalized to the case where  $A_i$  is allowed to be nonsquare. Furthermore, in Theorem 2.13 the requirement that  $B_i$  be of the form  $B_i = q_i(B)$  will be relaxed. A condition is given which is valid if it is only known that

the  $B_i$ 's commute. This is a true generalization since matrices  $B_1, \dots, B_k$  can commute without being polynomial in a fixed matrix  $B$  (see [2, Section IV]). Also, the method can be used to give universal solvability conditions for a continuous version of (1.5), viz.

$$(1.7) \quad \int_a^b A(t)X f(t,B)dt = C ,$$

see section 3, where this result is obtained as a special case of a more general type of equation.

Finally, in section 5, the results are extended to equations over an arbitrary commutative ring  $\mathcal{R}$ . Of course, in this general situation, it is not possible to give a condition in terms of eigenvalues. But the conditions given in Theorem 1.4 and its generalizations can be formulated in an "eigenvalue-free" way. E.g. introducing the polynomials

$$a(s) := \det A(s), \quad b(s) := \det(sI - B)$$

we can formulate the condition of Theorem 1.4 as:  $a(s)$  and  $b(s)$  have the bezoutian property, i.e. there exist polynomials  $u(s)$  and  $v(s)$  such that  $u(s)a(s) + v(s)b(s) = 1$ . It turns out that formulated this way, Theorem 1.4 extends to general commutative rings. It should be remarked that the more obvious condition: " $a(s)$  and  $b(s)$  are coprime" turns out to be too weak in general rings.

In the particular case that  $\mathcal{R} = \mathbb{C}[\underline{\xi}]$ , where  $\underline{\xi} = (\xi_1, \dots, \xi_v)$  (or, more generally,  $\mathcal{R} = K[\underline{\xi}]$ , where  $K$  is any algebraically closed field),  $a$  and  $b$  are polynomials  $a(\underline{\xi}, s)$  and  $b(\underline{\xi}, s)$  in  $\underline{\xi}$  and  $s$ . It follows from Hilbert's

Nullstellensatz (see [1, V 3.3]) that  $a(\xi, s)$  and  $b(\xi, s)$  have the bezoutian property iff they have no common zero. The absence of common zeroes thus will be a necessary and sufficient condition for the universal solvability of (1.5). This can be formulated as follows:

(1.8) THEOREM. Let  $A_i \in (\mathbb{C}[\xi])^{n \times n}$ ,  $B \in (\mathbb{C}[\xi])^{m \times m}$  and let  $q_i \in \mathbb{C}[\xi, s]$ .

Then the equation

$$\sum A_i(\xi) X(\xi) q_i(\xi, s) = C(\xi)$$

has a solution  $X(\xi) \in (\mathbb{C}[\xi])^{n \times m}$  for every  $C \in (\mathbb{C}[\xi])^{n \times m}$  iff the poly-  
mial matrix

$$A(\xi, s) := \sum A_i(\xi) q_i(\xi, s)$$

is nonsingular for every  $\xi \in \mathbb{C}^v$  and every eigenvalue  $s$  of  $B(\xi)$ .

In section 3 the individual solvability of (1.1) is investigated. For Sylvester's equation (1.2) a well-known condition was given by W. Roth in 1952 (see [11]).

Specifically:

(1.9) THEOREM (Roth). Given  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{p \times p}$  and  $C \in \mathbb{R}^{n \times m}$ , equation  
(1.2) has a solution if and only if the matrices

$$(1.10) \quad \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}, \quad \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

are similar .



An obvious question is how to generalize this result to equations of the form considered e.g. in Theorem 1.4 and its generalizations. A generalization in terms of similarity seems unlikely to be possible. However, according to ([4, VI §4 and 5]), two matrices  $M$  and  $N$  are similar iff  $sI-M$  and  $sI-N$  are  $\mathbb{R}[s]$ -equivalent, i.e. there exist  $\mathbb{R}[s]$ -invertible matrices  $P(s)$  and  $Q(s)$  such that  $P(s)(sI - M) = (sI - N)Q(s)$ . Consequently, the matrices (1.10) are similar iff

$$\begin{pmatrix} sI-A & -C \\ 0 & sI-B \end{pmatrix}, \begin{pmatrix} sI-A & 0 \\ 0 & sI-B \end{pmatrix}$$

are  $\mathbb{R}[s]$ -equivalent. In this formulation, Roth's theorem can be extended as follows

(1.11) THEOREM. Let  $A_i, B, q_i$  and  $A(s)$  be given as in Theorem 1.4. The following statements are equivalent:

- i) (1.5) has a solution,
- ii) The equation

$$(1.12) \quad A(s)U(s) + V(s)(sI - B) = C$$

has a solution  $(U(s), V(s)) \in (\mathbb{R}[s])^{n \times m} \times (\mathbb{R}[s])^{n \times m}$ ,

- iii) The matrices

$$\begin{pmatrix} A(s) & -C \\ 0 & sI-B \end{pmatrix}, \begin{pmatrix} A(s) & 0 \\ 0 & sI-B \end{pmatrix}$$

are  $\mathbb{R}[s]$ -equivalent.

The  $\mathbb{R}[s]$ -equivalence of two polynomial matrices can be checked by computing their invariant factors (see [4, VI, §3 Cor 1]).

Theorem 1.11 will be proved in section 4. In addition, some generalizations will be given. Finally, in section 5 the result will be generalized to equations over a commutative ring (based on a result by W. Gustafson [5]).

## 2. Universal solvability conditions

We start with a proof of Datuashvili's theorem 1.4:

PROOF. "if". The matrix  $A(s)$  is invertible as a rational matrix and we have the following relation

$$A(s)D(s) = a(s)I$$

where  $a(s) := \det A(s)$  and  $D(s)$  is the adjoint matrix. It is given that  $a(\mu) \neq 0$  for  $\mu \in \sigma(B)$ . Hence,  $a(B)$  is invertible. Define

$$C_1 := C(a(B))^{-1}$$

$$E(s) := D(s)C_1 .$$

Then

$$A(s)E(s) = \sum_{i=1}^k A_i E(s)q_i(s) = C_1 a(s) .$$

We substitute  $s = B$  from the right into this equation and denote by  $E(B)$  the result of substituting  $B$  from the right into  $E(s)$  (see [4,IV, §3]). The following equality results:

$$\sum_{i=1}^k A_i E(B)q_i(B) = C_1 a(B) = C ,$$

which shows that  $X := E(B)$  is a solution of (1.5).

"only if" It is easily seen that if (1.5) has a real solution for every real  $C$ , then it has a complex solution for every complex  $C$ . Suppose that for some  $\mu \in \sigma(B)$  the matrix  $A(\mu)$  is not invertible. Let  $p$  and  $q$  be non-zero vectors such that  $q'A(\mu) = 0$ ,  $Bp = \mu p$ . We claim that (1.5) has no

solution for  $C := qp'$ . In fact, for any matrix  $X$  we have

$$q' \sum A_i X q_i(B) p = q' A(\mu) X p = 0,$$

whereas  $q' C p = q' q p' p \neq 0$ . □

The following corollary is the result as it was actually stated by Datuashvili:

(2.1) COROLLARY. The spectrum of the map

$$\mathcal{L} : X \mapsto \sum A_i X q_i(B) : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}$$

is

$$\sigma(\mathcal{L}) := \bigcup_{\mu \in \sigma(B)} \sigma(A(\mu)).$$

This result is derived from Theorem 1.4 in the way suggested in the introduction, i.e. via the fact that  $\lambda \in \sigma(\mathcal{L})$  iff  $\mathcal{L} - \lambda I$  is not surjective.

The proof given above yields an explicit solution of equation (1.5). In the particular case of Sylvester's equation, we have  $A(s) = sI - A$  (apart from an irrelevant minus sign). Hence, the solution of (1.2) (under the assumption  $\sigma(A) \cap \sigma(B) = \emptyset$ ) is given by

$$(2.2) \quad X = (DC)(B) a^{-1}(B)$$

where  $(DC)(B)$  is the result of right substitution of  $B$  into the polynomial matrix  $D(s)C$ . (Note that  $B$  and  $a^{-1}(B)$  commute), and  $D(s)$  is the adjoint matrix of  $A$  (Compare [6, section 11], where the solution of Sylvester's

equation is expressed in the adjoint matrix under the assumption that A is simple). Using the algorithm of Souriau-Frame-Faddeev(see [4, IV, §5] or [10, Ch 1, section 2] ), equation (2.2) can be reduced to the following algorithm.

(2.3) COROLLARY. Consider Sylvester's equation (1.2), assume that  $\sigma(A) \cap \sigma(B) = \emptyset$  and define matrices  $L_k, M_k, Y_k$  for  $k = 0, \dots, n$  and numbers  $b_k$  for  $k = 0, \dots, n-1$  by

$$\begin{aligned} L_0 &:= M_0 := I, Y_0 := C, \\ b_k &:= -(k+1)^{-1} \operatorname{tr}(M_k A), \\ M_{k+1} &:= M_k A + b_k I, \\ L_{k+1} &:= L_k B + b_k I, \\ Y_{k+1} &:= Y_k B + M_k C \end{aligned}$$

for  $k = 0, \dots, n-1$ . Then  $X := Y_n L_n^{-1}$  is the solution of (1.2).

The following is a generalization of Theorem 1.4 to the case where  $A_i$ 's are not square.

(2.4) THEOREM. Let  $A_i \in \mathbb{R}^{n \times m}$ ,  $B \in \mathbb{R}^{p \times p}$ ,  $q_i \in \mathbb{R}[s]$ . The equation

$$(2.5) \quad \sum_{i=1}^k A_i X q_i(B) = C$$

is universally solvable if and only if

$$(2.6) \quad A(s) := \sum_{i=1}^k A_i q_i(s)$$

has full row rank for every  $s \in \sigma(B)$ .

PROOF. The necessity is proved the same way as in Theorem 1.4. For the sufficiency we can also use the same proof, provided we can find a polynomial matrix  $D(s)$  and a scalar polynomial  $a(s)$  such that

$$(2.7) \quad A(s) D(s) = a(s)I,$$

and  $a(B)$  is invertible. For this one can use the Smith canonical form for polynomial matrices (see [7, Theorem II,9]). In fact, we can write  $A = U \Delta V$  where  $U$  and  $V$  are  $\mathbb{R}[s]$ -invertible and  $\Delta = [\Delta_1, 0]$ ,  $\Delta_1 := \text{diag}(\psi_1, \dots, \psi_n)$ ,  $\psi_1 | \psi_2 | \dots | \psi_n$ . Let  $\Lambda_1$  be the diagonal matrix for which  $\Delta_1 \Lambda_1 = \psi_n I$ . Then we may choose  $D = V^{-1} \Lambda U^{-1}$ , where  $\Lambda := [\Lambda_1, 0]'$ , and  $a(s) = \psi_n(s)$ . Since  $a(s)$  is the G.C.D. of the  $n \times n$  minors of  $A(s)$ , we have that  $a(\mu) \neq 0$  for  $\mu \in \sigma(B)$ . □

An alternative construction for  $D$  and  $a$ , not depending on the Smith canonical form, will follow from the proof of Theorem 2.13 below.

In the following we replace the assumption  $B_i = q_i(B)$  by the weaker condition  $B_i B_j = B_j B_i$ . First we need some preliminary concepts and results.

(2.8) DEFINITION. Let  $B_1, \dots, B_k$  be commutative  $m \times m$  matrices. A vector  $\underline{\lambda} = (\lambda_1, \dots, \lambda_k) \in \mathbb{C}^k$  is a (joint)eigentuple of  $B_1, \dots, B_k$  if there exists a common corresponding eigenvector, i.e. if there exists  $v \neq 0$  such that

$$B_i v = \lambda_i v \quad (i = 1, \dots, k).$$

The proof of the following lemma shows in particular the existence of joint eigentuples.

(2.9) LEMMA. Let  $B_1, \dots, B_k$  be commutative matrices and let  $\psi(s_1, \dots, s_k) = \psi(\underline{s})$  be a polynomial. Then  $\psi(B_1, \dots, B_k)$  is nonsingular iff  $\psi(\underline{\lambda}) \neq 0$  for any joint eigentuple of  $B_1, \dots, B_k$ .

PROOF. First we observe that, if  $B \in \mathbb{R}^{m \times m}$  and  $w \in \mathbb{R}^m$ ,  $w \neq 0$ , then there exists a polynomial  $p(s)$  such that  $p(B)w$  is an eigenvector of  $B$ . In fact, if  $q(s)$  is a nonzero polynomial of minimal degree such that  $q(B)w = 0$ , then it is easily seen that  $\deg q \geq 1$ , so that we can find  $\lambda$  and  $p(s)$  such that  $q(s) = (s - \lambda)p(s)$ . Then  $v := p(B)w \neq 0$  (since  $\deg p < \deg q$ ) and  $(B - \lambda I)v = q(B)w = 0$ .

Now assume that  $\psi(\underline{\lambda}) = 0$  for some eigentuple  $\underline{\lambda}$  of  $B_1, \dots, B_k$ . Then  $B_i v = \lambda_i v$  for some  $v \neq 0$ . Hence  $\psi(B_1, \dots, B_k)v = \psi(\underline{\lambda})v = 0$ , so that  $\psi(B_1, \dots, B_k)$  is singular.

Conversely, assume that  $\psi(B_1, \dots, B_k)w_0 = 0$  for some  $w_0 \neq 0$ . By the above observation, there exists a polynomial  $p_1(s)$  such that  $w_1 := p_1(B_1)w_0$  is an eigenvector of  $B_1$ . Applying the observation repeatedly, we obtain a sequence of vectors  $w_i$  and numbers  $\lambda_i$  satisfying

$$w_i = p_i(B_i)w_{i-1} \neq 0,$$

$$B_i w_i = \lambda_i w_i.$$

Finally we find  $v = w_k$  and, by the commutativity of the  $B_i$ 's, it is easily seen that  $B_i v = \lambda_i v$  for  $i = 1, \dots, k$ . Hence  $\underline{\lambda} := (\lambda_1, \dots, \lambda_k)$  is an eigentuple. In addition

$$\psi(B_1, \dots, B_k)v = p_k(B_k) \dots p_1(B_1)\psi(B_1, \dots, B_k)w = 0.$$

But,  $\psi(B_1, \dots, B_k)v = \psi(\underline{\lambda})v$ , and, consequently  $\psi(\underline{\lambda}) = 0$ . □

(2.10) REMARK. In terms of the spectrum, we have

$$(2.11) \quad \sigma(\psi(B_1, \dots, B_k)) = \{\psi(\underline{\lambda}) \mid \underline{\lambda} \text{ is an eigentuple of } B_1, \dots, B_k\} .$$

This property is usually taken as a definition of the joint spectrum (see [12, §1]). Specifically,

$$\sigma(B_1, \dots, B_k) := \{\underline{\lambda} \in \mathbb{C}^k \mid \forall \psi \in \mathbb{C}[s_1, \dots, s_k] \psi(\underline{\lambda}) \in \sigma(\psi(B_1, \dots, B_k))\}$$

Equivalently one can say that  $\underline{\lambda} \in \sigma(B_1, \dots, B_k)$  iff for any polynomial  $\psi(\underline{s})$  we have that  $\psi(\underline{\lambda}) = 0$  implies that  $\psi(B_1, \dots, B_k)$  is singular.

Let us show that  $\underline{\lambda} \in \sigma(B_1, \dots, B_k)$  iff  $\underline{\lambda}$  is a joint eigentuple.

If  $\underline{\lambda}$  is a joint eigentuple, say  $B_i v = \lambda_i v$ , then  $\psi(B_1, \dots, B_k)v = \psi(\underline{\lambda})v$ , so that  $\psi(\underline{\lambda}) = 0$  implies that  $\psi(B_1, \dots, B_k)$  is singular.

Conversely, assume that there does not exist  $v$  such that  $B_i v = \lambda_i v$  for  $i = 1, \dots, k$ . Then

$$\text{rank} \begin{pmatrix} B_1 - \lambda_1 I \\ \vdots \\ B_k - \lambda_k I \end{pmatrix} = m .$$

It follows from the Lemma below that complex numbers  $\alpha_1, \dots, \alpha_k$  exist such that  $\sum_{i=1}^k \alpha_i (B_i - \lambda_i I)$  is nonsingular. Consequently, if we define

$$\psi(s_1, \dots, s_k) := \alpha_1 s_1 + \dots + \alpha_k s_k - (\alpha_1 \lambda_1 + \dots + \alpha_k \lambda_k) ,$$

then  $\psi(\underline{\lambda}) = 0$  and  $\psi(B_1, \dots, B_k)$  is nonsingular.



LEMMA. Let  $B_0, \dots, B_k$  be commutative  $m \times m$  matrices and let  $\text{rank}[B_0', \dots, B_k'] = m$ .

Then there exists  $\alpha$  such that  $B(\alpha) := \sum_0^k B_i \alpha^i$  is nonsingular.

PROOF. If  $B(\alpha)$  is singular for all  $\alpha$ , then  $B(s)$  is singular over the field of rational functions  $\mathbb{R}(s)$ . Hence, there exists a rational vector  $p(s) \neq 0$  such that  $B(s) p(s) = 0$ . We may assume that  $p(s)$  is polynomial. Let

$p(s) = \sum_{j=0}^v p_j s^j$ , where  $p_v \neq 0$ , and assume that  $p(s)$  is of minimal degree.

Then the equation  $B(s) p(s) = 0$  reads

$$(2.12) \quad \left\{ \begin{array}{l} B_0 p_0 = 0, \\ B_0 p_1 + B_1 p_0 = 0, \\ \vdots \\ B_0 p_v + B_1 p_{v-1} + \dots + B_k p_{v-k} = 0, \\ \vdots \\ B_k p_v = 0. \end{array} \right.$$

We notice that

$$B(s)(B_k p(s)) = B_k B(s) p(s) = 0,$$

and that  $\deg(B_k p(s)) < \deg p(s)$  because of  $B_k p_v = 0$ . By the minimality condition on  $p(s)$ , it follows that  $B_k p(s) = 0$ , i.e.,  $B_k p_i = 0$  for  $i = 0, \dots, v$ .

Considering the second last equation,  $B_k p_{v-1} + B_{k-1} p_v = 0$ , we see that  $B_{k-1} p_v = 0$ . Since  $B(s)(B_{k-1} p(s)) = 0$  we repeat the previous reasoning and conclude  $B_{k-1} p_i = 0$ . Thus continuing, we obtain

$$B_\ell p_v = 0 \quad (\ell = 0, \dots, k),$$

contradicting the assumption of the Lemma. □

Using Lemma 2.9, we are able to prove the following generalization of Theorem 2.4.

(2.13) THEOREM. Let  $A_i \in \mathbb{R}^{n \times m}$ ,  $B_i \in \mathbb{R}^{p \times p}$  and suppose that  $B_i B_j = B_j B_i$  ( $i, j = 1, \dots, k$ ). Then the equation

$$(2.14) \quad \sum_{i=1}^k A_i X B_i = C$$

is universally solvable iff  $A(\underline{\lambda})$  has full row rank for every joint eigentuple of  $B_1, \dots, B_k$ . Here

$$A(\underline{s}) = A(s_1, \dots, s_k) := \sum_{i=1}^k A_i s_i .$$

For the proof we need a multidimensional interpolation result:

(2.15) LEMMA. Let  $z_i \in \mathbb{C}^k$  ( $i = 0, \dots, \ell$ ) be distinct points. Then there exists  $\varphi(\underline{s}) \in \mathbb{C}[s_1, \dots, s_k]$  such that  $\varphi(\underline{z}_0) = 1$ ,  $\varphi(\underline{z}_i) = 0$  ( $i = 1, \dots, \ell$ ).

PROOF. Let  $\underline{z}_i = (z_{i,1}, \dots, z_{i,k})$ . By Lagrange's interpolation theorem there exists for  $j = 1, \dots, k$  a polynomial  $L_j \in \mathbb{C}[s]$ , such that  $L_j(z_{i,j}) = 1$  if  $i = 0$  and zero otherwise.

Now the function

$$\varphi(\underline{s}) = \prod_{j=1}^k L_j(s_j)$$

satisfies the requirements. □

PROOF of Theorem 2.13 : "if". If  $\underline{\lambda} = (\lambda_1, \dots, \lambda_k)$  is a joint eigentuple of  $B_1, \dots, B_k$ , then necessarily,  $\lambda_i \in \sigma(B_i)$ . Hence there are at most finitely many eigentuples, say  $\underline{\lambda}^{(1)}, \dots, \underline{\lambda}^{(\ell)}$ . Choose polynomials

$\varphi_i \in \mathbb{C}[s_1, \dots, s_k]$  such that  $\varphi_i(\underline{\lambda}^{(j)}) = \delta_{ij}$  ( $i, j = 1, \dots, \ell$ ). This is possible because of the previous Lemma. Define

$$F(\underline{s}) := \sum_{i=1}^{\ell} F_i \varphi_i(\underline{s}),$$

where  $F_i$  is an  $m \times n$  matrix such that  $A(\underline{\lambda}^{(i)})F_i = I$ . Such an  $F_i$  exists because of the assumption of the Theorem. Then  $A(\underline{s})F$  is an  $n \times n$  polynomial matrix invertible on the joint eigentuples of  $B_1, \dots, B_k$ . Let  $G(\underline{s})$  and  $a(\underline{s})$  be such that  $AFG = aI$  and  $a(\underline{\lambda}^{(i)}) \neq 0$  for  $i = 1, \dots, \ell$ . Because of Lemma 2.9,  $a(B_1, \dots, B_k)$  is invertible. Setting  $FG =: D$ , we can complete the "if"-part of the proof exactly as in the proof of Theorem 1.4.

"only if": This proof is completely similar to the corresponding proof of Theorem 1.4. □

### 3. Generalizations and applications

Consider a  $p \times p$  matrix  $B$  and an  $n \times n$ -matrix-valued function  $A(s)$  analytic on (a neighbourhood of)  $\sigma(B)$ . The right substitution of  $B$  into  $A(s)$  is defined by

$$(3.1) \quad A(B) := \int_{\Gamma} A(s) (sI - B)^{-1} ds$$

where  $\Gamma$  is a contour surrounding  $\sigma(B)$  and contained in the domain of analyticity of  $A(s)$ , and  $ds$  stands for  $ds/(2\pi i)$ . It is easily seen that in the case where  $A(s)$  is a polynomial, this definition coincides with the one used in section 2. If  $X$  is a constant  $m \times p$  matrix and  $A(s)$  an  $n \times m$ -matrix-valued map we define the  $n \times p$ -matrix-valued function  $AX$  by

$$(3.2) \quad (AX)(s) := A(s)X .$$

The equation in  $X$  that we consider in this section, is

$$(3.3) \quad (AX)(B) = C,$$

where  $C$  is a given  $n \times p$  matrix. This equation is readily seen to reduce to (1.5) when  $A(s)$  is defined by (1.6). The following result generalizes Theorem 2.4:

(3.4) THEOREM. Equation (3.3) is universally solvable if and only if  $A(s)$  has full row rank on  $\sigma(B)$ . In this case there exists a matrix function  $D(s)$ , analytic on  $\sigma(B)$ , such that  $A(s) D(s) = I$ . A solution of (3.3) is

$$(3.5) \quad X = (DC)(B) .$$

PROOF. "Only if": Suppose that for some  $\mu$  we have nonzero vectors  $v, w$  such that  $v'A(\mu) = 0, Bw = \mu w$ . Then

$$\begin{aligned} v'(AX)(B)w &= \int_{\Gamma} v'A(s)X(sI - B)^{-1}w \, ds = \\ &= \int_{\Gamma} v' \frac{A(s) - A(\mu)}{s - \mu} Xw \, ds = 0, \end{aligned}$$

since the integrand is analytic in the domain enclosed by  $\Gamma$ . Hence, (3.3) does not have a solution when  $v'cw \neq 0$ .

"If": Exactly as in the proof of Theorem 2.13, one can use interpolation to construct a matrix-valued function  $D(s)$  analytic on  $\sigma(B)$  and such that  $A(s)D(s) = I$ . We show that (3.5) is a solution of (3.3):

$$\begin{aligned} (AX)(B) &= \int_{\Gamma_s} A(s)(DC)(B)(sI - B)^{-1} \, ds = \\ &= \int_{\Gamma_s} \int_{\Gamma_z} A(s)D(z)C(zI - B)^{-1}(sI - B)^{-1} \, dz \, ds. \end{aligned}$$

We choose for  $\Gamma_z$  a contour surrounding  $\sigma(B)$  but contained in the domain enclosed by  $\Gamma_s$ . Using the well-known formula

$$(zI - B)^{-1}(sI - B)^{-1} = (z - s)^{-1}((sI - B)^{-1} - (zI - B)^{-1}),$$

we can write  $(AX)(B) = J_1 + J_2$ , where

$$J_1 := \int_{\Gamma_s} A(s) \int_{\Gamma_z} D(z)C(z - s)^{-1} \, dz (sI - B)^{-1} \, ds = 0$$

because  $z \mapsto (z - s)^{-1}$  is analytic in the domain enclosed by  $\Gamma_z$ . Furthermore,

$$\begin{aligned}
 J_2 &= - \int_{\Gamma_z} \int_{\Gamma_s} A(s) D(z) C(z-s)^{-1} ds (zI - B)^{-1} dz = \\
 &= \int_{\Gamma_z} A(z) D(z) C(zI - B)^{-1} dz = C . \quad \square
 \end{aligned}$$

In the proof of Theorem 2.4,  $D(s)$  and  $a(s)$  were chosen such that  $AD = aI$ . If one allows  $D$  to be an arbitrary analytic function instead of a polynomial, like we do here, we may replace  $D$  by  $D/a$ . In the particular case where  $m = n$ , the number of equations is equal to the number of unknowns. In this case, (3.5) is the unique solution of (3.3) (under the conditions of the theorem). Also, as in Corollary 2.1

$$\mathcal{L} : X \mapsto (AX)(B) : \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{n \times m}$$

has the spectrum

$$\sigma(\mathcal{L}) = \bigcup_{\mu \in \sigma(B)} \sigma(A(\mu)) .$$

We mention two special cases of equation (3.3):

(3.6) EXAMPLE. Assume that  $A_0, A_1, \dots$  is a sequence of  $n \times m$  matrices such that  $\sum_0^{\infty} \|A_i\| \alpha^i < \infty$ , where  $\alpha > 0$ . Let  $B$  be any  $p \times p$  matrix with spectral radius less than  $\alpha$ . Then

$$A(s) := \sum_0^{\infty} A_i s^i$$

is analytic for  $|s| < \alpha$  and  $(AX)(B)$  is defined for every  $m \times p$  matrix  $X$ .

It is not difficult to verify that

$$(AX)(B) = \sum_0^{\infty} A_i X B^i,$$

so that the equation reads

$$\sum_0^{\infty} A_i X B^i = C.$$

□

(3.7) EXAMPLE. Let

$$A(s) := \int_0^T L(t) e^{-st} dt.$$

Then  $A(s)$  is an entire function and

$$(AX)(B) = \int_0^T L(t) X e^{-tB} dt.$$

Consider the special case  $L(t) = e^{tA}$ . Then

$$A(s) = \int_0^T e^{(A-sI)t} dt$$

and  $A(\mu)$  is nonsingular iff

$$\int_0^T e^{(\lambda-\mu)t} dt \neq 0$$

for  $\lambda \in \sigma(A)$ , i.e. iff  $\lambda - \mu \neq 2\pi ik/T$  for any nonzero integer  $k$ . Hence the equation

$$\int_0^T e^{tA} X e^{-tB} dt = C$$

is universally solvable iff for nonzero  $k \in \mathbb{Z}$  we have  $2\pi ik/T \notin \sigma(A) - \sigma(B)$ .  
(Compare [9, §14.2]).

This example can be generalized in a straightforward way to the equation

$$\int_V L(v) X f(v, B) d\mu = C$$

where  $(V, \mathcal{F}, \mu)$  is a compact topological measure space and  $f : V \times \mathbb{C} \rightarrow \mathbb{C}$  is a continuous function such that  $s \mapsto f(v, s)$  is analytic on  $\sigma(B)$  for every  $v \in V$ . Here, of course,

$$A(s) = \int_V L(v) f(v, s) d\mu .$$

□

Let  $P(s)$  be an  $n \times n$ -matrix-valued function analytic in a certain domain  $\Omega$  in  $\mathbb{C}$ . Then  $P(s)$  defines a mapping

$$P : X \mapsto P(X)$$

for  $X \in \mathbb{C}^{n \times n}$  with spectrum contained in  $\Omega$ . We are interested in the question of when  $P$  is locally invertible (with  $\mathbb{C}^1$  inverse) at a given matrix  $B$ . For this we apply the implicit-function theorem. That is, we investigate whether the linearization of  $P$  at  $B$  is invertible. We have for small  $Y$ :

$$\begin{aligned} P(B + Y) - P(B) &= \int_{\Gamma} P(s) \{(sI - B - Y)^{-1} - (sI - B)^{-1}\} ds \\ &= \int_{\Gamma} P(s) (sI - B)^{-1} Y (sI - B - Y)^{-1} ds. \end{aligned}$$



It follows that the linearization of  $P(X)$  at  $B$  equals

$$\begin{aligned} \mathcal{L}(Y) &:= \int_{\Gamma^-} P(s)(sI - B)^{-1} Y(sI - B)^{-1} ds = \\ &= \int_{\Gamma} A(s) Y(sI - B)^{-1} ds = (AY)(B) , \end{aligned}$$

where

$$A(s) := (P(s) - P(B))(sI - B)^{-1}$$

is analytic on  $\Omega$ , and hence on  $\sigma(B)$  .

Here we use that

$$\int_{\Gamma} (sI - B)^{-1} Y(sI - B)^{-1} ds = 0 ,$$

as one can easily verify by letting  $\Gamma$  be a circle with radius tending to  $\infty$ .

We remark that we can write

$$A(s) = Q(s, B)$$

where

$$Q(s, z) := \begin{cases} (P(s) - P(z))/(s - z) & (s \neq z) \\ P'(s) & (s = z) . \end{cases}$$

According to Theorem 3.4, it follows that  $\mathcal{L}$  is invertible iff  $A(\mu)$  is invertible for  $\mu \in \sigma(B)$ .

In the particular case that  $P(s) = p(s)$  is a scalar analytic function, the condition can be further simplified. In this case, according to the spectral-mapping theorem,  $A(\mu)$  is nonsingular iff  $Q(\lambda, \mu) \neq 0$  for  $\lambda \in \sigma(B)$ .

Hence we find:

If  $P$  is defined by

$$P : X \mapsto p(X) ,$$

then  $P$  is locally invertible at  $B$  iff

- i)  $p(\lambda) \neq p(\mu)$        $(\lambda, \mu \in \sigma(B), \lambda \neq \mu)$
- ii)  $p'(\lambda) \neq 0$        $(\lambda \in \sigma(B))$ .

Notice that these conditions are exactly the conditions for the function  $p(s)$  to be locally invertible on  $\sigma(B)$ , i.e., for the existence of a function  $q(s)$  analytic on a neighbourhood of  $p(\sigma(B))$  such that  $q(p(s)) = s$ . Hence the inverse of  $P$  is given by:

$$Q : X \mapsto q(X) .$$

We conclude that we have the following:

If  $P : X \mapsto p(X)$  has a  $C^1$  inverse at a certain matrix  $B$ , then there is an inverse  $Q$  of the form  $Q : X \mapsto q(X)$ .

Notice that not every function analytic in a neighbourhood of a certain matrix has the representation  $q(X)$  (e.g.  $P(X) = X^T$ ).

4. Individual solvability conditions

PROOF of Theorem 1.11.

i)  $\Rightarrow$  ii): Let X be a solution of (1.5). Then

$$C - A(s)X = \sum A_i X (q_i(B) - q_i(s)I) = \sum A_i X V_i(s) (sI - B) ,$$

where  $V_i(s) := -\psi_i(s, B)$  and

$$\psi_i(s, z) := \frac{q_i(z) - q_i(s)}{z - s} .$$

Hence  $U(s) := X, V(s) := \sum A_i X V_i(s)$  form a solution of (1.12).

ii)  $\Rightarrow$  i): Right substitution of B into (1.12) yields

$$\sum A_i U(B) q_i(B) = C .$$

Hence  $X = U(B)$  is a solution of (1.5).

ii)  $\Leftrightarrow$  iii): In [11] it is shown that the polynomial equation

$$A(s) U(s) + V(s) B(s) = C(s)$$

has a solution iff

$$\begin{pmatrix} A(s) & C(s) \\ 0 & B(s) \end{pmatrix} , \begin{pmatrix} A(s) & 0 \\ 0 & B(s) \end{pmatrix}$$

are  $\mathbb{R}[s]$ -equivalent. Application of this to  $B(s) := sI - B$ ,  $C(s) := C$  yields the result. □

We mention two generalizations of Theorem 1.11.

(4.1) THEOREM. Let  $A_i, B_i$  be as in Theorem 2.13 and let  $C \in \mathbb{R}^{n \times p}$ . The following statements are equivalent :

- i) Equation (2.14) has a solution  $X$ .
- ii) The polynomial equation

$$A(\underline{s}) U(\underline{s}) + V(\underline{s}) \begin{bmatrix} s_1 I - B_1 \\ \vdots \\ s_k I - B_k \end{bmatrix} = C$$

has a solution  $U(\underline{s}) \in (\mathbb{R}[\underline{s}])^{n \times m}$ ,  $V(\underline{s}) \in (\mathbb{R}[\underline{s}])^{n \times m}$ . Here,  $\underline{s} = (s_1, \dots, s_k)$

- iii) The matrices

$$\begin{bmatrix} A(\underline{s}) & C \\ 0 & \begin{bmatrix} s_1 I - B_1 \\ \vdots \\ s_k I - B_k \end{bmatrix} \end{bmatrix}, \begin{bmatrix} A(\underline{s}) & 0 \\ 0 & \begin{bmatrix} s_1 I - B_1 \\ \vdots \\ s_k I - B_k \end{bmatrix} \end{bmatrix}$$

are  $\mathbb{R}[\underline{s}]$ -equivalent.

PROOF. i)  $\Leftrightarrow$  ii): similar to the previous proof. ii)  $\Leftrightarrow$  iii): Here one uses Gustafson's extension of Roth's theorem to general commutative rings (see [5]). In this generalization Gustafson, states that the matrix equation

$$AU + VB = C$$

over a commutative ring  $\mathcal{R}$  has a solution iff the matrices

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}, \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

are  $\mathcal{R}$ -equivalent. (See also section 5). □

(4.2) THEOREM. Let  $A(s)$ ,  $B$  be given as in Theorem 3.4, and let  $C \in \mathbb{R}^{n \times p}$ .

Then the following statements are equivalent:

i) Equation (3.3) has a solution,

ii) The equation

$$A(s) U(s) + V(s)(sI - B) = C$$

has a solution  $U(s)$ ,  $V(s)$  analytic on  $\Omega$ .

iii) The matrices

$$\begin{pmatrix} A(s) & C \\ 0 & sI - B \end{pmatrix}, \quad \begin{pmatrix} A(s) & 0 \\ 0 & sI - B \end{pmatrix}$$

are equivalent with respect to the ring of functions analytic on  $\Omega$ .

The proof of i)  $\Leftrightarrow$  ii) is based on contour-integral manipulations as in section 3. The proof of ii)  $\Leftrightarrow$  iii) depends again on Gustafson's result.

It is of interest to see whether the results of section 2 can be recovered from the previous results. The following lemma is instrumental.

(4.3) LEMMA. Let  $A(s)$  and  $B(s)$  be polynomial matrices. Then the following statements are equivalent:

i) The equation

$$(4.4) \quad A(s) U(s) + V(s) B(s) = C(s)$$

has a solution  $(U(s), V(s))$  for every polynomial matrix  $C(s)$  (of suitable dimensions).

ii) The equation

$$(4.5) \quad A(s) U(s) + V(s) B(s) = C$$

has a solution for every constant matrix  $C$ .

iii) For any  $s_0 \in \mathbb{C}$ ,  $A(s_0)$  has full row rank or  $B(s_0)$  has full column rank.

PROOF. i)  $\Rightarrow$  ii) is trivial.

ii)  $\Rightarrow$  iii) Suppose that for some  $s_0 \in \mathbb{C}$  there exist nonzero vectors  $v, w$  such that  $v'A(s_0) = 0, B(s_0)w = 0$ . Multiplying (4.4) from the left with  $v'$  and from the right with  $w$ , we find  $v'Cw = 0$ , which is not true for every  $C$ .

iii)  $\Rightarrow$  i) Using Smith canonical decompositions for  $A$  and  $B$  one can "diagonalize" equation (4.4), i.e. we may assume that  $A(s)$  and  $B(s)$  are diagonal. The  $(i, j)^{\text{th}}$  equation reads

$$a_i(s) u_{ij}(s) + b_j(s) v_{ij}(s) = c_{ij}(s) .$$

These equation have solutions, since iii) implies that  $a_i(s)$  and  $b_j(s)$  are coprime. □

Combining Lemma 4.3 , where  $B(s) := sI - B$ , with Theorem 1.11 we find a new proof of Theorem 2.4.

5. Matrix equations over rings

In this section  $\mathcal{R}$  denotes a commutative ring with unit element. We consider equation (1.1) again, but now we assume that the matrices  $A_i, B_i$  and  $C$  have entries in  $\mathcal{R}$  and we try to find a solution  $X$  with entries in  $\mathcal{R}$ . For individual solvability the result is straightforward.

(5.1) THEOREM. The equivalences as stated in Theorem 1.11 remain valid if  $\mathbb{R}$  is everywhere replaced by  $\mathcal{R}$ .

PROOF. The proof of i)  $\Leftrightarrow$  ii) carries over to the ring case. For the proof of the equivalence ii)  $\Leftrightarrow$  iii) we use Gustafson's generalization of Roth's theorem (see [5, Theorem 1], compare the proof of Theorem 4.1).  $\square$

Our next objective is the extension of Theorem 2.4 to the ring case. We say that polynomials  $a_0(s), \dots, a_\ell(s) \in \mathcal{R}[s]$  have the bezoutian property (or are bezoutian) if polynomials  $q_0(s), \dots, q_\ell(s) \in \mathcal{R}[s]$  exist such that  $\sum a_i(s) q_i(s) = 1$ , i.e., if  $a_0, \dots, a_\ell$  span the unit ideal in  $\mathcal{R}[s]$ .

(5.2) LEMMA. Let  $a_i \in \mathcal{R}[s]$  for  $i = 0, \dots, \ell$  and let  $a_0(s)$  be monic (i.e. with leading coefficient 1). Then  $a_0, \dots, a_\ell$  are bezoutian iff  $a_{0,\mu}, \dots, a_{\ell,\mu}$  are bezoutian in  $\mathcal{R}_\mu[s]$  for every maximal ideal  $\mu$  of  $\mathcal{R}$ . Here  $a_{i,\mu} = a_i \pmod{\mu}$  denotes the residue class of  $a_i$  modulo  $\mu$  and  $\mathcal{R}_\mu := \mathcal{R}/\mu$  is the quotient ring of  $\mathcal{R}$  with respect to  $\mu$ .

PROOF. We want to apply [1, Ch II, §3.3. Prop 11] : If  $M$  and  $N$  are  $\mathcal{R}$ -modules and  $N$  is finitely generated, then an  $\mathcal{R}$ -homomorphism  $A : M \rightarrow N$  is surjective iff for each maximal ideal  $\mu$  of  $\mathcal{R}$ , the map  $A_\mu : M_\mu \rightarrow N_\mu$  derived

from A by taking quotients, is surjective.

Here  $M_\mu = M/\mu M$ ,  $N_\mu := N/\mu N$ . One might be tempted to apply this result to  $M = (\mathcal{R}[s])^{\ell+1}$ ,  $N = \mathcal{R}[s]$  and

$$(5.3) \quad A : (u_0(s), \dots, u_\ell(s)) \mapsto \sum_0^\ell u_i(s) a_i(s) : M \rightarrow N.$$

Unfortunately, this  $N$  is not finitely generated as an  $\mathcal{R}$ -module. Therefore, we choose instead

$$(5.4) \quad M := \mathcal{R}[z|m] \times (\mathcal{R}[z|n-1])^\ell, \quad N := \mathcal{R}[z|m+n]$$

where  $n := \deg a_0$ ,  $m := \max_{i \geq 1} \deg a_i$  and  $\mathcal{R}[z|k]$  denotes the  $\mathcal{R}$ -module of polynomials of degree  $\leq k$ . Obviously,  $N$  is finitely generated. Also it is easily seen that  $A$  (defined by (5.3)) maps  $M$  into  $N$ . We show that  $A : M \rightarrow N$  is surjective iff  $a_0, \dots, a_\ell$  are bezoutian. To this extent we prove that if  $v \in \mathcal{R}[z|m+n]$  can be represented as

$$(5.5) \quad v = u_0 a_0 + \sum_{i=1}^\ell u_i a_i$$

then such a representation can be chosen in such a way that  $\deg u_i \leq n-1$ , ( $i = 1, \dots, \ell$ ). In fact, if one of the  $u_i$ 's contains a term with a factor  $s^n$ , we replace this factor with  $a_0 - b$ , where  $b := a_0 - s^n \in \mathcal{R}[z|n-1]$ . Then we obtain a term with a factor  $a_0$ , which we combine with  $u_0 a_0$ , and a term with a factor  $b$ , which is of lower degree than the original factor  $s^n$ . Repetition of this procedure eventually leads to a representation of the form (5.5) with  $\deg u_i \leq n-1$  for  $i = 1, \dots, \ell$ .

Now, if  $a_0, \dots, a_\ell$  are bezoutian, every  $v \in \mathcal{R}[z]$ , in particular every



element of  $\mathcal{R}[z|m+n]$ , has a representation of the form (5.5), where by the foregoing reasoning, we may assume that  $\deg u_i \leq n-1$ . But if  $v \in \mathcal{R}[z|m+n]$ , it follows that

$$\deg u_0 = \deg u_0 a_0 - n \leq \max\{\deg v, \deg \sum_{i=1}^{\ell} u_i a_i\} - n \leq m.$$

Hence  $(u_0, u_1, \dots, u_{\ell}) \in M$ . Consequently,  $A$  is surjective. The converse is obvious.

Similarly,  $a_{0,\mu}, \dots, a_{\ell,\mu}$  are  $\mathcal{R}$ -bezoutian iff

$$A_{\mu} : (u_{0,\mu}, \dots, u_{\ell,\mu}) \mapsto \sum_{i=0}^{\ell} u_{i,\mu} a_{i,\mu} : M_{\mu} \rightarrow N_{\mu},$$

is surjective. Now we can apply [1, Ch II §3.3. Prop 11] . □

(5.6) REMARK. If  $\mathcal{R} = K[x_1, \dots, x_v]$ , where  $K$  is an algebraically closed field, it follows from Hilbert's Nullstellensatz that Lemma 5.2 remains valid even if none of the polynomials is monic. For general rings, however, this condition cannot be omitted. For instance, if  $\mathcal{R} = \mathbb{C}[[x,y]]$ , the ring of formal power series in two variables, the polynomials  $a_0(s) = 1 + xs$ ,  $a_1(s) = 1 + ys$  are easily seen not to be bezoutian, but  $a_{0,\mu} = a_{1,\mu} = 1$  where  $\mu$  is the (unique) maximal ideal, generated by  $x$  and  $y$ . The monicity condition can be relaxed as follows: The leading coefficients of the polynomials  $a_0, \dots, a_{\ell}$  generate  $\mathcal{R}$ . The proof is obvious. □

Now we are in the position to formulate the desired generalization of Theorem 2.4.

(5.7) THEOREM. Let  $A_i \in \mathcal{R}^{n \times m}$ ,  $B \in \mathcal{R}^{p \times p}$ ,  $q_i(s) \in \mathcal{R}[s]$  for  $i = 1, \dots, k$ .

The equation

$$\sum_{i=1}^k A_i X q_i(B) = C$$

has a solution  $X \in \mathcal{R}^{m \times p}$  for every  $C \in \mathcal{R}^{n \times p}$  if and only if  $a_0(s), a_1(s), \dots, a_\ell(s)$  have the bezoutian property. Here

$$a_0(s) := \det(sI - B)$$

and  $a_1(s), \dots, a_\ell(s)$  are the  $n \times n$  minors of

$$A(s) := \sum A_i q_i(s) .$$

PROOF. Consider the map

$$\mathcal{L} : X \mapsto \sum_{i=1}^k A_i X q_i(B) : \mathcal{R}^{m \times p} \rightarrow \mathcal{R}^{n \times p} .$$

We have to show that  $\mathcal{L}$  is surjective iff  $a_0, \dots, a_\ell$  are bezoutian. This equivalence was shown for  $\mathcal{R} = \mathbb{C}$  in Theorem 2.4 and it is easily seen that this proof extends immediately to the case where  $\mathcal{R}$  is any algebraically closed field. We proceed in two steps.

First assume that  $\mathcal{R}$  is an (arbitrary) field. Let  $K$  be an algebraically closed field containing  $\mathcal{R}$ . The surjectivity of a map as well as the bezoutian property of a set of polynomials is invariant under field extensions. (Recall that in a field the bezoutian property is equivalent to coprimeness.) Hence the general-field case is reduced to the algebraically-closed-field case.

Now let  $\mathcal{R}$  be arbitrary. Again we apply [1, Ch II, §3.3 Prop. 11] :  $\mathcal{L}$  is

surjective iff

$$L_{\mu} : X_{\mu} \mapsto \sum A_{i,\mu} X_{\mu} q_{i,\mu} (B_{\mu}) : \mathcal{R}_{\mu}^{m \times p} \rightarrow \mathcal{R}_{\mu}^{n \times p}$$

is surjective for every maximal ideal  $\mu$  of  $\mathcal{R}$ .

Because  $\mathcal{R}_{\mu}$  is a field,  $L_{\mu}$  is surjective iff  $a_{0,\mu}, \dots, a_{\ell,\mu}$  are bezoutian.

Here

$$a_{0,\mu}(s) = \det(sI_{\mu} - B_{\mu})$$

is the residue modulo  $\mu$  of  $a_0(s)$ , and similarly for the  $a_{i,\mu}(s)$ . Hence  $a_{0,\mu}(s), \dots, a_{\ell,\mu}(s)$  are bezoutian iff  $a_0(s), \dots, a_{\ell}(s)$  are bezoutian over  $\mathcal{R}[s]$ , according to Lemma 5.2. □

In the particular case where  $m = n$ , i.e., in the case of Theorem 1.4, the result can be simplified and formulated differently.

(5.8) COROLLARY. Let  $A_i \in \mathcal{R}^{m \times m}$ ,  $B \in \mathcal{R}^{p \times p}$  and  $q_i(s) \in \mathcal{R}[s]$ ,  $i = 1, \dots, k$ .

Then the following statements are equivalent:

- i) The equation  $\sum A_i X q_i(B) = C$  is universally solvable.
- ii)  $a(s) := \det A(s)$  and  $b(s) := \det(sI - B)$  are bezoutian. Here  
 $A(s) := \sum A_i q_i(s).$
- iii)  $a(B)$  is  $\mathcal{R}$ -invertible. Here  $a(s)$  is defined as in ii).

The proof of ii)  $\Leftrightarrow$  iii) can again be given via maximal ideals.

Alternatively:

ii)  $\Rightarrow$  iii) follows after the substitution  $s = B$  into  $u(s)a(s) + v(s)b(s) = 1$ , because of the Cayley-Hamilton theorem.

iii)  $\Rightarrow$  i) can be proved as in the proof of Theorem 1.4 (see the beginning of section 2).

Also Theorem 2.3 can be generalized to the ring case:

(5.9) THEOREM. Let  $A_i \in \mathcal{R}^{n \times m}$ ,  $B_i \in \mathcal{R}^{p \times p}$ ,  $B_i B_j = B_j B_i$  for  $i, j = 1, \dots, k$ .

The equation

$$\sum_{i=1}^k A_i X B_i = C$$

has a solution  $X \in \mathcal{R}^{m \times p}$  for every  $C \in \mathcal{R}^{n \times p}$  if and only if  $a_1(\underline{s}), \dots, a_\ell(\underline{s})$ ,  $b_1(\underline{s}), \dots, b_r(\underline{s})$  have the bezoutian property. Here  $a_1(\underline{s}), \dots, a_\ell(\underline{s})$  are the  $n \times n$  minors of

$$A(\underline{s}) := \sum A_i s_i,$$

(where  $\underline{s} = (s_1, \dots, s_k)$ , and  $b_1(\underline{s}), \dots, b_r(\underline{s})$  are the  $p \times p$  minors of

$$\begin{pmatrix} s_1 I - B_1 \\ \vdots \\ s_k I - B_k \end{pmatrix}.$$

The proof of this theorem is similar to the proof of Theorem (5.7),  
except that Lemma 5.2 is replaced by

(5.10) LEMMA. Let  $a_1(\underline{s}), \dots, a_h(\underline{s}) \in \mathcal{R}[\underline{s}] = \mathcal{R}[s_1, \dots, s_k]$  and assume that  
for  $i = 1, \dots, k$  there is a polynomial  $p_i$  amongst the  $a_j$ 's such that  $p_i$   
is only dependent on  $s_i$  (and not on the other indeterminates) and is monic  
with respect to this variable. Then  $a_1(\underline{s}), \dots, a_h(\underline{s})$  are bezoutian iff  
 $a_{1,\mu}(\underline{s}), \dots, a_{h,\mu}(\underline{s})$  are bezoutian in  $\mathcal{R}[\underline{s}]$  for each maximal ideal  $\mu$  of  $\mathcal{R}$ .

REFERENCES

- [1] N. BOURBAKI, Elements of Mathematics, Commutative Algebra, Hermann, Paris, 1972.
- [2] B. CHARLES, "Sur l'algèbre des opérateurs linéaires", J. de Mathématiques pures et appliquées, 33: 81-145 (1954).
- [3] G.S. DATUASHVILI, "On the spectrum of a generalized matrix polynomial" (in Russian), Bulletin of the Academy of Sciences of the Georgian SSR, 44 : 7-9, (1966).
- [4] F.R. GANTMACHER, The Theory of Matrices I, Chelsea, New York, 1960.
- [5] W.H. GUSTAFSON, "Roth's theorems over commutative rings", Linear Algebra and Applications, 23 : 245-251, (1979).
- [6] P. LANCASTER, "Explicit solutions of linear matrix equations", SIAM Review, 12 : 544-566, (1970).
- [7] M. NEWMAN, Integral Matrices, Acad. Press, New York, 1972.
- [8] C.C. MACDUFFEE, The Theory of Matrices, Chelsea, New York, 1960.
- [9] M. ROSEAU, Vibrations non linéaires et théorie de la stabilité, Springer, Berlin, 1966.
- [10] H.H. ROSENBROCK, State-space and Multivariable Theory, Wiley, New York, 1970.
- [11] W.E. ROTH, "The equations  $AX-YB = C$  and  $AX - XB = C$  in matrices", Proc. Amer. Math. Soc. 3 : 292-396, (1952).
- [12] L. WAELEBROECK, "Le calcul symbolique dans les algèbres commutatives", J. de Math. pures et appliquées, 33 : 147-186, (1954).