POLYNOMIAL EQUATIONS: A TOOL FOR CONTROL SYSTEMS SYNTHESIS

V. Kučera

Institute of Information Theory and Automation Academy of Sciences of the Czech Republic CZ-182 08 Praha

and

Trnka Laboratory for Automatic Control Faculty of Electrical Engineering Czech Technical University CZ-166 27 Praha, Czech Republic

E-mail:-kucera@utia.cas.cz

Abstract: Motivated by the pole placement design techniques, linear polynomial (or diophantine) equations have become a popular design tool of control engineers. The paper presents the theory, properties, and computational algorithms for the equation ax + by = c in a tutorial manner. Recent results are presented on the parametrization of limited degrees solutions, of relatively prime solutions, and of solutions x, y such that y/x is proper rational. Parametric and non-parametric computational algorithms are discussed in detail. Copyright © 1998 IFAC

Résumé: Motivées par les méthodes de placement des pôles, les équations linéaires polynomiales (ou Diophantine) sont devenue un outil mathématique polulaire parmi les automaticiens. Cette communication a pour but de présenter, d'une manière synoptique, la théorie et le calcul numérique liés à l'équation ax+by=c. Plusieurs résultats récents sont présentés, sur la paramétrisation des solutions x,y de degré borné, sur les solutions x,y premières entre elles et sur les solutions telles que le rapport y/x est rationnel et propre. Plusieurs méthodes de résolution de cette équation, paramétriques ou non, sont discutées en détail.

Keywords: Linear control systems; pole assignment; polynomial methods; parametrization; numerical solutions; software tools.

1. MOTIVATION

We consider the linear equation

$$ax + by = c \tag{1}$$

where a, b, and c are given polynomials from R[s], the ring of polynomials in the indeterminate s over the field R of real numbers, and x and y are unknown polynomials in R[s].

Equation (1) has found application in several design problems for linear control systems (Kučera, 1993), including the pole placement design. This problem consists in the following: given a plant with strictly proper rational transfer function

$$G(s) = \frac{b(s)}{a(s)}$$

where a and b are relatively prime polynomials, one seeks to determine a controller with a proper rational transfer function, say

$$K(s) = -\frac{y(s)}{x(s)},$$

which processes the plant output to produce the control input in such a way that the closed-loop system has prespecified poles.

The pole polynomial of the cosed-loop system, say c, which specifies the poles desired, is given by

$$c(s) = a(s)x(s) + b(s)y(s).$$

Thus the pole placement design is based on equation (1). However, not all solution pairs x, y are of interest. One must take the ones for which y/x is proper rational and, frequently, additional con-

straints are to be satisfied which reflect additional design specifications.

This motivates an investigation into the relevant properties of the linear polynomial equation (1).

2. REVIEW OF THEORY

It is well known (Bourbaki, 1961) that R[s] is a principal ideal domain. That is why (1) is solvable if and only if any greatest common divisor of a and b divides the polynomial c. Writing d for a greatest common divisor of a and b and denoting

$$\bar{a} = \frac{a}{d}, \quad \bar{b} = \frac{b}{d}, \quad \bar{c} = \frac{c}{d}$$

one concludes that (1) has a solution if and only if \bar{c} is a polynomial. As a result, if a and b are relatively prime, then (1) has a solution for any polynomial c.

Suppose that \bar{x}, \bar{y} is a particular solution pair of (1). Since the equation is linear, any and all solution pairs of (1) are given by

$$x = \bar{x} - \bar{b}t, \quad y = \bar{y} + \bar{a}t$$

where t varies over R[s]. Thus the solution class of (1) is parametrized through t in a simple manner.

It is well known (Bourbaki, 1961) that R[s] is a euclidean domain. Therefore if (1) is solvable and $a \neq 0$, then there is a unique solution pair $x_1, y_{1 \min}$ of (1) such that either $y_{1 \min} = 0$ or $\deg y_{1 \min} < \deg \bar{a}$. Further if (1) is solvable and $b \neq 0$, then there is a unique solution pair $x_{2 \min}, y_2$ of (1) such that either $x_{2 \min} = 0$ or $\deg x_{2 \min} < \deg \bar{b}$. These two least-degree solution pairs coincide (Kučera, 1979) whenever the polynomial c is of low enough degree, specifically if $\deg \bar{a} + \deg \bar{b} > \deg \bar{c}$.

It follows that equation (1) with $a \neq 0$ and $b \neq 0$ can possess solution pairs x, y of arbitrarily high degree, limited only from below by $\deg x_{2 \min}$ and $\deg y_{1 \min}$.

3. CONSTRAINED DEGREE SOLUTIONS

We shall study the class of solutions whose degrees are limited from *above* (Kučera, 1992). We suppose that a, b, and c are non-zero polynomials from R[s] and denote

$$p = \deg a, \quad q = \deg b, \quad r = \deg c.$$

Given non-negative integers m, n the task is to find any and all solution pairs x, y of (1) for which

$$x = 0$$
 or $\deg x \le m$,
 $y = 0$ or $\deg y \le n$. (2)

Theorem 1. Let a and b in (1) be relatively prime and let $m \ge q-1$ and $n \ge p-1$. If $m \ge r-p$

then the set of solutions x, y of (1) that satisfy (2) is given as

$$x = x_1 - bt_1, \quad y = y_1 \min + at_1$$
 (3)

where t_1 varies over R[s] and either $t_1 = 0$ or

$$\deg t_1 \leq \min(m-q, n-p);$$

if $n \ge r - q$ then the set of solutions x, y of (1) that satisfy (2) is given as

$$x = x_{2 \min} - bt_2, \quad y = y_2 + at_2$$
 (4)

where t_2 varies over R[s] and either $t_2 = 0$ or

$$\deg t_2 \leq \min(m-q, \ n-p).$$

Proof: We note that at least one of the two conditions, $m \ge r - p$ and $n \ge r - q$, must be satisfied, for otherwise the set (2) is empty. Indeed, (1) and (2) jointly imply that

$$\max (p + m, q + n) \ge$$

$$\max (p + \deg x, q + \deg y) \ge r.$$

Now suppose that $m \ge r - p$ holds and consider (3). If $t_1 = 0$ then $\deg x = r - p$ and either y = 0 or $\deg y \le p - 1$. Hence the least-degree solution $x_1, y_1 \min$ of (1) satisfies (2). If $t_1 \ne 0$ then

$$\deg x \le \max (r - p, q + \deg t_1)$$
$$\deg y = p + \deg t_1$$

and any other solution x, y of (1) will satisfy (2) if and only if t_1 verifies

$$q + \deg t_1 \le m, \ p + \deg t_1 \le n.$$

Now suppose that $n \ge r - q$ holds and consider (4). If $t_2 = 0$ then either x = 0 or $\deg x \le q - 1$ and $\deg y = r - q$. Hence the least-degree solution $x_{2\min}, y_2$ of (1) satisfies (2). If $t_2 \ne 0$ then

$$\deg x = q + \deg t_2$$
$$\deg y \le \max(r - q, p + \deg t_2)$$

and again (2) is satisfied if and only if

$$q + \deg t_2 \le m, \quad p + \deg t_2 \le n.$$

Several comments are in order.

Firstly, the assumptions on m and n are to guarantee that the set of solutions x, y of (1) satisfying (2) is not empty: at least one of the two least-degree solutions is in the set. As the degree of $y_{1 \min}$ can drop below p-1 (and that of $x_{2 \min}$ below q-1), the set (2) may be non-empty even for some lower bounds m and n. This is not a generic situation, however. To illustrate, consider the equation

$$s^2x(s) - (s^2 - 1)y(s) = 1.$$

Theorem 1 allows choosing m and n as low as 1. The only solution which satisfies the constraint $\deg x \le 1, \deg y \le 1$, happens to be constant,

$$x = 1, y = 1.$$

So one could have chosen m = n = 0 in this case.

Secondly, (3) can be used to parametrize the solution set (2) even if m < r-p. Then, however, t_1 has a higher degree than shown and is not completely free in R[s]. An analogous statement is true of (4) when n < r - q. To illustrate, we parametrize the solution class of

$$x(s) + sy(s) = s^2$$

such that $\deg x \le 1$ and $\deg y \le 1$. Since $n \ge r - q$ holds, one can use (4) to obtain the parametrization

$$x(s) = -st_2, \quad y(s) = s + t_2$$

where t_2 is any polynomial of degree zero, hence any constant. Now $m \ge r - p$ fails to hold; still, (3) can be used to obtain the parametrization

$$x(s) = s^2 - t_1, \quad y(s) = t_1$$

where $t_1 = s + \tau$, τ constant. The parameter t_1 however, is not a free polynomial.

The utility of Theorem 1 is in that the set of solutions of a constrained degree can be parametrized through a free polynomial of a constrained degree, when one starts with an appropriate least-degree solution: $x_1, y_{1 \min}$ if $m \ge r - q$ and $x_{2 \min}, y_2$ if $n \ge r - p$. Otherwise the parameter polynomial has preconditioned high-degree coefficients.

4. RELATIVELY PRIME SOLUTIONS

Consider the polynomial equation (1), in which the given polynomials a, b are relatively prime. Under what condition a solution pair x, y is also relatively prime?

If c=1 then (1) is a Bézout equation and every solution pair of (1) is relatively prime. For a non-unit c there exist solution pairs which are not relatively prime. We shall parametrize any and all such solutions of (1).

Theorem 2. Let a and b in (1) be relatively prime and let d be any divisor of c, $c = c_0 d$. Then there exists a solution pair x, y of (1) such that d divides both x and y and the set of all such pairs is given by

$$x = (x_0 - bt) d, \quad y = (y_0 + at) d$$
 (5)

where x_0, y_0 is a particular solution of the equation

$$ax_0 + by_0 = c_0 \tag{6}$$

and t ranges over R[s].

Proof: Given that a and b are relatively prime, equation (6) is solvable. If x_0, y_0 is a particular solution pair of (6), then all solution pairs are generated as $x_0 - bt$, $y_0 + at$ where t ranges over R[s]. Then the polynomials x, y given by (5) satisfy equation (1) as well as the divisibility condition.

On the other hand, all solution pairs x, y of (1) divisible by d are generated in this way. Indeed, suppose that $x = x'_0 d$, $y = y'_0 d$. Then

$$ax_0'd + by_0'd = c_0d$$

and x'_0, y'_0 is a solution pair of (6).

It is much easier to parametrize the solution pairs which are not relatively prime than those which are. In fact, every equation (1) with a non-unit polynomial c has solution pairs that are divisible by any divisor of c. To illustrate, consider the equation

$$x(s) + sy(s) = s^2 - 1.$$

Then all solutions divisible by s+1 are given by

$$x(s) = -(1 + st_1)(s + 1)$$

$$y(s) = (1 + t_1)(s + 1),$$
(7)

all solutions divisible by s-1 are given by

$$x(s) = (1 - st_2)(s - 1)$$

$$y(s) = (1 + t_2)(s - 1),$$
(8)

and all solutions divisible by $s^2 - 1$ are given by

$$x(s) = (1 - st_3)(s^2 - 1)$$

$$y(s) = t_3(s^2 - 1),$$
(9)

where t_1 , t_2 and t_3 independently range over R[s].

No claim is made that d is the greatest common divisor of x and y shown in (5). For example, the solution set (7) contains a pair which is divisible by s-1 (put $t_1=-s$). This pair is of course included in (8) as well as in (9): just set $t_2=-s-2$ and $t_3=-1$.

5. PROPER SOLUTIONS

Consider the polynomial equation (1) in which the ratio of the given polynomials b/a is a proper rational function. Under what condition a solution pair defines also a proper rational function y/x? This question is clearly motivated by the pole placement design.

We suppose that a, b, and c in (1) are non-zero polynomials from R[s] and denote

$$p = \deg a, \quad q = \deg b, \quad r = \deg c$$

as before. We shall investigate the existence of solution pairs x, y of (1) such that y/x is proper rational and parametrize the set of such "proper" solutions.

Theorem 3. Let a and b in (1) be relatively prime and suppose that

$$p > q \tag{10}$$

$$r \ge 2p - 1. \tag{11}$$

Then $y_{1 \min}/x_1$ is proper rational and the set of all solution pairs x, y of (1) such that y/x is proper rational is given by

$$x = x_1 - bt, \ y = y_{1 \min} + at$$
 (12)

where t varies over R[s] and either t = 0 or

$$\deg t \le r - 2p. \tag{13}$$

Proof: First observe that the least-degree solution x_1 , $y_1 \min$ results in a proper rational function $y_1 \min/x_1$. Indeed, either $y_1 \min = 0$ or $\deg y_1 \min < p$ by definition and $x_1 \neq 0$ with $\deg x_1 = r - p \geq p - 1$ by (11).

Further observe that any solution x,y of (1) such that y/x is proper rational has $\deg x = r - p$. For suppose that $\deg x > r - p$. Then (12) implies that $\deg t = \deg x - q$ and, consequently, $\deg y = \deg x - q + p > \deg x$ in view of (10). Hence y/x is not proper, a contradiction. And one cannot have $\deg x < r - p$, either. In this case (12) implies that $\deg t = r - p - q$ and, accordingly, $\deg y = r - q > r - p$ in view of (10). Therefore y/x is not proper, either

Finally observe that all pairs x, y in (12) such that $\deg y \leq r - p$ are generated either by t = 0 (if r = 2p - 1) or by any polynomial t whose degree does not exceed r - 2p. The degree of the corresponding polynomial x remains r - p, as $\deg x_1 > \deg bt$. Therefore, we have obtained the set of proper solutions to (1).

The assumption (10) requires that b/a is strictly proper rational while (11) requires that the polynomial c is of high enough degree. These two assumptions are to guarantee that a proper solution pair of (1) exists independently of the coefficients of a, b, and c. Of course, a proper solution x, y of (1) can exist even if (10) or (11) fails to hold. This is not a generic situation, however. To illustrate, find the class of proper solutions of the equation

$$s^2x(s) + y(s) = s^2 + 1.$$

Though (11) fails to hold, there exists a proper solution pair, namely

$$x = 1, y = 1.$$

The utility of Theorem 3 is twofold. Firstly it provides conditions, namely (10) and (11), under which a proper solution exists. These conditions

involve the degrees rather than the coefficients of a, b, and c. Secondly, the set of proper solutions is parametrized through a free polynomial of limited degree, bounded by (13). As a simple example, consider

$$sx(s) + y(s) = s^2 + 2s + 1.$$

The general solution set is given by

$$x(s) = s + 2 - t$$
, $y(s) = 1 + st$

and the proper solution set corresponds to t having degree at most zero, hence an arbitrary constant.

6. EXAMPLE

Let us provide a system-theoretic interpretation of Theorems 1,2,3 by invoking the pole placement design problem.

Consider a water tank with the inflow-to-level transfer function

$$G(s) = \frac{1}{s+1}.$$

We seek to determine a first-order feedback controller, K(s), such that the poles of the closed-loop system are well damped, say -2 and -3, and the remaining degrees of freedom are used to meet further design specifications.

Clearly

$$K(s) = -\frac{y(s)}{x(s)}$$

where x, y is the solution set of the equation

$$(s+1)x(s) + y(s) = (s+2)(s+3)$$

which renders K proper rational. Theorem 3 is applicable and yields the set

$$x(s) = s + 4 - t$$

$$y(s) = 2 + (s+1)t$$

where t is any polynomial of degree not exceeding zero, hence any real number. The same result can be obtained from Theorem 1 by fixing m = n = 1.

Following Theorem 2, the above solution set contains pairs x, y that are not relatively prime; namely they are divisible by s+3 and s+2. These are

$$x(s) = s + 3, \quad y(s) = s + 3$$

and

$$x(s) = s + 2, \quad y(s) = 2s + 4$$

and correspond to t = 1 and t = 2, respectively. For these parameter values, one of the closed-loop poles is a hidden mode of the controller.

Suppose that level disburbances are to be regulated without offset. Then a PI controller should be used; one can obtain it for t=4 as

$$K(s) = -4 - \frac{6}{s}.$$

If instead the H_{∞} norm of the sensitivity function

$$S(s) = \frac{(s+1)(s+4-t)}{(s+2)(s+3)}$$

is not to exceed 1, the parameter t should stay within the interval $4 - \sqrt{12} \le t \le 4 + \sqrt{12}$.

7. METHODS OF SOLUTION

Equation (1) can be solved in several ways. One can distinguish parametric methods (where the polynomials are represented by their coefficients) and non-parametric ones (where the polynomials are represented by their functional values.)

We suppose that a, b, and c in (1) are non-zero polynomials with a and b relatively prime. Consequently (1) is solvable. For the sake of simplicity let

$$\deg a = \deg b = n$$
, $\deg c = 2n - 1$.

We shall first describe three parametric methods.

Method of Indeterminate Coefficients (Kučera, 1994) converts equation (1) into a system of 2n linear equations over the field of real numbers. Suppose we seek the least-degree solution pair x, y:

$$\deg x \le n-1$$
, $\deg y \le n-1$.

The 2n coefficients of x and y satisfy the system of equations

$$\begin{bmatrix} x_0 \dots x_{n-1} & y_0 \dots y_{n-1} \end{bmatrix} \times \begin{bmatrix} a_0 & \cdots & a_n & & & \\ & \ddots & & \ddots & & \\ & & a_0 & \cdots & a_n & \\ & & b_0 & \cdots & b_n & & \\ & & \ddots & & \ddots & \\ & & & b_0 & \cdots & b_n \end{bmatrix}$$

$$= \begin{bmatrix} c_0 & \cdots & c_{n-1} \end{bmatrix}$$

The system matrix is a Sylvester matrix and it has full rank since a and b are relatively prime.

Method of Elementary Operations (Ježek, 1982) consists of applying elementary polynomial operations on $[a \ b \ -c]$ of the following four types (the dots stand for appropriate monomials)

$$\begin{bmatrix} 1 & 0 & 0 \\ \cdot & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & \cdot & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 & \cdot \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \cdot \\ 0 & 0 & 1 \end{bmatrix}$$

to successively reduce the degrees of a,b,c until we obtain $[1\ 0\ 0]$. Thus

$$[a \ b \ -c]U = [1 \ 0 \ 0]$$

where the unimodular matrix

$$U = \left[\begin{array}{ccc} u & -b & x \\ v & a & y \\ 0 & 0 & 1 \end{array} \right]$$

provides the solution pair x, y (along with a solution p, q of the Bézout equation ap + bq = 1).

Method of State-space Realization (Emre, 1980) combines matrix and polynomial operations. We write (1) as

$$x + \frac{b}{a}y = \frac{c}{a} \tag{14}$$

and determine a reachable state-space realization (F, G, H, J) of the rational function b/a. The n coefficients of y satisfy the system of equations

$$\begin{bmatrix} y_0 \dots y_{n-1} \end{bmatrix} \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} = \begin{bmatrix} c'_0 \dots c'_{n-1} \end{bmatrix}$$

where the right-hand side contains the coefficients of the polynomial c reduced modulo a. The corresponding polynomial x is then recovered form (14). The system matrix is an observability matrix and it has full rank since a and b are relatively prime.

We shall also describe one non-parametric method of solution, the *Method of Polynomial Interpolation* (Antsaklis and Gao, 1993). Any polynomial f(s) of degree n can be uniquely represented by the n+1 interpolation pairs s_i , $f(s_i)$, i=1,2,...,n+1 where s_i are distinct (real or complex) scalars. Thus equation (1) is equivalent to the system of 2n linear equations

$$\begin{bmatrix} x_0 \dots x_{n-1} & y_0 \dots y_{n-1} \end{bmatrix} \times \begin{bmatrix} a(s_1) & \cdots & a(s_{2n}) \\ \vdots & & \vdots \\ s_1^{n-1} a(s_1) & \cdots & s_{2n}^{n-1} a(s_{2n}) \\ b(s_1) & \cdots & b(s_{2n}) \\ \vdots & & \vdots \\ s_1^{n-1} b(s_1) & \cdots & s_{2n}^{n-1} b(s_{2n}) \end{bmatrix} = \begin{bmatrix} c(s_1) & \cdots & c(s_{2n}) \end{bmatrix}.$$

The system matrix is a Vandermonde matrix and it is non-singular since a, b are relatively prime and $s_1, ..., s_{2n}$ are distinct.

8. NUMERICAL EXPERIENCE

The comparison of the solution methods with respect to the arithmetic complexity is quite clear (Kučera, Ježek and Krupička, 1991). The fastest is the method of elementary operations, where the

operations count is proportional to n^2 . The arithmetic complexity of the remaining three methods is proportinal to n^3 . The method of indeterminate coefficients is slower than that of state—space realization because it leads to a larger system of equations and the method of interpolation is still slower because the system matrix must be set up first.

The comparison of the methods from the precision point of view (Šebek and Henrion, 1998) is not that simple, however. Provided the polynomials a and b have no (especially multiple) roots close to each other, the precision of all four methods is alike. The faster methods tend to be more reliable in this case. The ill-conditioned data, however, make the method of elementary operations fail more often than that of indeterminate coefficients. In between stay the methods of state—space realization and polynomial interpolation.

9. CONCLUSIONS

The paper has reviewed the classical results on the polynomial diophantine equation (1) and presented several new ones in Theorems 1,2, and 3. As a motivation for the study of the equation, the pole placement design problem for linear control systems has been considered.

Four representative methods of numerically solving the equation (1) have been presented. Numerical experience has been used to compare the methods and to provide a guideline for the user. This author recommends the method of indeterminate coefficients, based on its conceptual simplicity and low sensitivity to data.

A real challenge is to generalize the results of this paper for various matrix versions of equation (1), namely AX + BY = C and XA + YB = C. The solvability conditions as well as the parametrization of the solution sets is discussed in Kučera (1991). The parametrization of constrained degree solutions seems to be difficult to obtain in the general case. The special case of constant solutions was treated by Kučera and Zagalak (1991). The parametrizations of proper and relatively prime solutions to matrix polynomial versions of (1) are currently under investigation.

Finally the attention of the practical user is drawn to the Polynomial Toolbox (Kwakernaak and Šebek, 1997), a package of m-files for MATLABTM. The toolbox can be used to perform computations with polynomials and polynomial matries. In particular, it provides efficient solvers for many types of linear polynomial (matrix) equations. A version of the toolbox is available as freeware along with its hypertext manual and tutorial.

ACKNOWLEDGEMENT

Supported by the Ministry of Education of the Czech Republic under project VS97/034, by the Grant Agency of the Czech Republic under project 102/97/0861, and by the Copernicus project DY-COMANS CP 94-1246.

REFERENCES

- Antsaklis, P. and Z. Gao (1993). Polynomial and rational matrix interpolation: theory and control applications. *Int. J. Control*, 58, 349-404.
- Bourbaki, N. (1961). Algèbre Commutative. Hermann, Paris.
- Emre, E. (1980). The polynomial equation $QQ_c + RP_c = \Phi$ with application to dynamic feedback. SIAM J. Control Optimiz., 18, 611-620.
- Ježek, J. (1982). New algorithm for minimal solution of linear polynomial equations. Kybernetika, 18, 505-516.
- Kučera, V. (1979). Discrete Linear Control: The Polynomial Equation Approach. Wiley, Chichester.
- Kučera, V. (1991). Analysis and Design of Discrete Linear Control Systems. Prentice-Hall, London.
- Kučera, V. (1992). Fixed degree solutions of polynomial equations. In: System Structure and Control (V. Strejc, Ed.), 24-26 Pergamon, Oxford.
- Kučera, V. (1993). Diophantine equations in control - a survey. Automatica, 29, 1361-1375.
- Kučera, V. (1994). The pole placement equation a survey. Kybernetika, 30, 578-584.
- Kučera, V. and P. Zagalak (1991). Constant solutions of polynomial equations. Int. J. Control, 53, 495-502.
- Kučera, V., J. Ježek and M. Krupička (1991). Numerical analysis of diophantine equations. In: Advanced Methods in Adaptive Control for Industrial Applications (K. Warwick, M. Kárný and A. Halousková, Eds.), 128-136. Springer, Berlin.
- Kwakernaak, H. and M. Šebek (1997). Polynomial Toolbox. URL
 - http://www.math.utwente.nl/polbox
- Šebek, M. and D. Henrion (1998). Private communication.