Petko Hr. Petkov; Nicolai D. Christov; Michail M. Konstantinov
Numerically stable algorithm for pole assignment of linear single-input systems

*Kybernetika*, Vol. 22 (1986), No. 4, 349--367

Persistent URL: http://dml.cz/dmlcz/124366

KYBERNETIKA — VOLUME 22 (1986), NUMBER 4

# NUMERICALLY STABLE ALGORITHM FOR POLE ASSIGNMENT OF LINEAR SINGLE-INPUT SYSTEMS

PETKO HR. PETKOV, NIKOLAI D. CHRISTOV, MIHAIL M. KONSTATINOV

An efficient computational algorithm for pole assignment of linear single-input systems is presented. It is based on orthogonal reduction of the closed-loop system matrix to upper (quasi) triangular form whose $1 \times 1$ or $2 \times 2$ diagonal blocks correspond to the desired poles. A detailed numerical analysis of the algorithm is made which shows that it is unconditionally stable. The number of computational operations is approximately $6n^3$, the necessary array storage being $2n^2 + 6n$ working precision words, where $n$ is the order of the system.

## 1. INTRODUCTION

In recent years great attention has been paid to the development of reliable and efficient numerical methods for analysis and design of linear control systems. However one of the important problems of the synthesis of linear systems — the pole assignment problem is not solved yet satisfactorily from computational point of view [1]. Most of the existing methods for pole assignment are numerically unstable and computationally expensive. For example, the methods based on reduction of the system into phase-variable or Luenberger canonical form are unstable since the Frobenius form of a matrix can not be obtained by stable similarity transformations [2]. The methods using the characteristic polynomial of the open-loop system are also unsatisfactory due to the absence of a reliable method for finding the characteristic polynomial of a matrix. From similar disadvantages are suffering the methods, based on reduction to Jordan canonical form.

Recently three efficient numerical methods for pole assignment have been proposed in [3], [4] and [5]. These methods exploit the (quasi) triangular (Schur) form of the closed-loop system matrix, which is preferable from computational point of view since it may be obtained by orthogonal transformations only. Unfortunately it is not know under which conditions these methods are numerically stable and that is why the problem is still far from its final solution.

In this paper a new efficient computational algorithm for pole assignment of linear single-input systems is presented (a brief version of the algorithm was given in [6]). It is based also on orthogonal reduction of the closed-loop system matrix to upper (quasi) triangular form whose $1 \times 1$ or $2 \times 2$ diagonal blocks correspond to the desired poles. The main feature of this algorithm is that it is unconditionally stable which makes it applicable to ill-conditioned and high order problems. The numerical stability is considered in the sense that the upper (quasi) triangular form obtained has the desired eigenvalues on its diagonal, and it is the exact (quasi) triangular form of a matrix which is close to the closed-loop system matrix. This quarantees that the gain matrix obtained is true for a problem near to the given one. Note however that this does not ensure small errors in the computed gain matrix and closeness of the eigenvalues of the closed-loop system matrix to the desired poles. The errors in the gain matrix obtained depend on the conditioning of the pole assignment problem which is not well studied and is still an open question [3].

The paper is organized in the following way. This section is concluded with some notations used throughout the paper. A brief statement of the problem is given in Section 2. Section 3 contains the main result – the determination of the gain matrix by orthogonal reduction of the closed-loop system matrix into (quasi) triangular form. The numerical properties of the algorithm are considered in Section 4. An example is given in Section 5 and some conclusions are made in Section 6.

Notations will be as follows. We use upper case for matrices and lower case for vectors and scalars. $\mathbb{R}^{n \times m}$ is the space of $n \times m$ real matrices ($\mathbb{R}^{n \times 1} = \mathbb{R}^n$), $A^{\mathrm{T}}$ is the transposed matrix $A$, $\mathbb{O}(n) \subset \mathbb{R}^{n \times n}$ is the group of orthogonal matrices, $\|\cdot\|$ denotes the Euclidean norm of a vector or a matrix.

## 2. PROBLEM STATEMENT

Consider the completely controllable time-invariant single-input linear system

$$\dot{x}(t) = Ax(t) + bu(t)\,, \tag{1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^1$ and $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$. The system (1) will be identified with the matrix pair $(A, b) \in \mathbb{LS}(n)$, where $\mathbb{LS}(n) \subset \mathbb{R}^{n \times n} \times \mathbb{R}^n$ is the set of matrix pairs $(A, b)$ with $A$ cyclic and $b$ a generator for $\mathbb{R}^n$ relative to $A$.

It is necessary to find a gain matrix $k \in \mathbb{R}^{1 \times n}$ such that the control law $u(t) = -kx(t)$ preassigns the spectrum $\operatorname{spect}(A_c)$ of the closed-loop system matrix $A_c = A - bk$: $\operatorname{spect}(A_c) = s$, where $s = (s_1, \ldots, s_n)$ is a given set of $n$ pair-wise complex conjugate numbers.

It is well known that for each $s$ there exists a solution to the pole assignment problem iff the system is completely controllable [7]. In the case of a single-input system this solution is unique.

In every algorithm for pole assignment the gain matrix $k$ is obtained as a solution of a system of linear algebraic equations. The construction of this system depends on the method used and may in turn deteriorate the solution. The algorithm described in the following section avoids the difficulties related to the solution of the pole assignment problem due to the construction and solving the equations for the elements of the gain matrix.

A preliminary step of the algorithm proposed is the reduction of the pair $(A, b)$ to the orthogonal canonical form $(\tilde{A}, \tilde{b}) = (P^T A P, P^T b)$, $P \in \mathbb{O}(n)$, where

$$\tilde{A} = \begin{bmatrix} & a_1 & \\ a_{21} & a_2 & \\ & \ddots & \\ 0 & a_{n,n-1} & a_n \end{bmatrix}, \quad \tilde{b} = \begin{bmatrix} b_{10} \\ \\ 0 \\ \end{bmatrix}, \quad a_i^T \in \mathbb{R}^{n-i+1} .$$

In view of the complete controllability of $(A, b)$ we have $b_{10} \neq 0$, $a_{i,i-1} \neq 0$, $i = 2, \ldots, n$.

Algorithms for obtaining the form $(\tilde{A}, \tilde{b})$ based on singular value decomposition have been proposed in [8]–[10]. This form may be obtained efficiently using $n - 1$ Householder reflections [11], [12] similarly to the reduction of a general matrix to its Hessenberg form [2]. The corresponding algorithm is numerically stable: it can be shown that the computed orthogonal canonical form $(\tilde{A}, \tilde{b})$ is exact for a pair $(A + \delta A, b + \delta b)$ which is close to $(A, b)$ –

$$\begin{aligned} \|\delta A\| &\leqq \text{eps}\,(6n^2 + \text{const. } n)\,\|A\| , \\ \|\delta b\| &\leqq \text{eps}\,(6n + \text{const})\,\|b\| , \end{aligned} \tag{2}$$

where eps is the relative machine precision of the computer used.

The number of the floating point operations (FLOPS) for this algorithm is approximately $5n^3/3$ (1 FLOP $\approx$ 1 addition + 1 multiplication), the necessary array storage being $2n^2 + 2n$ working precision words.

## 3. DETERMINATION OF THE GAIN MATRIX

The pole assignment algorithm presented in this section is based on the following idea. Since the open-loop and closed-loop system matrices $A$ and $A_c$ are in Hessenberg form and differ only in their first rows it is possible, by setting a desired pole, to find an eigenvector of the matrix $A_c$ before computing $k$. Using sequences of plane rotations, belonging to the group of orthogonal transformations, all but the first elements of the eigenvector may be annihilated. Then by necessity the first column of the transformed matrix $A_c$ will have zero elements below the first one which must be equal to the desired eigenvalue. This gives an equation for the first element of the transformed gain matrix. The key observation here is that after the transformation the matrices $A$ and $A_c$ remain in Hessenberg form which permits to work by the same way at the next step. At each step the algorithm works with a subsystem of decreasing order and the plane rotations are determined by the subsystem eigenvector. Since the subsystem matrices are in Hessenberg form their eigenvectors may be computed by solving triangular systems of linear equations.

Let the pair $(A, b)$ is preliminary reduced to the orthogonal canonical form $(\tilde{A}, \tilde{b})$ and let the set of desired poles be

$$s = (s_1, \ldots, s_r, p_1 + \mathrm{i}q_1, p_1 - \mathrm{i}q_1, \ldots, p_m - \mathrm{i}q_m)\,; \quad \mathrm{i}^2 = -1\,,$$
$$m = (n - r)/2\,.$$

The $n$ elements of the gain matrix $k$ can be computed by the following algorithm.

*Step* 1. The eigenvector of $A_c$ corresponding to $s_1$ is $v_1 = P\tilde{v}_1$, where

$$\tilde{A}_c\tilde{v}_1 = \tilde{v}_1 s_1\,, \tag{3}$$
$$\tilde{A}_c = \tilde{A} - \tilde{b}\tilde{k}\,, \quad \tilde{k} = kP\,.$$

The matrices $\tilde{A}$, $\tilde{A}_c$ are in Hessenberg form with non-zero subdiagonal elements, and differ only in their first rows. That is why the eigenvector $\tilde{v}_1$ may be determined from

$$T\bar{v}_1 = \bar{v}_1 s_1 - h_1\,, \tag{4}$$

where

$$\tilde{v}_1 = [v_{11}, v_{21}, \ldots, v_{n1}]^{\mathrm{T}} = \begin{bmatrix} \bar{v}_1 \\ \hline 1 \end{bmatrix} = \begin{bmatrix} \times \\ \hline \bar{\bar{v}}_1 \end{bmatrix}; \quad \bar{v}_1, \bar{\bar{v}}_1 \in \mathbb{R}^{n-1} \tag{5}$$

and the matrix $\tilde{A}$ is partitioned as

$$\tilde{A} = \left[\begin{array}{c|c} \times \;\; \times \;\; \ldots & \times \\ \hline T_1 & h_1 \end{array}\right]$$

with $T_1 \in \mathbb{R}^{(n-1)\times(n-1)}$ being a non-singular upper triangular matrix ($[\times]$ denotes non-referenced elements). Note that in view of the inequalities $a_{i,i-1} \neq 0$, $i = 2, \ldots, n$ the element $v_{n1}$ must be non-zero and hence is chosen equal to 1 in (5).

The linear triangular system of equations (4) may be solved by back substitution. However the elements of the eigenvector may be computed simultaneously with the transformation of this eigenvector exploiting the fact that some of the previous elements are already annihilated. This reduces the number of the computational operations and improves the accuracy of the eigenvector. That is why the following strategy is proposed.

*Step* 1.1: Compute the eigenvector elements $v_{n-1,1}$ and (if $n > 2$) $v_{n-2,1}$ from

$$v_{n-1,1} = (s_1 - a_{nn})\, v_{n1}/a_{n,n-1}\,,$$
$$v_{n-2,1} = ((s_1 - a_{n-1,n-1})\, v_{n-1,1} - a_{n-1,n}v_{n1})/a_{n-1,n-2}\,,$$

where $a_{ij}$ are the corresponding elements of $\tilde{A}$.

Construct a plane rotation

$$R_1 = \operatorname{diag}\left(I_{n-2}, \begin{bmatrix} c_1 & d_1 \\ -d_1 & c_1 \end{bmatrix}\right) \in \mathbb{O}(n)$$

in the $(n-1, n)$-plane such that $R_1\tilde{v}_1 = v_1^1$, where

$$v_1^1 = [\times, \ldots, \times, v_{n-2,1}, \tilde{v}_{n-1,1}, 0]^{\mathrm{T}}.$$

The numbers $c_1$ and $d_1$ are determined by $c_1 = v_{n-1,1}/z$, $d_1 = v_{n1}/z$, where $z = (v_{n-1,1}^2 + v_{n1}^2)^{1/2}$, $c_1^2 + d_1^2 = 1$, $\tilde{v}_{n-1,1} = z$.

Since $v_{n1} = 1$ this implies $\tilde{v}_{n-1,1} \geqq 1$. It follows from (3)

$$R_1\tilde{A}_c R_1^{\mathrm{T}} v_1^1 = v_1^1 s_1 . \tag{6}$$

The form of the matrices $R_1\tilde{A}R_1^{\mathrm{T}}$ and $R_1\tilde{A}_c R_1^{\mathrm{T}}$ is illustrated by the following 4th order example

$$\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}.$$

If $n > 2$ this transformation does not affect the vector $\tilde{b}$.

*Step* 1.2: If $n > 3$ compute the eigenvector element $v_{n-3,1}$ from

$$v_{n-3,1} = \left((s_1 - a_{n-2,n-2})\, v_{n-2,1} - a_{n-2,n-1}\, \tilde{v}_{n-1,1}\right)/a_{n-2,n-3} ,$$

where $a_{ij}$ are the $ij$ elements of $R_1\tilde{A}_c R_1^{\mathrm{T}}$ and hence of $R_1\tilde{A}R_1^{\mathrm{T}}$.

Construct a plane rotation $R_2 \in \mathcal{O}(n)$ in $(n-2, n-1)$-plane which annihilates the $(n-1)$ element of $v_1^1$, i.e. $R_2 v_1^1 = v_1^2$,

$$v_1^2 = [\times, \ldots, \times, v_{n-3,1}, \tilde{v}_{n-2,1}, 0, 0]^{\mathrm{T}}.$$

Since $\tilde{v}_{n-1,1} \geqq 1$ and the norm of $[v_{n-2,1}, \tilde{v}_{n-1,1}]^{\mathrm{T}}$ is preserved, it follows that $\tilde{v}_{n-2,1} \geqq 1$. From (6)

$$R_2 R_1 \tilde{A}_c R_1^{\mathrm{T}} R_2^{\mathrm{T}} v_1^2 = v_1^2 s_1 . \tag{7}$$

Returning to the 4th order example it may be observed that matrices $R_2R_1\tilde{A}_cR_1^{\mathrm{T}}R_2^{\mathrm{T}}$ and $R_2R_1\tilde{A}R_1^{\mathrm{T}}R_2^{\mathrm{T}}$ have the form

$$\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \otimes & \times & \times \end{bmatrix}.$$

Now it is easily verified that the $(n, n-2)$ – (encircled) elements of the above matrices must be zero. In fact the last equation in (7) yields $a_{n,n-2}\tilde{v}_{n-2,1} = 0$, where $a_{n,n-2}$ is the $(n, n-2)$-element of $R_2R_1\tilde{A}_cR_1^{\mathrm{T}}R_2^{\mathrm{T}}$, and since $\tilde{v}_{n-2,1} \neq 0$ this implies $a_{n,n-2} = 0$. The $(n, n-2)$ – element of $R_2R_1\tilde{A}R_1^{\mathrm{T}}R_2^{\mathrm{T}}$ also must be zero because it is not affected by the gain matrix.

*Step* 1.3: Using similar technique to annihilate the $(n-2)$-element of $v_1^2$ one obtains the matrices $R_3R_2R_1\tilde{A}_cR_1^{\mathrm{T}}R_2^{\mathrm{T}}R_3^{\mathrm{T}}$ and $R_3R_2R_1\tilde{A}R_1^{\mathrm{T}}R_2^{\mathrm{T}}R_3^{\mathrm{T}}$ whose $(n-1, n-3)$-elements are zero. For the example of 4th order this means that after Step 1.3 the matrix $R_3R_2R_1\tilde{A}R_1^{\mathrm{T}}R_2^{\mathrm{T}}R_3^{\mathrm{T}}$ is again in Hessenberg form.

*Step* 1. $(n-1)$: As a result of computations similar to those in 1.1, ..., 1 . $(n-2)$ one obtains the equation

$$Q_1^{\mathrm{T}}\tilde{A}_c Q_1 v_1^{n-1} = v_1^{n-1} s_1 \tag{8}$$

where $Q_1 = R_1^{\mathrm{T}} R_2^{\mathrm{T}} \dots R_{n-1}^{\mathrm{T}}, v_1^{n-1} = [\tilde{v}_{11}, 0, \dots, 0]^{\mathrm{T}}$, $\tilde{v}_{11} \geqq 1$. The matrix $Q_1^{\mathrm{T}}\tilde{A}Q_1$ may be represented as

$$Q_1^{\mathrm{T}}\tilde{A}Q_1 = \left[\begin{array}{c|cccc} a_{11} & \times & \times & \dots & \times \\ \hline a_{21} & & & & \\ \hline 0 & & \tilde{A}^{(2)} & & \end{array}\right],$$

where $\tilde{A}^{(2)} \in \mathbb{R}^{(n-1)\times(n-1)}$ is in Hessenberg form.

At this step the matrix $\tilde{b}$ is reduced to

$$Q_1^{\mathrm{T}}\tilde{b} = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \left[\begin{array}{c} \tilde{b}_1 \\ \hline \tilde{b}^{(2)} \end{array}\right].$$

With regard to (8) the transformed closed-loop system matrix is to be in the form

$$Q_1^{\mathrm{T}}\tilde{A}_c Q_1 = \left[\begin{array}{c|cccc} s_1 & \times & \times & \dots & \times \\ \hline 0 & & \tilde{A}_c^{(2)} & & \end{array}\right], \tag{9}$$

where $\tilde{A}_c^{(2)} \in \mathbb{R}^{(n-1)\times(n-1)}$ is a Hessenberg matrix. Since the closed-loop system is also completely controllable it follows that the element $\tilde{b}_2$ must be non-zero. Thus the matrices $\tilde{A}^{(2)}$ and $\tilde{A}_c^{(2)}$ differ in their first rows only.

The relation (9) yields

$$\tilde{b}_1 k_1 = a_{11} - s_1\,, \tag{10}$$

$$\tilde{b}_2 k_1 = a_{21}\,, \tag{11}$$

where $k_1$ is the first element of the row vector $\tilde{k}Q_1$. The equations (10) and (11) are algebraically consistent but in some cases (10) may be a zero identity. That is why it is reasonable to determine $k_1$ from

$$k_1 = (a_{11} - s_1)/\tilde{b}_1\,, \quad \text{if} \quad |\tilde{b}_1| \geqq |\tilde{b}_2|\,,$$
$$k_1 = a_{21}/\tilde{b}_2\,, \qquad\qquad \text{if} \quad |\tilde{b}_1| < |\tilde{b}_2|\,.$$

In this way as a result of Step 1 one element of the transformed gain matrix is obtained and the problem is reduced to a problem of dimension $n-1$. Since the matrices $\tilde{A}^{(2)}$ and $\tilde{A}_c^{(2)}$ of the $(n-1)$-order subsystem are in Hessenberg form it is possible to proceed further in the same way.

*Steps* 2, ..., $r$. The next $r-1$ elements of the gain matrix are determined. Every eigenvector is obtained as a solution of a three-diagonal system of linear equations and the number of necessary plane rotations decreases with 1 at each step. Note that the column transformations are to be performed on the whole $n \times n$ matrix. However they do not affect the elements of the gain matrix already computed.

Denoting the transformations at Steps 2, ..., $r$ with $Q_2, \ldots, Q_r \in \mathbb{O}(n)$ one obtains the matrices

$$Q_r^{\mathrm{T}} \ldots Q_1^{\mathrm{T}} \tilde{A} Q_1 \ldots Q_r = \begin{bmatrix} \times & \times & \ldots & & \ldots & \times \\ \times & \times & \ldots & & \ldots & \times \\ & \ddots & \ddots & & & \vdots \\ & 0 & \ddots & \times & \times \ldots & \times \\ & & & \times & \multicolumn{2}{|c}{\tilde{A}^{(r+1)}} \end{bmatrix},$$

$$Q_r^{\mathrm{T}} \ldots Q_1^{\mathrm{T}} \tilde{b} = [\underbrace{\times, \ldots, \times}_{r}, \tilde{b}_{r+1}, 0, \ldots, 0]^{\mathrm{T}},$$

where $\tilde{A}^{(r+1)} \in \mathbb{R}^{(n-r) \times (n-r)}$ is a Hessenberg matrix.

The closed-loop system matrix has the form

$$Q_r^{\mathrm{T}} \ldots Q_1^{\mathrm{T}} \tilde{A}_c Q_1 \ldots Q_r = \begin{bmatrix} s_1 & \times & \times & \ldots & \ldots & \times \\ & s_2 & \times & \ldots & \ldots & \times \\ & & \ddots & & & \\ & & & \ddots & & \\ & 0 & & & s_r & \times \ldots \times \\ & & & & & \tilde{A}_c^{(r+1)} \end{bmatrix}$$

and $\tilde{A}_c^{(r+1)} \in \mathbb{R}^{(n-r) \times (n-r)}$ is also a Hessenberg matrix. In view of the complete controllability of the system the subdiagonal elements of the matrices $\tilde{A}^{(r+1)}$, $\tilde{A}_c^{(r+1)}$ and the element $\tilde{b}_{r+1}$ must be non-zero.

It is clear that using complex plane rotations the above technique may also be applied to determine the elements of the gain matrix in the case of complex conjugate poles. However it is possible to solve the problem with slightly complicated technique using real arithmetic only. As a result the transformed closed-loop system matrix will have $2 \times 2$ blocks on its diagonal. This technique is described in the following double step.

*Steps* $(r + 1)$, $(r + 2)$. The computation of the real $x_1$ and the imaginary $y_1$ parts of the complex eigenvectors $x_1 + iy_1$, $x_1 - iy_1$ of the matrix $\tilde{A}_c^{(r+1)}$, corresponding to the poles $p_1 + iq_1$, $p_1 - iq_1$, may be performed by the equations

$$T_{r+1}[\bar{x}_1 \mid \bar{y}_1] = [\bar{\bar{x}}_1 \mid \bar{\bar{y}}_1] \, S_1 - [h_{r+1} \mid h_{r+1}], \tag{12}$$

where

$$x_1 = [x_{r+1}, \ldots, x_n]^{\mathrm{T}} = \begin{bmatrix} \bar{x}_1 \\ \hline 1 \end{bmatrix} = \begin{bmatrix} \times \\ \hline \bar{\bar{x}}_1 \end{bmatrix} \in \mathbb{R}^{n-r},$$

$$y_1 = [y_{r+1}, \ldots, y_n]^{\mathrm{T}} = \begin{bmatrix} \bar{y}_1 \\ \hline 1 \end{bmatrix} = \begin{bmatrix} \times \\ \hline \bar{\bar{y}}_1 \end{bmatrix} \in \mathbb{R}^{n-r},$$

$$S_1 = \begin{bmatrix} p_1 & q_1 \\ -q_1 & p_1 \end{bmatrix} \tag{13}$$

and the matrix $\tilde{A}^{(r+1)}$ is partitioned as

$$\tilde{A}^{(r+1)} = \left[\begin{array}{c|c} \times \quad \times \quad \dots & \dots \quad \times \\ \hline T_{r+1} & h_{r+1} \end{array}\right]$$

with $T_{r+1} \in \mathbb{R}^{(n-r-1)\times(n-r-1)}$ being non-singular upper triangular matrix.

The examination of equation (12) shows that in the case of small imaginary parts of the poles the vectors $x_1$ and $y_1$ will tend to be linearly dependent which will deteriorate the solution. For this reason equation (12) must be modified by taking

$$S_1 = \begin{bmatrix} p_1 & 1 \\ -q_1^2 & p_1 \end{bmatrix} \tag{14}$$

instead of (13). The matrices (13) and (14) have the same eigenvalues; however using (14) the vectors $x_1$ and $y_1$ will be linearly independent even if $q_1 = 0$. In this case $y_1$ will be determined as a generalized eigenvector. This simple device permits to avoid difficulties when the complex conjugate poles are close to the real axis.

Now two elements of the transformed gain matrix may be determined simultaneously applying plane rotation to annihilate appropriate elements of $x_1$ and $y_1$.

Similarly to the real case the element of the eigenvector may be computed simultaneously with the annihilation of previous elements thus reducing the number of the necessary operations.

*Step* $(r+1).1$: Compute the elements $x_{n-1,1}$, $y_{n-1,1}$ and (if $n > 2$) $x_{n-2,1}$, $y_{n-2,1}$ from

$$x_{n-1,1} = ((p_1 - a_{nn})\, x_{n1} - q_1^2 y_{n1})/a_{n,n-1}\,,$$

$$y_{n-1,1} = (x_{n1} + (p_1 - a_{nn})\, y_{n1})/a_{n,n-1}\,,$$

$$x_{n-2,1} = ((p_1 - a_{n-1,n-1})\, x_{n-1,1} - a_{n-1,n}x_{n1} - q_1^2 y_{n-1,1})/a_{n-1,n-2}\,,$$

$$y_{n-2,1} = (x_{n-1,1} + (p_1 - a_{n-1,n-1})\, y_{n-1,1} - a_{n-1,n}y_{n1})/a_{n-1,n-2}\,.$$

Construct a plane rotation $U_1$ in the $(n-1, n)$-plane such that $U_1 x_1 = x_1^1$, where $x_1^1 = [\times, \dots, \times, x_{n-2}, \tilde{x}_{n-1,1}, 0]^{\mathrm{T}}$. This transformation must be applied on the vector $y_1$. Let for example $\tilde{A}_c^{(r+1)}$ be a $4 \times 4$ matrix. Then $U_1 \tilde{A}_c^{(r+1)} U_1^{\mathrm{T}}$ will have the form

$$\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}.$$

*Step* $(r+1).2$: If $n > 3$ compute the elements $x_{n-3,1}$, $y_{n-3,1}$. Construct a plane rotation $U_2$ such that $U_2 x_1^1 = x_1^2$, where $x_1^2 = [\times, \dots, \times, x_{n-3,1}, \tilde{x}_{n-2,1}, 0, 0]^{\mathrm{T}}$. This transformation must be also applied on $y_1$.

After this step the matrix $U_2U_1\tilde{A}_c^{(r+1)}U_1^TU_2^T$ in the 4th order example takes the form

$$\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}.$$

*Step* $(r+2)$. 1: Construct a plane rotation $V_1$ in $(n-1, n)$-plane to annihilate the element $y_{n1}$, i.e. $V_1y_1 = y_1^1$, where $y_1^1 = [\times, \ldots, \times, \tilde{y}_{n-1,1}, 0]^T$. This transformation does not affect the transformed vector $x_1$. The matrix $V_1U_2U_1\tilde{A}_c^{(r+1)}U_1^T$ . $U_2^TV_1^T$ will have the form

$$\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix}.$$

*Step* $(r+1)$ . 3: Construct a plane rotation $U_3 \in \mathcal{O}(n-r)$ to annihilate the element $\tilde{x}_{n-2,1}$ and apply this transformation to $y_1$. This will not destroy the form of $y_1$. For our example the matrix $U_3V_1U_2U_1\tilde{A}_c^{(r+1)}U_1^TU_2^TV_1^TU_3^T$ will look as

$$\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \otimes & \times & \times & \times \end{bmatrix}.$$

where the (encircled) $n$ 1-element must be zero.

*Step* $(r+2)$ . 2: Compute a plane rotation $V_2$ to annihilate the element $\tilde{y}_{n-1,1}$. Then the matrix

$$V_2U_3V_1U_2U_1\tilde{A}_c^{(r+1)}U_1^TU_2^TV_1^TU_3^TV_2^T$$

will take the form

$$\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \otimes & \times & \times & \times \\ 0 & \otimes & \times & \times \end{bmatrix},$$

where the (encircled) elements in positions $(n, n-2)$ and $(n-1, n-3)$ must be zero. Thus the transformed open-loop and closed-loop system matrices will be again in Hessenberg form.

This process may be continued until the elements $\tilde{x}_{r+2}$ and $\tilde{y}_{r+3}$ are annihilated. The strategy of annihilation may be illustrated as

$$\begin{bmatrix} \otimes & \otimes \\ \otimes & \otimes \\ \otimes & \otimes \\ \otimes & \otimes \end{bmatrix}$$

and it preserves the Hessenberg form of the system matrices.

The above process clearly may be considered as $QR$-decomposition of the vectors $x_1, y_1$. Since these vectors are linearly independent one obtains

$$(15)\qquad W^{\mathrm{T}}_{r+1,r+2}\tilde{A}_c^{(r+1)}W_{r+1,r+2}\begin{bmatrix}\tilde{x}_{r+1} & \tilde{y}_{r+1}\\ 0 & \tilde{y}_{r+2}\\ 0 & 0\\ \vdots & \vdots\\ 0 & 0\end{bmatrix} = \begin{bmatrix}\tilde{x}_{r+1} & \tilde{y}_{r+1}\\ 0 & \tilde{y}_{r+2}\\ 0 & 0\\ \vdots & \vdots\\ 0 & 0\end{bmatrix} S_1$$

where $W_{r+1,r+2} \in \mathbb{O}(n-r)$ are the orthogonal transformations accumulated at the double step $(r+1)$, $(r+2)$. Note that $\tilde{x}_{r+1}, \tilde{y}_{r+2} \geqq 1$. The equation (15) may be written as

$$(16)\qquad W^{\mathrm{T}}_{r+1,r+2}\tilde{A}_c^{(r+1)}W_{r+1,r+2}\begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\\ \vdots & \vdots\\ 0 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\\ \vdots & \vdots\\ 0 & 0\end{bmatrix} \hat{S}_1 ,$$

where

$$(17)\qquad \hat{S}_1 = \begin{bmatrix}\tilde{x}_{r+1} & \tilde{y}_{r+1}\\ 0 & \tilde{y}_{r+2}\end{bmatrix}\begin{bmatrix}p_1 & 1\\ -q_1^2 & p_1\end{bmatrix}\begin{bmatrix}\tilde{x}_{r+1} & \tilde{y}_{r+1}\\ 0 & \tilde{y}_{r+2}\end{bmatrix}^{-1} =$$

$$= \begin{bmatrix}p_1 - q_1^2\tilde{y}_{r+1}/\tilde{x}_{r+1} & \tilde{x}_{r+1}/\tilde{y}_{r+2} + q_1^2\tilde{y}_{r+1}^2/(\tilde{x}_{r+1}\tilde{y}_{r+2})\\ -q_1^2\tilde{y}_{r+2}/\tilde{x}_{r+1} & p_1 + q_1^2\tilde{y}_{r+1}/\tilde{x}_{r+1}\end{bmatrix}.$$

Denoting $Q_{r+1,r+2} = \operatorname{diag}(I_r, W_{r+1,r+2}) \in \mathbb{O}(n)$ it follows from (16) that the transformed closed-loop system matrix is to be in the form

$$(18)\qquad Q^{\mathrm{T}}_{r+1,r+2}\ldots Q_1^{\mathrm{T}}\tilde{A}_cQ_1\ldots Q_{r+1,r+2} =$$

$$= \begin{bmatrix} s_1 & \times & \times & \cdots & & & \cdots & \times\\ & s_2 & \times & \cdots & & & \cdots & \times\\ & & \ddots & & & & & \\ & & & s_r & \times & \times & \cdots & \times\\ & \mathbf{0} & & & \hat{S}_1 & & \times \cdots & \times\\ & & & & & & \times \cdots & \times\\ & & & & & & \tilde{A}_c^{(r+3)} & \end{bmatrix},$$

where $\tilde{A}_c^{(r+3)}$ is a Hessenberg matrix. The vector $[\tilde{b}_{r+1}, 0, \ldots, 0]^{\mathrm{T}} \in \mathbb{R}^{n-r}$ is reduced to $[\times, \tilde{b}_{r+2}, \tilde{b}_{r+3}, 0, \ldots, 0]^{\mathrm{T}}$ and the complete controllability ensures that $\tilde{b}_{r+3} \neq 0$.

Now the equation (18) may be used to determine the elements $k_{r+1}, k_{r+2}$ of the transformed gain matrix $\tilde{k}Q_1 \ldots Q_{r+1,r+2}$. As a result one obtains

$$(19)\qquad \begin{aligned}\tilde{b}_{r+2}k_{r+1} &= a_{r+2,r+1} + q_1^2\tilde{y}_{r+2}/\tilde{x}_{r+1} ,\\ \tilde{b}_{r+2}k_{r+2} &= a_{r+2,r+2} - p_1 - q_1^2\tilde{y}_{r+1}/\tilde{x}_{r+1}\end{aligned}$$

and

$$\tilde{b}_{r+3}k_{r+1} = a_{r+3,r+1}\,, \tag{20}$$
$$\tilde{b}_{r+3}k_{r+2} = a_{r+3,r+2}\,,$$

where $a_{r+i,r+j}$; $i = 2, 3, j = 1, 2$ are the corresponding elements of the transformed open-loop system matrix

$$Q^{\mathrm{T}}_{r+1,r+2} \ldots Q^{\mathrm{T}}_1 \tilde{A} Q_1 \ldots Q_{r+1,r+2}\,.$$

The equations (19) and (20) are algebraically consistent and can be solved as in the real case.

It may be observed that at this step the real and the imaginary parts of the eigenvector are obtained as a solution of a four-diagonal system of linear equations.

In this way the complex conjugate poles are treated in a similar manner as the real poles at the cost of a small increase of the number of computational operations (an additional subdiagonal of the open-loop system matrix is used).

The next steps are performed in the same way. At steps $(n-1)$, $n$ the vector $x_m \in \mathbb{R}^2$ is transformed only once. No element of $y_m \in \mathbb{R}^2$ is to be annihilated. The elements $k_{n-1}$, $k_n$ may be obtained from equations of type (19) which cannot be zero identities since the closed-loop system must be completely controllable.

Finally one obtains

$$S = Q^{\mathrm{T}}\tilde{A}_{\mathrm{c}}Q = Q^{\mathrm{T}}P^{\mathrm{T}}A_{\mathrm{c}}PQ =$$

$$= \begin{bmatrix} s_1 & \times & \ldots & & & & \ldots & \times \\ & \ddots & & & & & & \\ & & \ddots & & & & & \\ & & & s_r & \times & \times & \ldots & \times \\ & & & & \hat{S}_1 & \times & \ldots & \times \\ & & & & & \times & \ldots & \times \\ & \mathbf{0} & & & & \ddots & & \\ & & & & & & \ddots & \\ & & & & & & & \hat{S}_m \end{bmatrix},$$

$$\tilde{k} = [k_1, \ldots, k_n]\, Q^{\mathrm{T}} \quad \text{and} \quad k = \tilde{k}P^{\mathrm{T}}\,, \quad \text{where}$$

$$Q = Q_1 Q_2 \ldots Q_{r+1,r+2} \ldots Q_{n-1,n}\,.$$

## 4. NUMERICAL CONSIDERATIONS

The algorithm presented in the previous section has many common with the deflation techniques [2], [13] used to eliminate a known eigenvalue from an eigenvalue problem. One of these techniques is of particular interest here. If an approximated eigenvector is known it is possible to construct an orthogonal transformation in order to produce a matrix of order one less than the original matrix that does not contain the eigenvalue corresponding to the known eigenvector. It is shown in [2] that this technique is very stable, although the approximate eigenvector may be far from the accurate one. This is because the errors in the transformed matrix depend not on the errors in the eigenvector $v_i$ but on the residual $Av_i - v_i s_i$ which may be very small even if the eigenvector is not very accurate.

In this section it will be shown that our algorithm has also very good numerical properties due to the fact that the computation of an eigenvector, its transformation and the determination of a gain matrix element correspond to a small residual in the equation for this eigenvector. In this way it will be proved that the subdiagonal elements of the triangular form obtained are negligible and since it is exact for a matrix which is close to the closed-loop system matrix, this leads to the numerical stability of the algorithm.

Denote with $\square$ any of the basic arithmetic operations $+, -, \times, /$. Further on we shall assume that in floating point arithmetic $\hat{z} = \mathrm{fl}\,(x \square y) = x \square y(1 + e)$, $|e| \leqq \mathrm{eps}$. In the derivation of the error bounds the second and higher order terms in eps will be neglected.

Consider for simplicity the case of real poles only. At the step $i$; $i = 1, \ldots, n$ the computed elements $\hat{v}_{n-1,i}$ and $\hat{v}_{n-2,i}$ of the eigenvector $\tilde{v}_i$ satisfy

(21)

$$\hat{v}_{n-1,i} = (s_i - a_{nn})\,\hat{v}_{ni}(1 + e_1)/(a_{n,n-1}(1 + e_2))\,,$$

$$\hat{v}_{n-2,i} = ((s_i - a_{n-1,n-1})\,\hat{v}_{n-1,i}(1 + e_3) - a_{n-1,n}\hat{v}_{ni}(1 + e_4))/(a_{n-1,n-2}(1 + e_5))\,,$$

where $a_{ij}$ are the corresponding elements of the open-loop subsystem matrix $\tilde{A}^{(i)}$, $\hat{v}_{ni} = 1$ and $|e_1|, |e_4| \leqq 2$ eps; $|e_2|, |e_5| \leqq$ eps; $|e_3| \leqq 3$ eps.

The equations (21) show that the elements $\hat{v}_{n-1,i}, \hat{v}_{n-2,i}$ may be considered as exact for a matrix $\tilde{A}^{(i)} + E_1$, where

$$\|E_1\| \leqq 3\,\mathrm{eps}\,((n - i + 1)^{1/2}\,|s_i| + \|\tilde{A}^{(i)}\|)$$
$$\leqq 3\,\mathrm{eps}\,(n^{1/2}|s_i| + \|\tilde{A}\|)\,.$$

Hence the equation for the unknown part $\tilde{k}^{(i)}$ of the gain matrix may be represented as

$$(\tilde{A}^{(i)} - \tilde{b}^{(i)}\,\tilde{k}^{(i)} + E_1)\,\hat{v}_i = \hat{v}_i s_i\,, \tag{22}$$

where $\tilde{b}^{(i)} = [\tilde{b}_i, 0, \ldots, 0]^{\mathrm{T}}$, $\hat{v}_i = [\times, \ldots, \times, \hat{v}_{n-2,i}, \hat{v}_{n-1,i}, \hat{v}_{ni}]^{\mathrm{T}}$.

Further on a plane rotation $R_1$ is implemented to annihilate the element $\hat{v}_{ni}$. Denote

$$v_i^1 = R_1\hat{v}_i = [\times, \ldots, \times, \hat{v}_{n-2,i}, \tilde{v}_{n-1,i}, 0]^{\mathrm{T}}\,,$$

$$\hat{v}_i^1 = \mathrm{fl}\,(\hat{R}_1\hat{v}_i) = [\times, \ldots, \times, \hat{v}_{n-2,i}, \bar{v}_{n-1,i}, 0]^{\mathrm{T}}\,,$$

where $\hat{R}_1$ is the computed plane rotation and

$$\tilde{v}_{n-1,i} = (\hat{v}_{n-1,i}^2 + \hat{v}_{ni}^2)^{1/2}\,.$$

Following [2] it may be shown that

$$(23) \qquad \hat{R}_1 = R_1 + Y_1\,, \quad \|Y_1\| \leqq 3(2)^{1/2}\,\mathrm{eps}\,,$$
$$\bar{v}_{n-1,i} = \tilde{v}_{n-1,i}(1 + e_6)\,,$$

where $|e_6| \leqq 6$ eps.

The transformed matrix of the open-loop subsystem satisfies the equation

$$(24) \qquad \hat{A}_1^{(i)} = \mathrm{fl}\,(\hat{R}_1 \tilde{A}^{(i)} \hat{R}_1^{\mathrm{T}}) = R_1 \tilde{A}^{(i)} R_1^{\mathrm{T}} + F_1\,.$$

The matrix $\hat{A}_1^{(i)}$ differs from Hessenberg form by the nonzero element in the position $(n, n-2)$.

The analysis made in [2] shows that $\|F_1\| \leqq 12$ eps $\|\tilde{A}\|$.

Note that this bound takes into account the errors, made during the column transformation of the whole matrix $A$.

If $n - i > 1$ it follows from (22) that

$$(25) \qquad (\hat{A}_1^{(i)} - \tilde{b}^{(i)} \tilde{k}_1^{(i)} + \tilde{E}_1 - F_1)\, v_i^1 = v_i^1 s_i\,,$$

where $\tilde{k}_1^{(i)} = \tilde{k}^{(i)} R_1^{\mathrm{T}}$, $\tilde{E}_1 = R_1 E_1 R_1^{\mathrm{T}}$, $\|\tilde{E}_1\| = \|E_1\|$. Since $\tilde{v}_{n-1,i} = \bar{v}_{n-1,i}(1 + e_7)$, $|e_7| \leqq 6$ eps, using a nonsingular transformation with the matrix $\mathrm{diag}\,(1, \ldots, 1, 1 + e_7, 1)$, the equation (25) may be represented as

$$(26) \qquad (\hat{A}_1^{(i)} + G_1 - (\tilde{b}^{(i)} + g_1)\, \tilde{k}_1^{(i)} + \tilde{E}_1 - F_1)\, \hat{v}_i^1 = \hat{v}_i^1 s_i\,.$$

Here $g_1 = [e_7 \tilde{b}_i, 0, \ldots, 0]^{\mathrm{T}}$, $\|g_1\| \leqq 6$ eps $\|\tilde{b}\|$, $\|G_1\| \leqq 6$ eps $\|\tilde{A}\|$.

The element $\hat{v}_{n-3,i}$ is computed so that

$$(27) \qquad \hat{v}_{n-3,i} = ((s_i - a_{n-2,n-2})\, \hat{v}_{n-2,i}(1 + e_8) - $$
$$- a_{n-2,n-1} \bar{v}_{n-1,i}(1 + e_9))/(a_{n-2,n-3}(1 + e_{10}))\,,$$

where $a_{ij}$ are now elements of $\hat{A}_1^{(i)}$ and $|e_8| \leqq 3$ eps, $|e_9| \leqq 2$ eps, $|e_{10}| \leqq$ eps. The next operation is the annihilation of $\bar{v}_{n-1,i}$ by the plane rotation $R_2$. This leads to the equation

$$(\hat{A}_2^{(i)} + \tilde{G}_1 + G_2 - (\tilde{b}^{(i)} + g_1 + g_2)\, \tilde{k}_2^{(i)} + \hat{E}_1 + \tilde{E}_2 - \tilde{F}_1 - F_2)\, \hat{v}_i^{(2)} = \hat{v}_i^{(2)} s_i\,,$$

where

$$\hat{A}_2^{(i)} = \mathrm{fl}\,(\hat{R}_2 \hat{A}_1^{(i)} \hat{R}_2^{\mathrm{T}}) = R_2 \hat{A}_1^{(i)} R_2^{\mathrm{T}} + F_2\,, \quad \tilde{k}_2^{(i)} = \tilde{k}_1^{(i)} R_2^{\mathrm{T}}\,,$$
$$\hat{v}_i^{(2)} = [\times, \ldots, \times, \hat{v}_{n-3,i}, \bar{v}_{n-2,i}, 0, 0]^{\mathrm{T}}\,,$$
$$\bar{v}_{n-2,i} = \mathrm{fl}\,((\hat{v}_{n-2,i}^2 + \bar{v}_{n-1,i}^2)^{1/2}\,,$$
$$\|\tilde{G}_1\| = \|G_1\|,\ \|G_2\| \leqq 6\,\mathrm{eps}\,\|\tilde{A}\|,\ \|g_2\| \leqq 6\,\mathrm{eps}\,\|\tilde{b}\|\,,$$
$$\|\hat{E}_1\| = \|\tilde{E}_1\|,\ \|\tilde{E}_2\| \leqq 3\,\mathrm{eps}\,(n^{1/2}|s_i| + \|\tilde{A}\|)\,,$$
$$\|\tilde{F}_1\| = \|F_1\|,\ \|F_2\| \leqq 12\,\mathrm{eps}\,\|\tilde{A}\|$$

and $\hat{R}_2$ is the computed plane rotation.

The matrix $\hat{A}_2^{(i)}$ differs from Hessenberg form by the nonzero elements $a_{n-1,n-3}$

and $a_{n,n-2}$. Since $\bar{v}_{n-2,i} \neq 0$ it follows that the $(n, n-2)$-element of the matrix

$$\hat{A}_2^{(i)} + \bar{G}_1 + G_2 - (\bar{b}^{(i)} + g_1 + g_2)\,\bar{k}_2^{(i)} + \hat{E}_1 + \bar{E}_2 - \bar{F}_1 - F_2\,,$$

which has the same form as $\hat{A}_2^{(i)}$, must be equal to zero. This element is not affected by the gain matrix and that is why

$$|a_{n,n-2}| \leqq \|\bar{G}_1 + G_2 + \hat{E}_1 + \bar{E}_2 - \bar{F}_1 - F_2\| \leqq$$
$$\leqq 6 \text{ eps } n^{1/2}|s_i| + 42 \text{ eps } \|\bar{A}\|\,.$$

The last inequality shows that the element $a_{n,n-2}$ is negligible for any reasonable $|s_i|$. It may be shown in the same way that after the next step it will be possible to neglect the element $a_{n-1,n-3}$ and so on, i.e. the transformed matrix keeps its Hessenberg form. It must be noted that this is valid independently from the fact that the eigenvector itself may be not very accurate – the cause is that this approximate eigenvector at every step satisfies equation of the type (26) for slightly perturbed matrices $\hat{A}^{(i)}$ and $\bar{b}^{(i)}$.

After $n - i$ transformation the vector $\bar{b}^{(i)}$ is reduced to

$$\hat{b}^{(i)} = \text{fl}\,(\hat{R}_{n-i}\bar{b}^{(i)}) = R_{n-i}\bar{b}^{(i)} + f^{(i)}\,, \tag{28}$$

where $\hat{b}^{(i)} = [\hat{b}_i, \hat{b}_{i+1}, 0, \ldots, 0]^{\mathrm{T}}$, $R_{n-i}\bar{b}^{(i)} = [b_i, b_{i+1}, 0, \ldots, 0]^{\mathrm{T}}$, $\hat{b}_i = b_i(1 + e_{11})$, $\hat{b}_{i+1} = b_{i+1}(1 + e_{12})$; $|e_{11}|, |e_{12}| \leqq 6$ eps, $f^{(i)} = [e_{11}b_i, e_{12}b_{i+1}, 0, \ldots, 0]^{\mathrm{T}}$, $\|f^{(i)}\| \leqq \leqq 6 \text{ eps } \|\bar{b}^{(i)}\|$.

Extending this procedure and combining the resulted equations one gets

$$(\hat{A}^{(i)} - (\hat{b}^{(i)} - f^{(i)} + g^{(i)})\,\bar{k}_{n-i}^{(i)} + E^{(i)} - F^{(i)} + G^{(i)})\,\hat{v}_i^{n-i} = \hat{v}_i^{n-i}s_i\,, \tag{29}$$

where $\hat{A}^{(i)} = Q_i^{\mathrm{T}}\bar{A}^{(i)}Q_i + F^{(i)}$ is the reduced open-loop subsystem matrix, $\bar{k}_{n-i}^{(i)} = \bar{k}^{(i)}Q_i$, $Q_i = R_1^{\mathrm{T}} \ldots R_{n-i}^{\mathrm{T}}$, $\hat{v}_i^{n-i} = [\times, 0, \ldots, 0]^{\mathrm{T}}$ and

$$\begin{aligned}
\|g^{(i)}\| &\leqq 6(n - i) \exp \|\bar{b}^{(i)}\|\,,\\
\|E^{(i)}\| &\leqq 3(n - i) \text{ eps } (n^{1/2}|s_i| + \|\bar{A}\|)\,,\\
\|F^{(i)}\| &\leqq 12(n - i) \text{ eps } \|\bar{A}\|\,,\\
\|G^{(i)}\| &\leqq 6(n - i) \text{ eps } \|\bar{A}\|\,.
\end{aligned}$$

The computed product of the plane rotation implemented satisfy

$$\hat{Q}_i = Q_i + Y^{(i)}\,, \quad \|Y^{(i)}\| \leqq 6(n - i)\, n^{1/2} \text{ eps}\,. \tag{30}$$

Suppose that $|\hat{b}_i| \geqq |\hat{b}_{i+1}|$ (similar results may be obtained for the case $|\hat{b}_i| < < |\hat{b}_{i+1}|$). Then the computed element of $\bar{k}_{n-i}^{(i)}$ satisfies

$$\hat{k}_i = k_i(1 + e_{13})\,, \tag{31}$$

where $k_i = (a_{ii} - s_i)/\hat{b}_i$, $|e_{13}| \leqq 2$ eps and $a_{ii}$ is the corresponding element of $\hat{A}^{(i)}$.

The $(i, i)$- and $(i + 1, i)$-elements of the first column of the transformed closed-loop subsystem matrix are obtained as follows. Instead of computing $\text{fl}\,(a_{ii} - \hat{b}_i\hat{k}_i)$

the $(i, i)$-element is set equal to $s_i$. The $(i+1, i)$-element satisfies

$$\mathrm{fl}\,(a_{i+1,i} - \hat{b}_{i+1}\hat{k}_i) = a_{i+1,i} - \hat{b}_{i+1}\hat{k}_i + h_1^{(i)}$$
$$= a_{i+1,i} - \hat{b}_{i+1}k_i + h_2^{(i)}\,,$$

where

$$h_1^{(i)} = a_{i+1,i}e_{15} - \hat{b}_{i+1}\hat{k}_i(e_{14} + e_{15})\,,$$
$$h_2^{(i)} = a_{i+1,i}e_{15} - \hat{b}_{i+1}k_i(e_{13} + e_{14} + e_{15});\ |e_{14}|,\ |e_{15}| \leqq \mathrm{eps}$$

and since $b_{i+1}k_i = b_{i+1}(a_{ii} - s_i)(1 + e_{13})/b_i$ it follows that

$$|h_1^{(i)}| \leqq 2\,\mathrm{eps}\,(n^{1/2}|s_i| + \|\tilde{A}\|)\,,$$
$$|h_2^{(i)}| \leqq 4\,\mathrm{eps}\,(n^{1/2}|s_i| + \|\tilde{A}\|)\,.$$

Denote $k_{n-i}^{(i)} = [k_i, \times, \ldots, \times]$, $\hat{k}_{n-i}^{(i)} = [\hat{k}_i, \times, \ldots, \times]$.

The matrix $\hat{A}_c^{(i)}$ satisfies

$$\hat{A}_c^{(i)} = \mathrm{fl}\,(\hat{A}^{(i)} - \hat{b}^{(i)}\hat{k}_{n-i}^{(i)}\hat{A}) = \hat{A}^{(i)} - \hat{b}^{(i)}\hat{k}_{n-i}^{(i)} + H_1^{(i)}$$
$$= \hat{A}^{(i)} - \hat{b}^{(i)}k_{n-i}^{(i)} + H_2^{(i)}\,, \tag{32}$$

where

$$H_1^{(i)} = \begin{bmatrix} -\hat{b}_i\hat{k}_ie_{13} & \times & \ldots & \times \\ h_1^{(i)} & \times & \ldots & \times \\ 0 & \times & \ldots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \ldots & \times \end{bmatrix}, \quad H_2^{(i)} = \begin{bmatrix} 0 & \times & \ldots & \times \\ h_2^{(i)} & \times & \ldots & \times \\ 0 & \times & \ldots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \ldots & \times \end{bmatrix}.$$

From (29) and (32)

$$(\hat{A}_c^{(i)} + M^{(i)})\,\hat{v}_i^{n-i} = \hat{v}_i^{n-i}s_i\,, \tag{33}$$

where $M^{(i)} = E^{(i)} - F^{(i)} + G^{(i)} - H_2^{(i)} + (f^{(i)} - g^{(i)})\,k_{n-i}^{(i)}$. The vector $f^{(i)} - g^{(i)}$ has the form $f^{(i)} - g^{(i)} = D_i\tilde{b}^{(i)}$, where the elements of $D_i$ have modules less than $6(n - i + 1)$ eps.

Let the matrix $\hat{A}_c^{(i)}$ is represented as

$$\hat{A}_c^{(i)} = \left[\begin{array}{c|ccc} s_i & \times & \ldots & \times \\ \hline & \times & \ldots & \times \\ r_i & \vdots & & \vdots \\ & \times & \ldots & \times \end{array}\right].$$

Then from (33) one obtains $\|r_i\| \leqq \|E^{(i)}\| + \|F^{(i)}\| + \|G^{(i)}\| + |h_2^{(i)}| + \|D_i\tilde{b}^{(i)}k_i\|$. Since $\|D_i\tilde{b}^{(i)}k_i\| \leqq 6(n - i + 1)\,\mathrm{eps}\,(n^{1/2}|s_i| + \|\tilde{A}\|)$ it follows that

$$\|r_i\| \leqq (9(n - i) + 10)\,n^{1/2}\,\mathrm{eps}\,|s_i| + (27(n - i) + 10)\,\mathrm{eps}\,\|\tilde{A}\|\,. \tag{34}$$

The bound (34) shows that the subdiagonal elements of the matrix $\hat{A}_c^{(i)}$ may be considered as negligible for any reasonable desired eigenvalue $s_i$. Ususall $|s_i|$ does not exceed $\|\tilde{A}\|$.

Combining these results for $i = 1, 2, \ldots, n$ it may be proved that

$$\hat{S} + M = \begin{bmatrix} s_1 & \times & \cdots & \cdots & \times \\ & s_2 & \times & \cdots & \times \\ & & \ddots & & \vdots \\ & \mathbf{0} & & \ddots & \times \\ & & & & s_n \end{bmatrix},$$

where $\hat{S}$ is the reduced closed-loop system matrix,

$$\|M\| \leqq \|r_1\| + \|r_2\| + \ldots + \|r_n\|$$
$$\leqq (4{\cdot}5n^{5/2} + 5{\cdot}5n^{3/2}) \operatorname{eps} |s_{\max}| + (13{\cdot}5n^2 - 3{\cdot}5n) \operatorname{eps} \|\tilde{A}\|$$

and $s_{\max}$ is the eigenvalue with maximal module. Hence the computed form of the closed-loop system matrix is almost upper triangular. It should be noted however that the eigenvalues of $\hat{S}$ are not necessarily close to $s_1, \ldots, s_n$.

In the next part of the analysis it will be shown that the matrix $Q\hat{S}Q^{\mathrm{T}}$ is close to the matrix $A - b\bar{k}$, where $\bar{k}$ is the computed gain matrix and $Q$ is the product of exact orthogonal transformations.

It follows from (32) that for $i = 1, \ldots, n$

(35) $$\hat{A}_c^{(i)} = Q_i^{\mathrm{T}}(\tilde{A}^{(i)} - \tilde{b}^{(i)}\hat{k}_{n-i}^{(i)}Q_i^{\mathrm{T}})\, Q_i + T^{(i)},$$

where $T^{(i)} = F^{(i)} + H_1^{(i)} - f^{(i)}k_{n-i}^{(i)}$. Hence

(36) $$\hat{S} = Q^{\mathrm{T}}(\tilde{A} - \tilde{b}\hat{k}Q^{\mathrm{T}})\, Q + T + U,$$

where $\|T\| \leqq 8n^{3/2} \operatorname{eps} |s_{\max}| + (6n^{5/2} + 2n) \operatorname{eps} \|\tilde{A}\|$ and the term $U$, $\|U\| \leqq \leqq 2 \operatorname{eps} \|\tilde{A} - \tilde{b}\hat{k}Q^{\mathrm{T}}\|$, takes into account the errors made in the computation of the elements $a_{ij} - \hat{b}_i\hat{k}_j$ for $i = 1, \ldots, j - 1$; $j = 2, \ldots, n$.

In accordance with (30) the computed transformation matrix satisfies

(37) $$\hat{Q} = Q + Y,$$

where $\|Y\| \leqq 3(n^{5/2} - n^{3/2}) \operatorname{eps}$.

The gain matrix is computed as

(38) $$\bar{k} = \mathrm{fl}(\hat{k}\hat{Q}^{\mathrm{T}}) = \hat{k}(\hat{Q}^{\mathrm{T}} + Z), \quad \|Z\| \leqq n \operatorname{eps}.$$

From (37) and (38)

(39) $$\hat{k}Q^{\mathrm{T}} = \bar{k}(I_n - V), \; \|V\| = \|Y^{\mathrm{T}} + Z\| \leqq (3n^{5/2} - 3n^{3/2} + n) \operatorname{eps}.$$

Hence

(40) $$\hat{S} = Q^{\mathrm{T}}(\tilde{A} - \tilde{b}\bar{k} + W)\, Q,$$

where

(41) $$\|W\| \leqq \|T\| + \|U\| + \|\tilde{b}\bar{k}V\| \leqq (3n^{5/2} - 3n^{3/2} + n + 2) \operatorname{eps} \|\tilde{A} - \tilde{b}\bar{k}\| +$$
$$(9n^{5/2} - 3n^{3/2} + 3n) \operatorname{eps} \|\tilde{A}\| + 8n^{3/2} \operatorname{eps} |s_{\max}|.$$

The equation (40) and the bound (41) show that the upper triangular form obtained is exact for a matrix which is close to $\tilde{A} - \tilde{b}\tilde{k}$, i.e. the algorithm is stable with respect to the determination of the gain matrix.

Similar results may be obtained for the case of complex conjugate desired poles.

The following remark to the above analysis can be made. Since the effect of rounding errors was overestimated several times the bounds obtained (although quite satisfactory) tend to be very pessimistic especially for higher $n$. It is not possible, however, to obtain more precise bounds without significant complication of the analysis.

The analysis of the accuracy of the computed gain matrix is related to the conditioning of the pole assignment problem which may be stated in the following way. The pole assignment problem is well-conditioned if small perturbations in $A$, $b$ and $s$ lead to small changes in $k$. It is evident that if the problem is well conditioned the bounds obtained guarantee that the computed gain matrix will be close to the accurate one.

This section will be concluded with an approximative operation count for the algorithm (as usual only the terms of order $n^3$ are considered).

| | OPERATIONS |
|---|---|
| 1. Row transformation of $A$ | $2n^3/3$ |
| 2. Column transformation of $A$ | $4n^3/3$ |
| 3. Accumulation of the transformations | $2n^3$ |
| Total | $4n^3$ |

Adding to this figure the number of necessary operations for reducing the system into orthogonal canonical form one can find $17n^3/3$ operations. With respect to the array storage the algorithm requires $2n^2 + 6n$ working precision words.

## 5. AN EXAMPLE

In this section an example is given in order to illustrate the implementation of the algorithm and its numerical properties.

The computations were carried out on a VAX 11/780 using VAX-11 FORTRAN V 2·0−2 and single precision arithmetic (the relative machine precision is eps = $= 2^{-24} \approx 0{\cdot}6 \,.\, 10^{-7}$).

Consider a 10th order system with matrices

$$A = \begin{bmatrix} 1 & & & & \\ 1 & 1 & & \mathbf{0} & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & \mathbf{0} & & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and all desired poles equal to $-0{\cdot}1$.

The controllability matrix of the system is upper triangular with units on the diagonal. Its condition number, estimated by the subroutine STRCO from LINPACK

[14] is $\approx 1{\cdot}75 . 10^5$. This shows that approximately five significant digits may be lost if the controllability matrix is used in the reduction of the system.

Using the algorithm proposed the following gain matrix is obtained (all results are rounded to six decimal digits)

$$k = [11{\cdot}000,\ 54{\cdot}4500,\ 159{\cdot}720,\ 307{\cdot}461,\ 405{\cdot}849,\ 372{\cdot}028,\ 233{\cdot}846,\ 96{\cdot}4616,\ 23{\cdot}5795,\ 2{\cdot}59370] .$$

The elements of the first row of the matrix $S = Q^T A_c Q$ are

$$-0{\cdot}100000,\ -0{\cdot}417867,\ -0{\cdot}448388,\ -0{\cdot}977623,\ -1{\cdot}52878,\ -3{\cdot}22388,\ -7{\cdot}17335,\ -19{\cdot}2290,\ -64{\cdot}1255,\ -309{\cdot}039 .$$

The correctness of the results was checked by using double precision arithmetic (in this case eps $= 2^{-56} \approx 0{\cdot}14 . 10^{-16}$) and comparing both gain matrices. Their elements coincided up to five digits, the norm of the difference being $1{\cdot}08 . 10^{-4}$.

To demonstrate the stability of the method it will be shown that the problem is ill-conditioned. In fact, a perturbation $10^{-3}$ in the element $a_{10,10}$ leads to the gain matrix

$$k = [11{\cdot}0010,\ 54{\cdot}4610,\ 159{\cdot}775,\ 307{\cdot}621,\ 406{\cdot}157.\ 372{\cdot}434,\ 234{\cdot}219,\ 96{\cdot}6958,\ 23{\cdot}6762,\ 2{\cdot}61739]$$

whose elements are correct again up to five digits. Hence a perturbation of $10^{-3}$ in the data leads to a change in the gain matrix of a norm $\approx 0{\cdot}702$. This shows that the effect of the errors, introduced by the algorithm is equivalent to perturbations in the data of order much less than $10^{-3}$, i.e. the algorithm does not increase the sensitivity of the problem.

Compare now the results with the bounds predicted by the numerical analysis. The bound for the subdiagonal elements is

$$\|r_i\| \leqq 0{\cdot}68 . 10^{-4}$$

and the norm of the matrix $W$ must satisfy

$$\|W\| \leqq 0{\cdot}52 . 10^{-4}\|A - b\bar{k}\| + 0{\cdot}17 . 10^{-3}\|A\| + 0{\cdot}15 . 10^{-7} \approx \approx 0{\cdot}52 . 10^{-4}\|A - b\bar{k}\| ,$$

since $\|A\| < 10^{-3}\|A - b\bar{k}\|$.

The actual norms are

$$\|r_i\| \leqq 0{\cdot}23 . 10^{-6}$$

and

$$\|W\| \leqq 0{\cdot}1 . 10^{-6}\|A - b\bar{k}\|$$

which confirms that the a priori bounds are pessimistic.

The algorithm was also tested with various examples of order between 2 and 50 with real and complex, distinct and multiple poles. In all examples it was observed that $\|W\|/\|A_c\|$ was of order $n$ eps and the subdiagonal elements of $S$ were of the order of the machine precision.

## 6. CONCLUSIONS

An efficient computational algorithm for pole assignment of linear single-input systems, based on an orthogonal triangularization of the closed-loop system matrix, is presented. The algorithm is numerically stable with respect to the determination of the gain matrix and performs equally well with real and complex, distinct and multiple desired poles. It is applicable to ill-conditioned and high order problems and may be used for synthesis of continuous as well as discrete time systems. The algorithm is implemented as a FORTRAN program which is used for solving various problems of different orders.

(Received February 1, 1985.)

## REFERENCES


[1] A. J. Laub: Survey of computational methods in control theory. In: Electric Power Problems: The Mathematical Challenge (A. M. Erisman et al., eds.), SIAM, Philadelphia 1980, pp. 231–260.

[2] J. H. Wilkinson: The Algebraic Eigenvalue Problem. Clarendon Press, Oxford 1965.

[3] M. M. Konstantinov, P. Hr. Petkov and N. D. Christov: A Schur approach to pole assignment problem. Preprints IFAC 8th Congress, Kyoto, Aug. 1981, vol. 11, pp. 1–6.

[4] P. Hr. Petkov, M. M. Konstantinov and N. D. Christov: Pole assignment by orthogonal transformations. System Sci. *8* (1982), 85–94; also 7th Internat. Conf. Syst. Sci., Wroclaw, Sept. 1981, Paper C. 77.

[5] G. S. Miminis and C. C. Paige: An algorithm for pole assignment of time invariant linear systems. Internat. J. Control *35* (1982), 341–354.

[6] P. Hr. Petkov, N. D. Christov and M. M. Konstantinov: A computational algorithm for pole assignment of linear single-input systems. IEEE Trans. Automat. Control *AC-29* (1984), 1045–1048.

[7] W. M. Wonham: Linear Multivariable Control: A Geometric Approach. Springer-Verlag, New York 1979.

[8] M. M. Konstantinov, P. Hr. Petkov and N. D. Christov: Synthesis of linear multivariable systems with prescribed equivalent form. System Sci. *5* (1979), 381–394; also 5th Internat. Conf. Syst. Sci., Wroclaw, Sept. 1978, Paper C 46.

[9] A. Varga: Numerically reliable algorithm to test controllability. Electron. Lett. *15* (1979), 452–453.

[10] M. M. Konstantinov, P. Hr. Petkov and N. D. Christov: Synthesis of linear system with desired equivalent form. J. Comp. Appl. Math. *6* (1980), 27–35.

[11] M. M. Konstantinov, P. Hr. Petkov and N. D. Christov: Orthogonal invariants and canonical forms for linear controllable systems. Preprints IFAC 8th Congress, Kyoto, Aug. 1981, vol. 1, pp. 50–55.

[12] M. M. Konstantinov, P. Hr. Petkov and N. D. Christov: Invariants and canonical forms for linear multivariable systems under the action of orthogonal transformation groups. Kybernetika *17* (1981), 413–424.

[13] G. W. Stewart: Introduction to Matrix Computations. Academic Press, New York 1973.

[14] J. J. Dongarra, J. R. Bunch, C. B. Moler and G. W. Stewart: LINPACK User's Guide. SIAM, Philadelphia 1979.


*Dr. Mihail M. Konstantinov, Institute of Engineering Cybernetics and Robotics — Bulgarian Academy of Sciences, Akad. G. Bonchev Str. Bl. 2, 1113 Sofia. Bulgaria.*

*Dr. Petko Hr. Petkov, Dr. Nikolai D. Christov, Department of Automatics — Higher Institute of Mechanical and Electrical Engineering, Bl. 2, 1156 Sofia. Bulgaria.*