УДК 512.643

W. Auzinger, R. Stolyarchuk

A TOPIC IN THE STABILITY ANALYSIS ASSOCIATED WITH COMPANION MATRICES

W. Auzinger, R. Stolyarchuk. *Stability analysis of companion matrices*, Mat. Stud. **46** (2016), 115–120.

Assume that a family of (non-normal) matrices has stable spectra contained in the complex unit circle. This does not necessarily imply that the 2-norm of such matrices is small, and the question is under what 'natural' similarity transformation the transformed matrices will have 2-norm smaller or equal to 1. Already for the 2×2-case this is a nontrivial question, involving the analysis of a function in two complex variables (the eigenvalues) and a positive scaling parameter. We discuss and explain an approach to this problem which has been used before in the analysis of companion matrices. For the 3D case we present a numerical example.

1. Introduction and motivation. In the analysis of difference equations, in particular arising from multistep methods applied to initial value problems, the property of step-by-step stability is essential. A standard approach is to reformulate an n-step difference equation as a one-step recursion for n-dimensional vectors representing n successive iterates, see [3]. For linear constant coefficient problems the iteration matrix of this one-step recursion is a nonderogatory companion matrix C whose eigenvalues coincide with the roots of the characteristic polynomial of the multistep method under consideration. The asymptotic stability of the method depends on the location of these roots, i.e., they are required to be contained in the complex unit disc. On the other hand, for quantitative convergence investigations the *norm* of the iteration matrix C is of interest. But since a companion matrix is not normal, its norm (e.g., its spectral norm) will be larger (and possibly significantly larger) than 1 in general, even if the spectrum of C is stable.

In [2] a special bidiagonal normal form was derived for companion matrices C, which for the case of a stable spectrum yields stability in the maximum norm for the transformed system. This transformation is an alternative to the Jordan form of C and – in contrast to the Jordan form – it behaves in a regular, continuous way when transition between diagonalizable and non-diagonalizable situations is taken into consideration.

For spectral-type norms this question is much more difficult: What 'natural' transformation leads to the desired result, namely contractivity in the Euclidean norm for the transformed system?

2010 *Mathematics Subject Classification:* 15A21, 15A60.
Keywords: companion matrix; stable spectrum; similarity; contraction.
doi:10.15330/ms.46.2.115-120

In this short note we concentrate on some technical issues arising in such a stability investigation which were used in [1] without further comment. In Sec. 2 we explain how to proceed for the case of dimension $n = 2$. This purely symbolic approach is very difficult to generalize to higher dimensions n, but it may be used as the basis for an algorithm involving solution of a polynomial system, see Sec. 3.

Of course there is abundant literature dealing with similar questions. Here we only note the connection of our problem with the so-called Kreiss Matrix Theorem (see, e.g., [4]): This theorem implies that for any matrix $A \in \mathbb{C}^{n\times n}$ with a stable spectrum there exists a similarity transformation X such that $\|X^{-1} A\, X\|_2 < 1$. However, there is no general purpose algorithm to find X.

2. Problem setting and approach. The two-dimensional case. We consider a family of companion matrices

$$C = \begin{pmatrix} 0 & 1 \\ -c_0 & -c_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\zeta_1 \zeta_2 & \zeta_1 + \zeta_2 \end{pmatrix} \in \mathbb{C}^{2\times 2},$$

with complex spectra $\{\zeta_1, \zeta_2\}$ satisfying a stability condition w.r.t. the complex unit circle, i.e.,

$$|\zeta_1| \leq 1, \quad |\zeta_2| \leq 1, \quad \text{and} \quad |\zeta_1| < 1 \;\text{ if }\; \zeta_1 = \zeta_2. \tag{1}$$

We impose no other conditions on the eigenvalues ζ_1 and ζ_2, and they are also allowed to coincide. This is called the confluent case, where C not diagonalizable, and this special case is seamlessly integrated in our analysis.

Since the matrices C from our family are not normal, we generally have $\|C\|_2 > 1$. Now we aim for finding a similarity transformation of C, depending on ζ_1 and ζ_2, such that the transformed matrix T satisfies $\|T\|_2 \leq 1$. For $\zeta_1 \neq \zeta_2$, a straightforward idea would be use the diagonalization of C,

$$C = V\, Z\, V^{-1}, \tag{2a}$$

with

$$V = \begin{pmatrix} 1 & 1 \\ \zeta_1 & \zeta_2 \end{pmatrix}, \text{ a Vandermonde matrix,} \quad \text{and} \quad Z = \begin{pmatrix} \zeta_1 & 0 \\ 0 & \zeta_2 \end{pmatrix}. \tag{2b}$$

Here $\|Z\|_2 = \max\{|\zeta_1|, |\zeta_2|\} < 1$. However, For $\zeta_1 = \zeta_2$ (the non-diagonalizable case) this does not work, and also for $\zeta_1 \approx \zeta_2$ it is unnatural because V is very ill-conditioned in this case and becomes singular for $\zeta_1 \to \zeta_2$, see Example 1 below.

The following alternative transformation was proposed in [1]. It is based on a modified QR-decomposition of the transposed Vandermonde matrix V^T,

$$V^T = P^T R, \quad \text{with} \quad P = \begin{pmatrix} 1 & 1 \\ \frac{1}{2}(\zeta_1 - \zeta_2) & \frac{1}{2}(\zeta_2 - \zeta_1) \end{pmatrix}, \quad R = \begin{pmatrix} 1 & \frac{1}{2}(\zeta_1 + \zeta_2) \\ 0 & 1 \end{pmatrix},$$

where the rows of P are orthogonal to each other. (However, P is not unitary.)

Assume $\zeta_1 \neq \zeta_2$ for the moment. Then, with Z from (2b), we have

$$C = L\, T L^{-1}, \tag{3a}$$

with $L = R^T$ and $T = P\, Z\, P^{-1}$, which evaluate to

$$L = \begin{pmatrix} 1 & 0 \\ \mu & 1 \end{pmatrix}, \quad T = \begin{pmatrix} \mu & 1 \\ \sigma & \mu \end{pmatrix}, \quad \text{with } \mu = \tfrac{1}{2}(\zeta_1 + \zeta_2), \quad \sigma = \tfrac{1}{4}(\zeta_1 - \zeta_2)^2. \tag{3b}$$

Moreover, it is easy to verify that the identity (3a), with L and T defined by (3b), also remains valid for the confluent case $\zeta_1 = \zeta_2$ (even though V and P are singular)!

We also introduce a scaling parameter $\delta > 0$. With

$$\Delta := \begin{pmatrix} 1 & 0 \\ 0 & \delta \end{pmatrix} \tag{4}$$

we replace L by $L\,\Delta$ and T by $\Delta^{-1}\,T\,\Delta$. For simplicity of notation we denote these re-scaled versions again by L and Δ, i.e., we have $C = L\,T L^{-1}$ with

$$L = \begin{pmatrix} 1 & 0 \\ \mu & \delta \end{pmatrix}, \quad T = \begin{pmatrix} \mu & \delta \\ \frac{\sigma}{\delta} & \mu \end{pmatrix}.$$

Now the aim is to choose the scaling parameter $\delta > 0$ in dependence of ζ_1 and ζ_2, balancing the off-diagonal entries of T in a nontrivial way, such that the desired property

$$\|T\|_2^2 = \rho(T\,T^*) \leq 1 \tag{5}$$

holds. A 'brute-force' approach to this problem would be to symbolically compute the spectrum of $T\,T^*$ in dependence of ζ_1, ζ_2 and δ and to analyze its behavior in dependence of these parameters. However, this leads to rather complicated symbolic expressions.

Here we propose an alternative, more convenient approach. We proceed from the fact that (5) is equivalent to

$$\big\langle T\,T^* u, u \big\rangle_2 \leq \big\langle u, u \big\rangle_2 \quad \text{for all} \;\; u \in \mathbb{C}^2, \tag{6a}$$

or equivalently,

$$\big\langle T\,T^* X v, X v \big\rangle_2 \leq \big\langle X v, X v \big\rangle_2 \quad \text{for all} \;\; v \in \mathbb{C}^2, \tag{6b}$$

for any regular matrix $X \in \mathbb{C}^2$.

The matrix $T\,T^*$ contains an entry with a factor $1/\delta^2$. In order to simplify our problem we use a trick:

In (6b) *we choose* $X = \Delta$ *from* (4), *where the parameter* δ *is still unspecified.*

Then, for any $\delta \neq 0$,[1]

$$\begin{aligned} & \big\langle T\,T^* u, u \big\rangle_2 \leq \big\langle u, u \big\rangle_2 \\ \Leftrightarrow \quad & \big\langle T\,T^* \Delta v, \Delta v \big\rangle_2 \leq \big\langle \Delta v, \Delta v \big\rangle_2 \\ \Leftrightarrow \quad & \big\langle (\Delta\,T)(\Delta\,T)^* v, v \big\rangle_2 \leq \big\langle \Delta^2\, v, v \big\rangle_2, \end{aligned} \tag{7}$$

which is equivalent to the requirement

$$S := \Delta^2 - (\Delta\,T)(\Delta\,T)^* \geq 0 \quad \text{positive semidefinite.} \tag{8a}$$

The matrix S evaluates to

$$S = \begin{pmatrix} 1 - |\mu|^2 & -\mu\overline{\sigma} \\ -\overline{\mu}\sigma & -|\sigma|^2 \end{pmatrix} + \delta^2 \begin{pmatrix} -1 & -\overline{\mu} \\ -\mu & 1 - |\mu|^2 \end{pmatrix}, \tag{8b}$$

[1] There is a typing error in [1, equation (3.4)]. Equation (7) is the correct version.

with μ, σ from (3b). The determinant

$$\det S = -\delta^4 + \big(1 - 2\,|\mu|^2 + |\mu^2 - \sigma|^2\big)\delta^2 - |\sigma|^2,$$

is a strictly concave quadratic polynomial in δ^2 and assumes its maximal value for

$$\delta = \delta_{opt}(\zeta_1, \zeta_2) = \sqrt{\tfrac{1}{2}(1 - |\zeta_1|^2)(1 - |\zeta_2|^2) + \tfrac{1}{4}|\zeta_1 - \zeta_2|^2}, \tag{9a}$$

and, indeed, as shown in [1], under the stability assumption (1) we have

$$\delta_{opt}(\zeta_1, \zeta_2) > 0, \tag{9b}$$

and with this choice for δ it can be verified that

$$S \text{ is positive [semi]definite,}$$

as required in (8a). Thus, for δ from (9a) we indeed obtain $\|T\|_2 \leq 1$.

Example 1. A numerical example with real data: Let $\zeta_1 = 0.9999$, $\zeta_2 = 0.99990001$. Here, $\|C\|_2 \approx 2.4$. For the Vandermonde matrix V we have $\|V\|_2\,\|V^{-1}\|_2 \approx 4 \cdot 10^8$, i.e., the transformation (2) to diagonal form is very ill-conditioned.

Proceeding as described above we obtain $C = L\,T\,L^{-1}$ with $\delta_{opt}(\zeta_1, \zeta_2) \approx 2.8 \cdot 10^4$, and

$$T \approx \begin{pmatrix} 0.999900005 & 0.1414072\,\text{e–3} \\ 0.1767944\,\text{e–12} & 0.999900005 \end{pmatrix}, \quad \text{with} \quad \|T\|_2 \approx 0.99997 < 1.$$

The basis transformation matrix $L = R^T\Delta$ has a significantly smaller condition number than V, namely $\|L\|_2\,\|L^{-1}\|_2 \approx 1.4 \cdot 10^4$.

In the limit $\zeta_1 \to \zeta_2$ (confluent case) this converges to a properly rescaled Jordan form of the non-diagonalizable matrix C; see [1].

We also note the effect of non-normality and near-confluence for this example, which gives rise to a significant, approximately linear growth of $\|C^k\|_2$ over a long range of values of k. Although $\rho(C) < 1$, we have $\|C^k\|_2 > 1$ up to $k \approx 10000$. The maximal value $\|C^k\|_2 \approx 735$ is attained near $k = 1000$, and it becomes smaller than 1 only beginning with $k \approx 10000$.

Remark. In [1] an analogous analysis was performed for companion matrices with spectra satisfying a stability condition with respect to the left complex half-plane.

3. A numerical 3×3 example.

The symbolic computations described Sec. 2, for dimension $n = 2$ and for arbitrary $\zeta_1, \zeta_2 \in \mathbb{C}$ satisfying (1), are already rather intricate and cannot readily be generalized to higher dimensions $n > 2$. In particular, the confluent case where two or more roots ζ_j are identical cannot be dealt with in a uniform way for $n > 2$.

For the case where the different roots ζ_j are numerically specified, we now propose a numerical procedure and we describe it for the case $n = 3$.

Let pairwise distinct values $\zeta_1, \zeta_2, \zeta_3 \in \mathbb{C}$ be given numerically, satisfying the stability condition $|\zeta_j| \leq 1, j = 1, 2, 3$.

1. Compute the QR-decomposition, $V^T = Q^T R$, with Q unitary, of the transpose V^T of the Vandermonde matrix V associated with the given roots ζ_j,

$$V = \begin{pmatrix} 1 & 1 & 1 \\ \zeta_1 & \zeta_2 & \zeta_3 \\ \zeta_1^2 & \zeta_2^2 & \zeta_3^2 \end{pmatrix},$$

Then, with

$$Z := \begin{pmatrix} \zeta_1 & 0 & 0 \\ 0 & \zeta_2 & 0 \\ 0 & 0 & \zeta_3 \end{pmatrix},$$

the companion matrix C associated with the ζ_j satisfies

$$C = V Z V^{-1} = R^T Q\, Z\, Q^* R^{-T}.$$

2. With the 2-parameter ansatz for the diagonal scaling matrix

$$\Delta := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \delta_1 & 0 \\ 0 & 0 & \delta_2 \end{pmatrix},$$

we write C in the form [2]

$$C = (R^T \Delta)\, (\Delta^{-1} Q\, Z\, Q^* \Delta)\, (R^T \Delta)^{-1} =: L\, T\, L^{-1}.$$

Now, using the same trick as in Sec. 2, the desired property $\|T\|_2 \leq 1$ is seen to be equivalent to

$$S := \Delta^2 - (\Delta\, T)(\Delta\, T)^* \geq 0 \quad \text{positive semidefinite,} \tag{10}$$

where the parameters δ_1, δ_2 are still unspecified. The matrix S depends in s simple way on $\delta_j^2 =: \omega_j$, $j = 1, 2$, and $\det S =: \varphi(\omega_1, \omega_2)$ is a polynomial of degree 3 in the parameters ω_j.

3. In general, $\varphi(\omega_1, \omega_2)$ will not be globally concave, but we propose to perform a numerical search for local maxima (ω_1, ω_2) of φ satisfying $\omega_j > 0$, $j = 1, 2$, and to check whether the resulting parameter values $\delta_j = \sqrt{\omega_j}$, $j = 1, 2$, result in a matrix S satisfying (10).

Example 2. For

$$\{\zeta_1, \zeta_2, \zeta_3\} = \big\{\tfrac{9}{10},\ -\tfrac{2}{3} + \tfrac{2}{3}\,\mathrm{i},\ \tfrac{2}{3} - \tfrac{1}{2}\,\mathrm{i}\big\},$$

the associated companion matrix C has the norm $\|C\|_2 \approx 2.05$. Using the approach described above we find a local maximum of φ at $(\omega_1, \omega_2) \approx (1.07, 2.06)$, and for the resulting parameter values $(\delta_1, \delta_2) \approx (1.03, 1.44)$ we obtain $S > 0$ and $\|T\|_2 \approx 0.96 < 1$.

Remark 1. It is not guaranteed that this 'searching algorithm' always leads to the desired result, but we have tested it successfully on some more examples.

For the case where at least two of the roots ζ_j coincide, this does not work in its present form since L becomes singular. In this respect the case $n = 2$ is rather special, see Sec. 2. The generalization the approach from Sec. to higher dimension (perhaps at least $n = 3$) seems to be an interesting problem in symbolic computation.

R. Stolyarchuk gratefully acknowledges financial support by TU Wien, Austria, and the Austrian Federal Ministry of Foreign Affairs in course of a 'Maidan Scholarship'.

It can be shown (for arbitrary dimension n) that T is lower Hessenberg. If all ζ_j are real, T is tridiagonal.

REFERENCES

1. W. Auzinger, *A note on similarity to contraction for stable 2x2 companion matrices*, Ukr. Mat. Zh., **68** (2016), №3, 400–407.
2. A. Eder, G. Kirlinger, *A normal form for multistep companion matrices*, Math. Models and Methods in Applied Sciences, **11** (2001), №1, 57–70.
3. E. Hairer, G. Wanner, Solving ordinary differential equations II. Stiff and Differential-Algebraic Problems, Springer Series in Computational Mathematics, V.14, 2nd rev. edn. Springer-Verlag Berlin, Heidelberg, 1996.
4. J.C. Strikwerda, B.A. Wade, *A survey of the Kreiss matrix theorem for power bounded families of matrices and its extensions*, in: Linear Operators, Banach Center Publ., **38** (1997), 339–360.

Institut für Analysis und Scientific Computing
Technische Universität Wien, Austria
w.auzinger@tuwien.ac.at

Institute for Applied Mathematics and Fundamental Sciences
Lviv Polytechnic National University
sroksolyana@yahoo.com

Received 30.10.2016