Generalized Standard Triples

Robert M. Corless

June 9, 2023

Editor-in-Chief, Maple Transactions
Western University, Canada

talk intended for the celebration of FMD60
Happy Birthday Froilán!
These slides available at rcorless.github.io

This talk is based on the following papers

1. Eunice Chan & RMC, A New Kind of Companion Matrix (ELA 2017)
2. Eunice Chan & RMC, Minimal Height Companion Matrices for Euclid Polynomials (Math. Comput. Sci. 2019)
3. Eunice Chan et al, Algebraic Linearizations (LAA 2019)
4. Eunice Chan, RMC, & Leili Rafiee Sevyeri, Generalized Standard Triples (ELA 2021)

Contributions of many: Neil Calkin, Lalo Gonzalez-Vega, Don Knuth, Piers Lawrence, Juana Sendra, Rafa Sendra, and Steven Thornton, are gratefully acknowledged. I also thank Froilán Dopico for exceptionally detailed and patient editorial work for that last paper!

Mandelbrot polynomials and Matrices

The talk is also related to *Mandelbrot polynomials and matrices*.

1. Piers Lawrence & RMC, The Largest Root of the Mandelbrot Polynomials (Jonfest proceedings, 2013)
2. Bini and Robol's MPSolve paper (JCAM 2014) (version 1 was 2000, Bini & Fiorentino)
3. Neil J Calkin, Eunice Chan, & RMC, Some Facts and Conjectures about Mandelbrot Polynomials (Maple Transactions 2021)
4. Neil Calkin et al, A Fractal Eigenvector (American Math Monthly 2022)

Piers Lawrence had the fundamental idea which opened the door to these results.

NB: There is also a strongly related paper from 2017 by Robol, Vandebril, and Van Dooren.

Theorem

Let $P(z) \in \mathbb{C}^{n\times n}$ be a regular matrix polynomial expressed in terms of a polynomial basis $\{\phi_i(z)\}_{i=0}^{\ell}$ i.e. $P(z) = \sum_{k=0}^{\ell} P_k\phi_k(z)$. Consider a linearization $L(z) = zB - A$ of $P(z)$ such that

$$L(z)\left(\boldsymbol{\Phi}_\ell(z) \otimes I_n\right) = \left(\mathrm{e}_1 \otimes I_n\right) P(z)\,, \tag{1}$$

where $\mathrm{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{C}^\ell$ and $\boldsymbol{\Phi}_\ell(z) = \begin{bmatrix} \phi_{\ell-1}(z) & \cdots & \phi_0(z) \end{bmatrix}^T$. Let $\boldsymbol{x}$ be a vector such that $\boldsymbol{x}\boldsymbol{\Phi}_\ell(z) = 1$ and define $X = \boldsymbol{x} \otimes I_n$ and $Y = \boldsymbol{e}_1 \otimes I_n$. Then

$$P(z)^{-1} = XL(z)^{-1}Y\,. \tag{2}$$

Equation (1) can be generalized using an anszatz of Faßbender and Saltenberger (2017).

Proof

Premultiplying eq (1) by $\boldsymbol{L}^{-1}(z)$ and post-multiplying by $\boldsymbol{P}^{-1}(z)$, we have

$$\boldsymbol{L}^{-1}(z)\begin{bmatrix} \boldsymbol{I}_n \\ \boldsymbol{0}_n \\ \vdots \\ \boldsymbol{0}_n \end{bmatrix} = \begin{bmatrix} \phi_{\ell-1}(z)\boldsymbol{I}_n \\ \phi_{\ell-2}(z)\boldsymbol{I}_n \\ \vdots \\ \phi_0(z)\boldsymbol{I}_n \end{bmatrix} \boldsymbol{P}^{-1}(z)\,. \tag{3}$$

If $1 = \sum_{k=0}^{\ell-1} x_k\phi_k(z)$ is the expression of 1 in that basis, then premultiplying both sides by

$$\boldsymbol{X} = \begin{bmatrix} x_{\ell-1}\boldsymbol{I}_n & x_{\ell-2}\boldsymbol{I}_n & \ldots & x_0\boldsymbol{I}_n \end{bmatrix}$$

gives the theorem.

Tricky bits

The basis is $\{\phi_i(z)\}_{i=0}^{\ell}$ but we only use up to grade $\ell-1$ in that vector. For degree-graded bases (monomial, Chebyshev, Newton, and the like) this is trivial. However, it's not a given for (e.g.) Bernstein basis (Mackey & Perović (2016)) that we may express 1 only using part of the basis. We can, though, by a trick: If $B_k^\ell = \binom{\ell}{k} z^k (1-z)^{\ell-k}$ for $0 \le k \le \ell$ is the Bernstein basis of grade ℓ, then by "degree elevation" using

$$(j+1)B_{j+1}^{\ell}(z) + (\ell-j)B_j^{\ell}(z) = \ell B_j^{\ell-1}(z)\ , \tag{4}$$

we can still do it. The result is

$$X = \left[\frac{1}{\ell}, \frac{2}{\ell}, \ldots, \frac{\ell}{\ell}\right] \otimes I_n . \tag{5}$$

Time permitting I will show how to do Lagrange and Hermite interpolational bases as well.

Algebraic companions

Suppose we have local linearizations (A_a,B_a) for dimension n matrix polynomial $a(z)$, and (A_b, B_b) for $b(z)$ (same dimension), with

$$\begin{aligned} E_a(z)(zB_a - A_a)F_a(z) &= \mathrm{diag}(a(z), I_{N_a-n}) \\ E_b(z)(zB_b - A_b)F_b(z) &= \mathrm{diag}(b(z), I_{N_b-n}) \end{aligned} \tag{6}$$

and we wish to construct a local linearization (A_c,B_c) for $c(z) = za(z)b(z) + d$.

Suppose that we do not wish to expand this out, because we are afraid of making the conditioning worse.

Theorem 1.7 in the GST paper

Let $E_a(z)$ and $F_a(z)$ be rational matrices such that if $z \in \Sigma_a$ (ie the region in which the local linearization of $\boldsymbol{a}$ is valid) then $E_a(z)$ and $F_a(z)$ are invertible and $E_a(z)(zB_a - A_a)F_a(z) = \mathrm{diag}(\boldsymbol{a}(z), I_{N_a-n})$, and likewise let $E_b(z)$ and $F_b(z)$ be rational matrices such that if $z \in \Sigma_b$ then $E_b(z)$ and $F_b(z)$ are invertible and $E_b(z)(zB_b - A_b)F_b(z) = \mathrm{diag}(\boldsymbol{b}(z), I_{N_b-n})$.

Then the pencil $zB_c - A_c$ is a local linearization of $\boldsymbol{c}(z) = z\boldsymbol{a}(z)\boldsymbol{b}(z) + \boldsymbol{d}$ for $z \in \Sigma_a \cap \Sigma_b$, where the matrices B_c and A_c are given on the next slides:

The constructed (block upper Hessenberg) linearization

$$B_c = \begin{bmatrix} B_a & & \\ & I_n & \\ & & B_b \end{bmatrix} \tag{7}$$

and

$$A_c = \begin{bmatrix} A_a & 0_{N_a,n} & -Y_a d X_b \\ -X_a & 0_n & 0_{n,N_b} \\ 0_{N_b,N_a} & -Y_b & A_b \end{bmatrix}. \tag{8}$$

Here $X_a = [I_n, 0, \ldots, 0] F_a^{-1}(z)$, $Y_a = E_a^{-1}(z)[I_n, 0, \ldots, 0]^T$ and likewise $X_B = [I_n, 0, \ldots, 0] F_b^{-1}(z)$, and $Y_b = E_a^{-1}(z)[I_n, 0, \ldots, 0]^T$ give the elements of the (generalized) standard triples for $a(z)$ and $b(z)$.

Why this might be interesting

This gives a whole different class of possible linearizations[1]. For instance, consider a variation of Newton's example polynomial, namely $p(x) = x^3 - Tx - 5 = x(x - \sqrt{T})(x + \sqrt{T}) - 5$. Algebraic linearization gives

$$A = \begin{bmatrix} \sqrt{T} & 0 & 5 \\ -1 & 0 & 0 \\ 0 & -1 & -\sqrt{T} \end{bmatrix} \tag{9}$$

as a companion matrix. Computing the eigenvalues of *this* matrix, when $T = 2 \cdot 10^5$, results in a relative error of $1.4 \cdot 10^{-13}$ in the smallest eigenvalue, whereas using the Frobenius companion forces an error of about 10^{-9}.

[1] The theory is actually in Gohberg, Lancaster, and Rodman, though!

Varying T

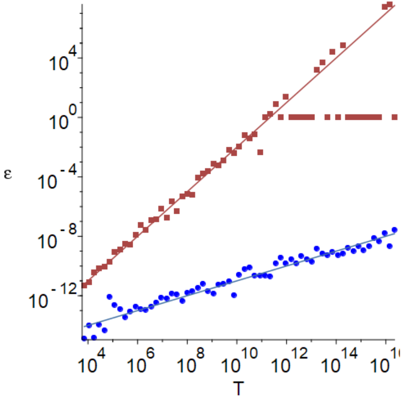

Figure 1: Relative error in smallest eigenvalue: Algebraic Linearization vs Frobenius Linearization, as the parameter T varies in $x^3 - Tx - 5$. Fits: $10^{-16} \cdot \sqrt{T}$ (blue, Algebraic), $10^{-17} \cdot T^{3/2}$ (red, Frobenius).

A general cubic

If instead we are given $P(z) = z^3 A_3 + z^2 A_2 + zA_1 + A_0$, we may write it[2] as $P(z) = z(A_3 z - B_1)(zI - B_2) + A_0$ where

$$A_1 = B_1 B_2$$
$$A_2 = -\left(B_1 + A_3 B_2\right)$$

Solving the second for B_1 and substituting into the first leads to a matrix quadratic equation for B_2:

$$A_1 = -\left(A_2 + A_3 B_2\right) B_2 \,. \tag{10}$$

So, if it's worthwhile to do this to find B_1 and B_2 as a preprocessing step, then we have another potential linearization to use. This seems that it would be valuable only in cases where the original was poorly scaled.

[2] Matrix polynomials can have nonunique factorizations!

A bigger example

We created an $n = 3$, grade 5 example by choosing a grade 2 $\boldsymbol{A}$ and a grade 2 $\boldsymbol{B}$ and a $\boldsymbol{D}$ and forming $\boldsymbol{C} = z\boldsymbol{A}\boldsymbol{B} + \boldsymbol{D}$. We perturbed it in two different ways, and compared the algebraic linearization (Frobenius for $\boldsymbol{A}$ and $\boldsymbol{B}$) to the ordinary (2nd) Frobenius linearization for the explicitly expanded $\boldsymbol{C}$.

Preliminary results

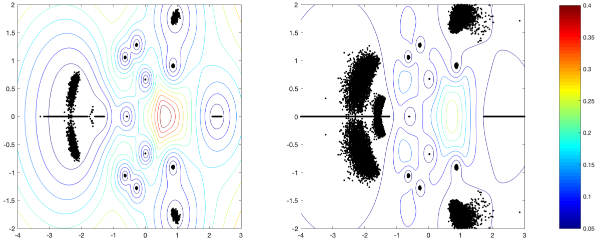

Figure 2: Pseudospectra of two different kinds of linearizations for our test equation which is expressed in the monomial basis. The linearization constructions used are algebraic linearization (left) and Frobenius linearization (right). [Graph courtesy Eunice Y. S. Chan.]

Unresolved questions

We *think* that the potentially improved numerical stability arises because the *height* of the new matrices can be lower.

$\text{Height}(A) := \|\mathrm{vec}(A)\|_\infty$ is a matrix norm, but not a submultiplicative one. For instance, consider

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} . \tag{11}$$

The height of AB is not necessarily less than the height of A times the height of B.

Also, the height of a matrix can be forced to 1 by scaling, so we are really worrying about the smallest nonzero elements after such a scaling.

Minimal height companions/linearizations

If we are given a recursive construction, this idea makes sense. But if we are given a fully formed matrix polynomial $\mathbf{P}(z)$, can we *construct* factors in a reasonable way? And how far can this be taken?

An alternative question: if the entries of the (matrix) polynomial coefficients are integers, what is the *minimal height* linearization? And how do we compute it? This looks like a discrete optimization problem. [I have asked some of my friends for advice but so far they have all looked rather helplessly at me.]

NB: As exemplified by the Mandelbrot matrices, the *minimal height* may be *exponentially smaller* than the size of the coefficients of the original polynomial.

Thank you!

Happy to take questions!

This work was partially supported by NSERC grant RGPIN-2020-06438, and partially supported by the grant PID2020-113192GB-I00 (Mathematical Visualization: Foundations, Algorithms and Applications) from the Spanish MICINN. I also thank CUNEF Universidad for the financial support to attend this event, and the organizers for including me.

Happy Birthday, Froilán!