On a root-finding approach to the polynomial eigenvalue problem

Dario A. Bini

Dipartimento di Matematica, Università di Pisa
www.dm.unipi.it/~bini

Joint work with Vanni Noferini and Meisam Sharify

Limoges, May, 10, 2012
The problem

Given \( m \times m \) matrices \( A_i, i = 0, \ldots, n \) compute the eigenvalues \( \lambda \) such that

\[
A(\lambda)v = 0, \quad v \neq 0,
\]

where \( A(\lambda) = \sum_{i=0}^{n} \lambda^i A_i \), \( \det A(\lambda) \neq 0 \)

Standard matrix approach: linearization \((A - \lambda B)w = 0\), for suitable \( mn \times mn \) matrices \( A, B \)

Polynomial approach: solving the scalar equation

\[
a(\lambda) = 0, \quad \text{where} \quad a(\lambda) = \det(A(\lambda))
\]
Structures matrix polynomials

- Skew-symmetric coefficients of even size: eigenvalues with even multiplicities since $a(x) = q(x)^2$
  
  roots appear in pairs $(\lambda, \lambda)$

- Palindromic polynomials: $A_i = A_{n-i}$
  
  T-palindromic polynomials: $A_i = A_{n-i}^T$

  roots appear in pairs $(\lambda, 1/\lambda)$ including eigenvalues at infinity

- Hamiltonian polynomials: $A_{2i}$ is Hamiltonian, $A_{2i+1}$ is skew-Hamiltonian
  
  roots appear in pairs $(\lambda, -\lambda)$

- General property of the roots: roots appear in pairs $(\lambda, f(\lambda))$, 
  
  where $f(z)$ is such that $f(f(z)) = z$

Problem: to design a polynomial eigensolver which exploits the structure, and delivers approximations in pairs $(\lambda, f(\lambda))$
The Ehrlich-Aberth iteration

Given a scalar polynomial $a(x)$ of degree $N$, given approximations $x_1^{(0)}, \ldots, x_N^{(0)}$ to the roots $\lambda_1, \ldots, \lambda_N$, the Ehrlich-Aberth method generates the sequence

$$x_i^{(\nu+1)} = x_i^{(\nu)} - \frac{\mathcal{N}(x_i^{(\nu)})}{1 - \mathcal{N}(x_i^{(\nu)}) \sum_{j=1, j \neq i}^{N} \frac{1}{x_i^{(\nu)} - x_j^{(\nu)}}}, \quad i = 1, \ldots, N$$

where $\mathcal{N}(x) = \frac{a(x)}{a'(x)}$ is the Newton correction

This iteration is just Newton’s iteration applied to the rational functions (implicit deflation)

$$g_i(x) = \frac{a(x)}{\prod_{j=1, j \neq i}^{n} (x - x_j)}$$
Some features of the E-A iteration

- Local convergence to simple roots is cubic
- Local convergence to simple roots is more than cubic in the Gauss-Seidel fashion
- Local convergence to multiple roots is linear
- Global convergence is not proved: practically, convergence always holds if initial approximations are equally placed along a circle
- Cost per iteration: $O(N^2) + N \times \text{Cost of computing } \mathcal{N}(x)$
- Number of iterations for arriving at numerical convergence is practically bounded under a suitable choice of starting approximations
Problems related to the E-A implementation

- computing $N(x)$
- choosing initial approximations
- design an effective stop condition
- design a posteriori error bounds
- design effective strategies for accelerating convergence in case of clusters or multiple roots

These issues were treated in the package MPSolve

We try to do the same for (structured) matrix polynomials
Exploiting the structure: Newton’s correction

For a general $m \times m$ matrix polynomial $A(x)$ of degree $n$, set $a(x) = \det A(x)$. Then

$$\mathcal{N}(x) = \frac{a(x)}{a'(x)} = \frac{1}{\text{trace}(A^{-1}(x)A'(x))}$$

Computational cost: $O(m^3 + m^2 n)$ ops
Structured case (palindromic and T-palindromic polynomials): Assume that $N = mn$ is even, set $z = z(x) = x + 1/x$ then

$q(z) = x(z)^{-mn/2}a(x(z))$ is a **polynomial** of degree $mn/2$,

where $x(z) = (z - \sqrt{z^2 - 4})/2$ or $x(z) = (z + \sqrt{z^2 - 4})/2$ is any of the two branches of the inverse of the function $z(x)$. 
Exploiting the structure: Newton’s correction

Structured case (palindromic and T-palindromic polynomials):
Assume that \( N = mn \) is even, set \( z = z(x) = x + 1/x \) then

\[
q(z) = x(z)^{-mn/2} a(x(z)) \quad \text{is a polynomial of degree } mn/2,
\]

where \( x(z) = (z - \sqrt{z^2 - 4})/2 \) or \( x(z) = (z + \sqrt{z^2 - 4})/2 \) is any of the two branches of the inverse of the function \( z(x) \). Moreover,

\[
\frac{q(z)}{q'(z)} = \frac{1 - 1/x^2}{g(x) - mn/(2x)}, \quad g(x) = \text{trace}(A(x)^{-1} A'(x))
\]

This way, the computation of the roots of the palyndromic polynomial \( a(x) \) of degree \( mn \) is reduced to computing the roots of the polynomial \( q(z) \) of degree \( mn/2 \).
Structured case (palindromic and T-palindromic polynomials):
Assume $mn$ odd

In this case $-1$ is root. Moreover the matrix polynomial

$$\hat{A}(x) = (x + 1)A(x)$$

is such that $(n + 1)m$ is even and $-1$ has multiplicity at least $m + 1$.

Aberth iteration can be applied to $\hat{A}(x)$ with $m + 1$ components of the initial vector equal to $-1$. 
Exploiting the structure: Newton’s correction

Structured case: Hamiltonian polynomials (\(mn\) is even)

Set \(z = x^2\). For \(x(z) = \sqrt{z}\) or \(x(z) = -\sqrt{z}\),

\[ q(z) = a(x(z)) \text{ is a polynomial of degree } mn/2. \]

Moreover,

\[ \frac{q(z)}{q'(z)} = \frac{2x}{g(x) - \frac{1}{x}}, \quad g(x) = \text{trace}(A(x)^{-1}A'(x)) \]

If \(mn\) is odd, then \(q(z) = a(x(z))/x(z)\) is a polynomial and

\[ \frac{q(z)}{q'(z)} = \frac{2x^3}{g(x) - 1}, \quad g(x) = \text{trace}(A(x)^{-1}A'(x)) \]

Once the roots \(z_1, \ldots, z_{mn/2}\) have been computed, the roots of \(a(x)\) are given by the pairs \((\sqrt{z_i}, -\sqrt{z_i}), i = 1, \ldots, mn/2\)
Exploiting the structure: Newton’s correction

General case: roots in pairs \((x, f(x))\), \(f(x) = (\alpha x + \beta)/(\gamma x + \alpha)\), \(\alpha^2 + \beta\gamma \neq 0\)

Set \(z(x) = xf(x) = \frac{\alpha x^2 + \beta x}{\gamma x - \alpha}\) and denote \(x(z)\) one of the two branches of the inverse of \(z(x)\). If there are no eigenvalues with odd multiplicity, then

\[
q(z) = \frac{a(x(z))}{(\gamma x(z) - \alpha)^{mn/2}}
\]

is a polynomial.

Dario A. Bini  
Polynomial eigenvalue problems
Computational cost

The overall cost of E-A iteration applied to an \( m \times m \) matrix polynomial of degree \( n \) is \( O(m^4 n + m^3 n^2) \) ops.

The overall cost of the technique based on linearization is \( O(m^3 n^3) \) ops.

From the complexity point of view the E-A approach is more convenient if \( n > m \).

The cost of E-A, i.e., the computation of \( \text{trace}(A(x)^{-1}A'(x)) \), can be reduced to \( O(m^2 n^2) \) ops, by using linearization. This leads to the overall cost of \( O(m^3 n^3) \) ops.

The drawback is that linearization may increase the condition number of eigenvalues.
Choosing initial approximations

The choice of initial approximations is crucial to arrive at numerical convergence with a small number of iterations.

This is particularly important for polynomials having very unbalanced roots.

The package MPSolve for computing polynomial roots to any guaranteed precision, relies on a very effective tool: the Newton polygon construction.

We synthesize the main ideas of this tool and then extend it to the case of matrix polynomials.
Choosing initial approximations

The Rouché theorem:

If \( p(x) \) and \( q(x) \) are polynomials such that \( |p(x)| > |q(x)| \) for \( |x| = r \) then \( p(x) \) and \( p(x) + q(x) \) have the same number of roots in the open disk of center 0 and radius \( r \).

Given \( a(x) = \sum_{i=0}^{n} a_i x^i \) apply Rouché theorem with \( p(x) = a_k x^k \) and \( q(x) = \sum_{i=0, i \neq k}^{n} x^i a_i \). It follows that if

\[
|p(x)| = |a_k||x|^k > \sum_{i=0, i \neq k}^{n} |a_i||x|^i \geq |q(x)|, \quad |x| = r
\]

then \( a(x) = p(x) + q(x) \) has \( k \) roots of modulus less than \( r \).
Choosing initial approximations

Properties: The equation

\[ |a_k|r^k = \sum_{i=0, i\neq k}^{n} |a_i|r^i \]

has

- one positive solution \( t_0 \) if \( k = 0 \)
- one positive solution \( s_n \) if \( k = n \)
- either no positive solutions or two positive solutions \( s \leq t \) otherwise

Pellet’s Theorem, 1881

If \( h_1, \ldots, h_p \) are the values of \( k \) for which the above equation has either one or two positive solutions \( s_{h_i} \leq t_{h_i} \) then

- the open annulus with radii \( s_{h_i}, t_{h_i} \) contains no roots of any polynomial equimodular with \( p(x) \)
- the closed annulus with radii \( t_{h_{i-1}}, s_{h_i} \) contains \( h_i - h_{i-1} \) roots
Choosing initial approximations
Choosing initial approximations: strategy of MPSolve

Computing the positive roots of $n$ polynomials is expensive.

An alternative solution relies on the following properties:

- Any positive solution $r$ of the blue equation, if it exists, is such that

\[
 u_k := \max_{i<k} \left| \frac{a_i}{a_k} \right|^{\frac{1}{k-i}} < r < \min_{i>k} \left| \frac{a_i}{a_k} \right|^{\frac{1}{k-i}} =: v_k
\]

- Let $k_i, i = 1, \ldots, q$ be such that $u_{k_i} < v_{k_i}$. Then $v_{k_i} = u_{k_{i+1}}$, moreover any interval $[s_{h_j}, t_{h_j}]$ does not contain any $u_{k_i}$

- The indices $k_i$ are the abscissas of the Newton polygon: upper convex hull of the set $\{(i, \log|a_i|), \ i = 0, \ldots, n\}$

- The values $u_{k_i}$ are the exponential of the slopes of the edges of the Newton polygon. They are called tropical roots
Choosing initial approximations: strategy of MPSolve

\[ \log |a_k| \]

- k0
- k1
- k2
- k3
- k4
- k5
- k6

Dario A. Bini  
Polynomial eigenvalue problems
Choosing initial approximations: strategy of MPSolve

\[ t_0, s_{k1}, s_{k2}, s_n \]

\[ k_3-k_2 \text{ approximations} \]

\[ k_4-k_3 \text{ approximations} \]

\[ k_5-k_4 \text{ approximations} \]

\[ k_5-k_2=h_2-h_1 \]
Starting approximations: an example

\[ p(x) = x^{10} + (1000x + 1)^3 \]

Coefficient vector: \( (1, 3000, 3000000, 10000000000, 0, 0, 0, 0, 0, 0, 1) \)

There are:

3 roots of modulus close to \( 1/1000 \),
7 roots of modulus close to \( 19.3065 \)

The Newton polygon has three vertices besides 0 and \( n \). The values of \( u_k \) are \( 0.0003, 0.001, 0.003, 19.3069 \)

7 initial approximations are placed in a circle of radius \( 19.3069 \)
3 initial approximations are placed inside a disk of radius \( 3/1000 \)
The Sharify improvement

What happens inside a blue annulus?

If a vertex of the polynomial is sufficiently sharp then we may arrive at a strong localization of the roots

Denote

\[ r_i = u_{k_i} \] the value of the radius corresponding to the vertex \( k_i \)

\[ m_i = k_{i+1} - k_i \] its “multiplicity”

\[ \delta_i = r_i / r_{i+1} \] the ratio of two consecutive radii

**Theorem (Sharify)**

If \( \delta_i, \delta_{i-1} < 1/9 \) then the polynomial \( a(x) \) has \( m_i \) roots in the annulus with radii \( r_i/3 \) and \( 3r_i \).
Let $A, B$ Hermitian matrices, define $A \succ B$ if $A - B$ is positive definite, and $A \succeq B$ if $A - B$ is positive semidefinite.

Let $F(x), G(x)$ be matrix polynomials.

**Theorem (Rouché theorem for matrix polynomials)**

If $F(x)^* F(x) \succ G(x)^* G(x)$ for $|x| = r$ then $F(x)$ and $F(x) + G(x)$ have the same number of roots in the disc of center 0 and radius $r$.

**Corollary**

If $(A_k^* A_k) r^{2k} \succ (\sum_{i=0, i \neq k}^n A_i x^i)^* (\sum_{i=0, i \neq k}^n A_i x^i)$, for $|x| = r$ then $\det A(x)$ has $kn$ roots inside the disk of center 0 and radius $r$.

The above condition requires that $\det A_k \neq 0$ and is implied by

$$ r^k \succ \sum_{i=0, i \neq k}^n \| A_k^{-1} A_i \| r^i $$
Generalization

Theorem

1. If $0 < k < n$ is such that $\det A_k \neq 0$ then the equation

$$ r^k = \sum_{i=0, i \neq k} ||A_k^{-1}A_i||r^i $$

has either no real positive solution or two real positive solutions $s_k \leq t_k$.

2. In the latter case, the polynomial $a(x) = \det A(x)$ has no roots in the open annulus of radii $s_k, t_k$, while it has $mk$ roots in the disk of radius $s_k$.

3. If $k = 0$ and $\det A_0 \neq 0$ then (1) has only one real positive root $t_0$, moreover, the polynomial $a(x)$ has no root in the open disk of radius $t_0$.

4. If $k = n$ and $\det A_n \neq 0$, then (1) has only one real positive solution $s_n$ and the polynomial $a(x)$ has no root outside the disk of radius $s_n$. 

Dario A. Bini

Polynomial eigenvalue problems
Corollary (Generalized Pellet Thm.)

Let \( h_0 < h_1 < \ldots < h_q \) be the values of \( k \) such that \( \det A_k \neq 0 \) and there exist positive real solution(s) \( s_{h_i} \leq t_{h_i} \) of (1). Then

1. \( t_{h_{i-1}} \leq s_{h_i}, \ i = 1, \ldots, q; \)
2. there are \( m(h_i - h_{i-1}) \) roots of \( a(x) \) in the annulus of radii \( t_{h_{i-1}}, s_{h_i} \);
3. there are no roots of the polynomial \( a(x) \) in the open annulus of radii \( s_{h_i}, t_{h_i} \), where \( i = 0, 1, \ldots, q \) and we assume that \( s_0 = 0, t_n = \infty \).
Theorem

If $k$ is such that (1) has two real positive solutions $s_k \leq t_k$, then $u_k \leq s_k \leq t_k \leq v_k$ where

$$u_k := \max_{i < k} \| A_k^{-1} A_i \|^{1/(k-i)} ,$$

$$v_k := \min_{i > k} \| A_k^{-1} A_i \|^{1/(k-i)} .$$

If for $k = 0$ (1) has a solution $t_0$ then

$$t_0 \leq v_0 := \min_{i > 0} \| A_0^{-1} A_i \|^{-1/i} .$$

If for $k = n$ (1) has a solution $s_n$ then

$$s_n \geq u_n := \max_{i < n} \| A_n^{-1} A_i \|^{1/(n-i)} .$$

Moreover, $v_{k_i} \leq u_{k_{i+1}}$, $i = 1, \ldots, p - 1$. 
A special class

Consider the class $Q_{m,n}$ of matrix polynomials $A(x) = \sum_{i=0}^{n} x^i A_i$, where

$$A_i = \alpha_i Q_i, \quad Q_i^* Q_i = I, \quad \alpha_i \geq 0$$

The condition given in the blue equation (1) turns into

$$\alpha_k r^k \sum_{i=0, i \neq k}^{n} \alpha_i r^i$$

and leads to the machinery valid in the scalar case for the location of the roots of a polynomial by means of the Newton polygon.

This is the best result that we can obtain for polynomials in the class $Q_{m,n}$. In fact, for $Q_i = I$ we obtain exactly $n$ copies of the scalar polynomial $p(x) = \sum_{i=0}^{n} \alpha_i x^i$ for which the inclusion result is optimal.
Theorem

Let $S = \{h_i, \quad i = 0, \ldots, p\}$ be such that $u_k < v_k$ if and only if $k \in S$. Then,

1. $\{k_0, \ldots, k_q\} \subset \{h_0, \ldots, h_p\}$, moreover
   \[\{h_0, \ldots, h_p\} \cap [s_{k_i}, t_{k_i}] = \emptyset;\]
2. $v_{k_i} \leq u_{k_{i+1}}$;
3. if $A(x) \in \mathcal{Q}_{m,n}$, then $v_{k_i} = u_{k_{i+1}}$ and $v_{k_i}$ coincide with the vertices of the Newton polygon of the polynomial
   \[w(x) = \sum_{i=0}^{n} \|A_i\| x^i = \sum_{i=0}^{n} \alpha_i x^i.\]
The Sharify bound for matrix polynomials

Let $A(x) = \sum_{i=0}^{n} x^i A_i \in \mathbb{Q}_{m,n}$, i.e., $A_i \in \mathbb{C}^{m \times m}$, $A_i = \alpha_i Q_i$, $Q_i^* Q_i = I$. Let $a(x) = \det A(x)$.

Define $r_i$ the radii (tropical roots) corresponding to the $i$-th vertex of the Newton polygon of the polynomial

$$\sum_{i=0}^{n} x^i \| A_i \|_2$$

and $m_i$ their multiplicities, define $\delta_i = r_i/r_{i-1}$

**Theorem**

For $A(x) \in \mathbb{Q}_{m,n}$, if $\delta_i, \delta_{i-1} < 1/12.12$ then the polynomial $a(x) \det A(x)$ has $mm_i$ roots in the annulus with radii $r_i/4.37$ and $4.37r_i$. 
Orthogonal random coefficients, degree $n = 13$, scaling factors:

$$\alpha = [1, 3 \cdot 10^3, 3 \cdot 10^6, 10^9, 0, \ldots 0, 10^4, 0, 0, 1]$$

<table>
<thead>
<tr>
<th>$m$</th>
<th>Newton polygon simul_it</th>
<th>aver_it</th>
<th>Unit circle simul_it</th>
<th>aver_it</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>8</td>
<td>5.4</td>
<td>243</td>
<td>191</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>5.5</td>
<td>444</td>
<td>375</td>
</tr>
<tr>
<td>20</td>
<td>11</td>
<td>5.6</td>
<td>855</td>
<td>738</td>
</tr>
<tr>
<td>40</td>
<td>13</td>
<td>6.1</td>
<td>1594</td>
<td>1466</td>
</tr>
</tbody>
</table>
Some numerical experiments: Counting \# of iterations

Non-orthogonal random coefficients, degree $n = 13$, scaling factors:

$$\alpha = [1, \ 3 \cdot 10^3, \ 3 \cdot 10^6, \ 10^9, \ 0, \ \ldots 0, \ 10^4, \ 0, \ 0, \ 1]$$

<table>
<thead>
<tr>
<th>$m$</th>
<th>Newton polygon</th>
<th>Unit circle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>simul_it</td>
<td>aver_it</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>6.8</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>7.7</td>
</tr>
<tr>
<td>20</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>40</td>
<td>16</td>
<td>10.4</td>
</tr>
</tbody>
</table>
Some numerical experiments: Counting # of iterations

General random coefficients, scaling factors:
coefficient $i$, factor: $\left(\text{rand} < 1/4\right) \times 10^{12}$
Number of average iterations with $m = 10$ and $n = 30$
Newton’s Polygon 7.2
Unit circle 111
Some numerical experiments: Approximation error

Ehrlich-Aberth: *
Polyeig: +
Problem: Sommerfeld problem
Some numerical experiments: Approximation error

Ehrlich-Aberth: ∗
Polyeig: +
Problem: Plasma drift
Some numerical experiments: Approximation error

Ehrlich-Aberth: *
Polyeig: +
QuadEig (Hammarling-Munro-Tisseur): ×

Problem: hospital
Some numerical experiments: Approximation error

Ehrlich-Aberth: ∗
Polyeig: +
QuadEig (Hammarling-Munro-Tisseur): ×

Problem: power plant
Some numerical experiments: Approximation error

Structured Ehrlich-Aberth: ○
Ehrlich-Aberth: ★
Polyeig (QZ): +
Structured URV (Schroeder): ×

Problem: butterfly
Some numerical experiments: Approximation error

Structured Ehrlich-Aberth: ○
Ehrlich-Aberth: *
Polyeig (QZ): +
Structured URV (Schroeder): ×

Problem: wiresaw1
Some numerical experiments: Approximation error

Structured Ehrlich-Aberth: ○
Ehrlich-Aberth: *
Polyeig (QZ): +
Problem with symmetry \((\lambda, \frac{\lambda+1}{\lambda-1})\)
Some references


Some references