New Methods for Computing Roots of Polynomials

Joab R. Winkler
Department of Computer Science
The University of Sheffield
United Kingdom
CONTENTS

• Difficulties of computing polynomial roots
• The geometry of ill-conditioned polynomials
• A simple polynomial root finder
• Approximate greatest common divisors
• The calculation of the polynomial roots
• Examples
• Further research
• Summary
1. DIFFICULTIES OF COMPUTING POLYNOMIAL ROOTS

There exist many algorithms for computing the roots of a polynomial:

- Bairstow, Graeffe, Jenkins-Traub, Laguerre, Müller, Newton, . . .

These methods yield satisfactory results if:

- The polynomial is of moderate degree
- The roots are simple and well-separated
- A good starting point in the iterative scheme is used

This heuristic has exceptions:

\[ f(x) = \prod_{i=1}^{20} (x - i) = (x - 1)(x - 2) \cdots (x - 20) \]
Example 1.1 Consider the polynomial

\[ x^4 - 4x^3 + 6x^2 - 4x + 1 = (x - 1)^4 \]

whose root is \( x = 1 \) with multiplicity 4. MATLAB returns the roots

1.0002, \( 1.0000 + 0.0002i \), \( 1.0000 - 0.0002i \), 0.9998

Example 1.2 The roots of the polynomial \((x - 1)^{100}\) were computed by MATLAB.

Figure 2.1: The computed roots of \((x - 1)^{100}\).
Figure 1.1: The root distribution of four perturbed polynomials.
2. THE GEOMETRY OF ILL-CONDITIONED POLYNOMIALS

- A root $x_0$ of multiplicity $r$ introduces $(r - 1)$ constraints on the coefficients.
- A monic polynomial of degree $m$ has $m$ degrees of freedom.
- The root $x_0$ lies on a manifold of dimension $(m - r + 1)$ in a space of dimension $m$.
- This manifold is called a pejorative manifold because polynomials near this manifold are ill-conditioned.
- A polynomial that lies on a pejorative manifold is well-conditioned with respect to (the structured) perturbations that keep it on the manifold, which corresponds to the situation in which the multiplicity of the roots is preserved.
- A polynomial is ill-conditioned with respect to perturbations that move it off the manifold, which corresponds to the situation in which a multiple root breaks up into a cluster of simple roots.
Example 2.1 Consider a cubic polynomial \( f(x) \) with real roots \( x_0, x_1 \) and \( x_2 \)

\[
(x - x_0)(x - x_1)(x - x_2) = x^3 - (x_0 + x_1 + x_2)x^2 + (x_0x_1 + x_1x_2 + x_2x_0)x - x_0x_1x_2
\]

- If \( f(x) \) has one double root and one simple root, then \( x_0 = x_1 \neq x_2 \) and thus \( f(x) \) can be written as

\[
x^3 - (2x_1 + x_2)x^2 + (x_1^2 + 2x_1x_2)x - x_1^2x_2
\]

The pejorative manifold of a cubic polynomial that has a double root is the surface defined by

\[
\begin{pmatrix}
-(2x_1 + x_2) & (x_1^2 + 2x_1x_2) & -x_1^2x_2
\end{pmatrix}
\begin{pmatrix}
x_1 \\
\end{pmatrix}
\neq x_2
\]
• If $f(x)$ has a triple root, then $x_0 = x_1 = x_2$ and thus $f(x)$ can be written as

$$x^3 - 3x_0x^2 + 3x_0^2x - x_0^3$$

The pejorative manifold of a cubic polynomial that has a triple root is the curve defined by

$$\left( \begin{array}{ccc} -3x_0 & 3x_0^2 & -x_0^3 \end{array} \right)$$
Theorem 2.1  The condition number of the real root $x_0$ of multiplicity $r$ of the polynomial $f(x) = (x - x_0)^r$, such that the perturbed polynomial also has a root of multiplicity $r$, is

$$\rho(x_0) := \frac{\Delta x_0}{\Delta f} = \frac{1}{r |x_0|} \frac{\|(x - x_0)^r\|}{\|(x - x_0)^{r-1}\|} = \frac{1}{r |x_0|} \left( \frac{\sum_{i=0}^{r} \binom{r}{i}^2 (x_0)^{2i}}{\sum_{i=0}^{r-1} \binom{r-1}{i}^2 (x_0)^{2i}} \right)^{\frac{1}{2}}$$

where $\|\cdot\| = \|\cdot\|_2$ and

$$\Delta f = \frac{\|\delta f\|}{\|f\|} \quad \text{and} \quad \Delta x_0 = \frac{|\delta x_0|}{|x_0|}$$
Example 2.2 The condition number $\rho(1)$ of the root $x_0 = 1$ of $(x - 1)^r$ is

$$\rho(1) = \frac{1}{r} \left( \frac{\sum_{i=0}^{r} \binom{r}{i}^2}{\sum_{i=0}^{r-1} \binom{r-1}{i}^2} \right)^{\frac{1}{2}}$$

This expression reduces to

$$\rho(1) = \frac{1}{r} \sqrt{\frac{\binom{2r}{r}}{\binom{2(r-1)}{r-1}}} = \frac{1}{r} \sqrt{\frac{2(2r - 1)}{r}} \approx \frac{2}{r} \quad \text{for large } r$$

Compare with the componentwise and normwise condition numbers

$$\kappa_c(1) \approx \frac{|\delta x_0|}{\varepsilon_c} \quad \text{and} \quad \kappa_n(1) \approx \frac{|\delta x_0|}{\varepsilon_n}$$

- $\rho(1)$ is independent of the noise level (assumed to be small)
- $\rho(1)$ decreases as the multiplicity $r$ of the root $x_0 = 1$ increases

$\blacksquare$
3. A SIMPLE POLYNOMIAL ROOT FINDER

Let \( w_i(x) \) be the product of all factors of degree \( i \) of \( f(x) \)

\[
f(x) = w_1(x)w_2^2(x)w_3^3(x) \cdots w_{r_{\text{max}}}^{r_{\text{max}}}(x)
\]

Perform a sequence of GCD computations

\[
\begin{align*}
q_1(x) &= \text{GCD} \left( f(x), f^{(1)}(x) \right) = w_2(x)w_3^2(x)w_4^3(x) \cdots w_{r_{\text{max}}}^{r_{\text{max}}-1}(x) \\
q_2(x) &= \text{GCD} \left( q_1(x), q_1^{(1)}(x) \right) = w_3(x)w_4^2(x)w_5^3(x) \cdots w_{r_{\text{max}}}^{r_{\text{max}}-2}(x) \\
q_3(x) &= \text{GCD} \left( q_2(x), q_2^{(1)}(x) \right) = w_4(x)w_5^2(x)w_6^3(x) \cdots w_{r_{\text{max}}}^{r_{\text{max}}-3}(x) \\
q_4(x) &= \text{GCD} \left( q_3(x), q_3^{(1)}(x) \right) = w_5(x)w_6^2(x)w_7^3(x) \cdots w_{r_{\text{max}}}^{r_{\text{max}}-4}(x) \\
&\vdots
\end{align*}
\]

The sequence terminates at \( q_{r_{\text{max}}}(x) \), which is a constant.
A set of polynomials $h_i(x), i = 1, \ldots, r_{\text{max}},$ is defined such that

$$
\begin{align*}
  h_1(x) &= \frac{f(x)}{q_1(x)} = w_1(x)w_2(x)w_3(x) \cdots \\
  h_2(x) &= \frac{q_1(x)}{q_2(x)} = w_2(x)w_3(x) \cdots \\
  h_3(x) &= \frac{q_2(x)}{q_3(x)} = w_3(x) \cdots \\
  &\vdots \\
  h_{r_{\text{max}}}(x) &= \frac{q_{r_{\text{max}}-2}(x)}{q_{r_{\text{max}}-1}(x)} = w_{r_{\text{max}}}(x)
\end{align*}
$$

The functions, $w_1(x), w_2(x), \cdots, w_{r_{\text{max}}}(x),$ are determined from

$$
\begin{align*}
  w_1(x) &= \frac{h_1(x)}{h_2(x)}, \quad w_2(x) = \frac{h_2(x)}{h_3(x)}, \quad \cdots, \quad w_{r_{\text{max}}-1}(x) = \frac{h_{r_{\text{max}}-2}(x)}{h_{r_{\text{max}}}(x)}
\end{align*}
$$

until

$$
  w_{r_{\text{max}}}(x) = h_{r_{\text{max}}}(x)
$$
The equations

\[ w_1(x) = 0, \quad w_2(x) = 0, \quad \cdots, \quad w_{r_{\text{max}}}(x) = 0 \]

contain only simple roots, and they yield the simple, double, triple, etc., roots of \( f(x) \).

- If \( x_0 \) is a root of \( w_i(x) \), then it is a root of multiplicity \( i \) of \( f(x) \).

Mathematical operations performed in this root finder:

- GCD computations
- Polynomial division
- Solution of simple polynomial equations
3.1 Discussion of method

- The computation of the GCD of two polynomials is an ill-posed problem because it is not a continuous function of their coefficients:
  - The polynomials \(f(x)\) and \(g(x)\) may have a non-constant GCD, but the perturbed polynomials \(f(x) + \delta f(x)\) and \(g(x) + \delta g(x)\) may be coprime.

- The determination of the degree of the GCD of two polynomials reduces to the determination of the rank of a resultant matrix, but the rank of a matrix is not defined in a floating point environment. The determination of the rank of a noisy matrix is a challenging problem that arises in many applications.

- Polynomial division, which reduces to the deconvolution of their coefficients, is an ill-conditioned problem that must be implemented with care in order to obtain a computationally reliable solution.
4. APPROXIMATE GREATEST COMMON DIVISORS

If \( f(x) \) is exact and all computations are performed in a symbolic environment, the GCD of \( f(x) \) and its derivative \( f^{(1)}(x) \) can be computed by the Sylvester resultant matrix \( S(f, f^{(1)}) \).

The polynomial \( f(x) \) is rarely known exactly, and so the given data is

\[
\tilde{f}(x) = f(x) + \delta f(x)
\]

and \( \tilde{f}(x) \) and \( \tilde{f}^{(1)}(x) \) are (with probability almost 1) coprime.

- The polynomials \( \tilde{f}(x) \) and \( \tilde{f}^{(1)}(x) \) have an approximate greatest common divisor.

- Use the method of structured total least norm applied to \( S(\tilde{f}, \tilde{f}^{(1)}) \) to compute the smallest perturbation of \( S(\tilde{f}, \tilde{f}^{(1)}) \) such that its perturbed form is singular, which implies that the perturbed form of \( \tilde{f}(x) \) has a multiple root.
Example 4.1 If \( m = 4 \) and \( n = 3 \), then

\[
f(x) = \sum_{i=0}^{4} a_i x^i
\]

the Sylvester resultant matrix \( S(f, f^{(1)}) \) is

\[
S(f, f^{(1)}) = \begin{bmatrix}
a_4 & 4a_4 \\
a_3 & a_4 & 3a_3 & 4a_4 \\
a_2 & a_3 & a_4 & 2a_2 & 3a_3 & 4a_4 \\
a_1 & a_2 & a_3 & a_1 & 2a_2 & 3a_3 & 4a_4 \\
a_0 & a_1 & a_2 & a_1 & 2a_2 & 3a_3 & 4a_4 \\
a_0 & a_1 \\
a_0 & a_1 & a_2 & a_1 & 2a_2 \\
a_0 & a_1 & a_2 & a_1 & 2a_2 \\
a_0 & a_1 & a_2 & a_1 & 2a_2 \\
\end{bmatrix}
\]
4.1 The non-uniqueness of the Sylvester resultant matrix

- An approximate GCD of $f(x)$ and $g(x)$ is equal to, up to a scalar multiplier, an approximate GCD of $f(x)$ and $\alpha g(x)$, where $\alpha$ is an arbitrary non-zero constant.

- The resultant matrix $S(f, \alpha g)$ should be used when it is desired to compute an approximate GCD of $f(x)$ and $g(x)$.

- Since $S(f, \alpha g) \neq \alpha S(f, g)$, the inclusion of $\alpha$ permits a family of approximate GCDs, rather than only one approximate GCD, to be computed.
Example 4.2 Consider the exact polynomials

\[ \hat{f}_1(x) = (x - 0.25)^8(x - 0.5)^9(x - 0.75)^{10}(x - 1)^{11}(x - 1.25)^{12} \]
\[ \hat{g}_1(x) = (x + 0.25)^4(x - 0.25)^5(x - 0.5)^6 \]

with signal-to-noise ratio \( \mu = 10^8 \).
Figure 4.1: The normalised singular values of the Sylvester resultant matrix, on a logarithmic scale, for (i) the theoretically exact data $S(\hat{f}_1, \hat{g}_1)$, ◇; (ii) the given inexact data $S(f_1, g_1)$, □; (iii) the computed data $S(\tilde{f}_1, 0, \tilde{g}_1, 0)$, ×, for $\alpha = 10^{-0.6}$.  □
5. THE CALCULATION OF THE POLYNOMIAL ROOTS

- Use the method of least squares to perform the polynomial division.
- Obtain initial estimates of the roots of the polynomial by solving a set of polynomial equations, each of whose roots is simple.
- Refine these estimates by using the method of non-linear least squares.
6. EXAMPLES

Example 6.1  Consider the Sylvester resultant matrix of \( f(x) \) and its derivative \( f^{(1)}(x) \), which has rank 26.

\[
f(x) = (x - 0.1)^8(x - 0.5)^8(x - 0.9)^8
\]

Figure 6.1: (i) The rank estimate of \( S(f, f^{(1)}) \) from the principle of maximum likelihood, and (ii) the singular values of \( S(f, f^{(1)}) \), in the absence of noise.
Figure 6.2: (i) The rank estimate of \( S(f, f^{(1)}) \) from the principle of maximum likelihood, and (ii) the singular values of \( S(f, f^{(1)}) \), for a signal-to-noise ratio of \( 10^9 \).
Example 6.2 Consider the polynomial

\[ f(x) = (x - 1)^{20}(x - 2)^{15}(x - 3)^{10}(x - 4)^{5} \]

The computed roots are

<table>
<thead>
<tr>
<th>Multiplicity</th>
<th>Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
7. FURTHER RESEARCH

- Use the displacement structure of matrices to optimise the computational efficiency of the method.

- Optimise the calculation of the scale parameter $\alpha$ in the Sylvester resultant matrix.

- Investigate the use of threshold independent methods, for example, the principle of maximum likelihood, for the estimation of the rank of a matrix.

- Consider the problem that occurs when there are bounds on the displacement of each coefficient.

- Extend to bivariate and trivariate polynomials.
8. SUMMARY

- A radically new method of solving polynomial equations has been described.
- This method first calculates the multiplicities of the roots, and then computes their values.
- The method is a computational implementation of the theory of pejorative manifolds.
- Computational experiments show that it is very successful.