

Trace matrices and clusters of zeroes

Agnes Szanto

North Carolina State University

In collaboration with

Itzmit Janovitz-Freireich (Cinvestav, Mexico City)

Bernard Mourrain (GALAAD, INRIA),

Lajos Rónyai (Hungarian Academy of Sciences and Budapest University of Technology and Economics)

The problem

Given: $f_1, \dots, f_s \in \mathbb{C}[\mathbf{x}]$ polynomials in $\mathbf{x} = (x_1, \dots, x_m)$ generating an ideal I , with finitely many roots in \mathbb{C}^m .
Suppose I either has roots with multiplicities or roots form clusters with radius $\varepsilon > 0$, clusters well separated.

The problem

Given: $f_1, \dots, f_s \in \mathbb{C}[\mathbf{x}]$ polynomials in $\mathbf{x} = (x_1, \dots, x_m)$ generating an ideal I , with finitely many roots in \mathbb{C}^m .

Suppose I either has roots with multiplicities or roots form clusters with radius $\varepsilon > 0$, clusters well separated.

We compute an **approximate radical** of I , an ideal which has exactly one root for each cluster, corresponding to the arithmetic mean of the cluster.

The problem

Given: $f_1, \dots, f_s \in \mathbb{C}[\mathbf{x}]$ polynomials in $\mathbf{x} = (x_1, \dots, x_m)$ generating an ideal I , with finitely many roots in \mathbb{C}^m .

Suppose I either has roots with multiplicities or roots form clusters with radius $\varepsilon > 0$, clusters well separated.

We compute an **approximate radical** of I , an ideal which has exactly one root for each cluster, corresponding to the arithmetic mean of the cluster.

The method's computationally most expensive part is computing a **matrix of traces**.

The problem

Given: $f_1, \dots, f_s \in \mathbb{C}[\mathbf{x}]$ polynomials in $\mathbf{x} = (x_1, \dots, x_m)$ generating an ideal I , with finitely many roots in \mathbb{C}^m .

Suppose I either has roots with multiplicities or roots form clusters with radius $\varepsilon > 0$, clusters well separated.

We compute an **approximate radical** of I , an ideal which has exactly one root for each cluster, corresponding to the arithmetic mean of the cluster.

The method's computationally most expensive part is computing a **matrix of traces**.

We propose a simple method to compute matrices of traces, and future directions for improvements.

Example

$$f(x) := (x - 1)^8(x + 2)^5(x - 3)^3(x + 4) + 10^{-5}R(x)$$

Example

$$f(x) := (x - 1)^8(x + 2)^5(x - 3)^3(x + 4) + 10^{-5}R(x)$$

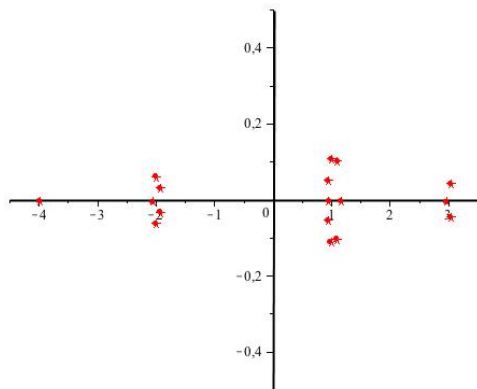


Figure: 4 clusters of size < 0.2

Example cont.

→ The corresponding 17×17 eigen-problem is *ill-conditioned!*

Example cont.

→ The corresponding 17×17 eigen-problem is *ill-conditioned!*

Instead, we compute a 4×4 companion matrix which is “square-free”:

$$M = \begin{bmatrix} 0 & 0 & 0 & -23.9974 \\ 1 & 0 & 0 & 14.0028 \\ 0 & 1 & 0 & 12.9990 \\ 0 & 0 & 1 & -2.0004 \end{bmatrix} + 10^{-10}E,$$

Example cont.

→ The corresponding 17×17 eigen-problem is *ill-conditioned!*

Instead, we compute a 4×4 companion matrix which is “square-free”:

$$M = \begin{bmatrix} 0 & 0 & 0 & -23.9974 \\ 1 & 0 & 0 & 14.0028 \\ 0 & 1 & 0 & 12.9990 \\ 0 & 0 & 1 & -2.0004 \end{bmatrix} + 10^{-10}E,$$

with eigenvalues

$$2.9998, .99986, -2.0001, -4.00006$$

Multiplication matrices

Definition

Let $I = \langle f_1, \dots, f_s \rangle$ be an ideal for which $A = \mathbb{C}[\mathbf{x}]/I$ is finite dimensional. Let $B = [b_1, \dots, b_n]$ be a basis of A . The **multiplication matrix** M_h is the matrix of the map

$$m_h : A \rightarrow A, \quad [g] \mapsto [hg]$$

written in the basis B .

Multiplication matrices

Definition

Let $I = \langle f_1, \dots, f_s \rangle$ be an ideal for which $A = \mathbb{C}[x]/I$ is finite dimensional. Let $B = [b_1, \dots, b_n]$ be a basis of A . The **multiplication matrix** M_h is the matrix of the map

$$m_h : A \rightarrow A, \quad [g] \mapsto [hg]$$

written in the basis B .

Example

$f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_0$, $B = [1, x, \dots, x^{n-1}]$ basis for $\mathbb{C}[x]/\langle f \rangle$. The multiplication matrix M_x is the companion matrix of f :

$$M_x = \begin{bmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{bmatrix}.$$

Expressions in the roots

Let $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{C}^m$ be the roots of I and $B = [b_1, \dots, b_n]$ be a basis of $A = \mathbb{C}[\mathbf{x}]/I$. Define the Vandermonde matrix

$$V := [b_j(\mathbf{z}_i)]_{i,j=1}^n \in \mathbb{C}^{n \times n}.$$

Expressions in the roots

Let $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{C}^m$ be the roots of I and $B = [b_1, \dots, b_n]$ be a basis of $A = \mathbb{C}[\mathbf{x}]/I$. Define the Vandermonde matrix

$$V := [b_j(\mathbf{z}_i)]_{i,j=1}^n \in \mathbb{C}^{n \times n}.$$

Fact

If V is invertible then

$$M_h = V \operatorname{diag}(h(\mathbf{z}_1), \dots, h(\mathbf{z}_n)) V^{-1},$$

i.e. the multiplication matrices M_h are simultaneously diagonalizable with $h(\mathbf{z}_1), \dots, h(\mathbf{z}_n)$ eigenvalues.

Expressions in the roots

Let $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{C}^m$ be the roots of I and $B = [b_1, \dots, b_n]$ be a basis of $A = \mathbb{C}[\mathbf{x}]/I$. Define the Vandermonde matrix

$$V := [b_j(\mathbf{z}_i)]_{i,j=1}^n \in \mathbb{C}^{n \times n}.$$

Fact

If V is invertible then

$$M_h = V \operatorname{diag}(h(\mathbf{z}_1), \dots, h(\mathbf{z}_n)) V^{-1},$$

i.e. the multiplication matrices M_h are simultaneously diagonalizable with $h(\mathbf{z}_1), \dots, h(\mathbf{z}_n)$ eigenvalues.

Note: If I has multiple roots then M_h is not diagonalizable.

Expressions in the roots

Let $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{C}^m$ be the roots of I and $B = [b_1, \dots, b_n]$ be a basis of $A = \mathbb{C}[\mathbf{x}]/I$. Define the Vandermonde matrix

$$V := [b_j(\mathbf{z}_i)]_{i,j=1}^n \in \mathbb{C}^{n \times n}.$$

Fact

If V is invertible then

$$M_h = V \operatorname{diag}(h(\mathbf{z}_1), \dots, h(\mathbf{z}_n)) V^{-1},$$

i.e. the multiplication matrices M_h are simultaneously diagonalizable with $h(\mathbf{z}_1), \dots, h(\mathbf{z}_n)$ eigenvalues.

Note: If I has multiple roots then M_h is not diagonalizable.

Goal: Compute multiplication matrices for the radical \sqrt{I} .

Matrix of traces

Definition

Let $B = [b_1, \dots, b_n]$ be a basis of $A = \mathbb{C}[\mathbf{x}]/I$. The **matrix of traces** is the $n \times n$ symmetric matrix:

$$R = [Tr(b_i b_j)]_{i,j=1}^n$$

where $Tr(b_i b_j)$ is the trace of the multiplication matrix $M_{b_i b_j}$.

Matrix of traces

Definition

Let $B = [b_1, \dots, b_n]$ be a basis of $A = \mathbb{C}[x]/I$. The **matrix of traces** is the $n \times n$ symmetric matrix:

$$R = [Tr(b_i b_j)]_{i,j=1}^n$$

where $Tr(b_i b_j)$ is the trace of the multiplication matrix $M_{b_i b_j}$.

Example

For $f(x) := (x - 1)^8(x + 2)^5(x - 3)^3(x + 4)$

$$Tr(M_1) = Tr(I_{17}) = 17, \quad Tr(M_x) = -a_{16} = 3, \quad Tr(M_{x^2}) = Tr(M_x^2) = 71, \quad Tr(M_{x^3}) = -15, \dots$$

so for the basis $B = [1, x, x^2, \dots, x^{16}]$ we get a 17×17 trace matrix

$$R = \begin{bmatrix} 17 & 3 & 71 & -15 & \dots \\ 3 & 71 & -15 & \dots \\ 71 & -15 & \dots \\ \vdots & & & \end{bmatrix}.$$

Matrix of traces cont.

Fact

- $\text{rank}(R) = \#\text{distinct roots in } \mathbb{C}^m$
- Let $R_{x_k} := [\text{Tr}(x_k b_i b_j)]_{i,j=1}^n$. If \tilde{R} is a maximal non-singular submatrix of R , and \tilde{R}_{x_k} is the same submatrix of R_{x_k} , then the solution M_{x_k} of

$$\tilde{R}M_{x_k} = \tilde{R}_{x_k}$$

is a multiplication matrix of x_k for the radical of \sqrt{I} .

Matrix of traces cont.

Fact

- $\text{rank}(R) = \#\text{distinct roots in } \mathbb{C}^m$
- Let $R_{x_k} := [\text{Tr}(x_k b_i b_j)]_{i,j=1}^n$. If \tilde{R} is a maximal non-singular submatrix of R , and \tilde{R}_{x_k} is the same submatrix of R_{x_k} , then the solution M_{x_k} of

$$\tilde{R}M_{x_k} = \tilde{R}_{x_k}$$

is a multiplication matrix of x_k for the radical of \sqrt{I} .

Example

For $f(x) := (x-1)^8(x+2)^5(x-3)^3(x+4)$ we choose \tilde{R} to be the principal 4×4 submatrix of the 17×17 matrix of traces of f :

$$\begin{bmatrix} 17 & 3 & 71 & -15 \\ 3 & 71 & -15 & 587 \\ 71 & -15 & 587 & -447 \\ -15 & 587 & -447 & 6611 \end{bmatrix} \cdot M_x = \begin{bmatrix} 3 & 71 & -15 & 587 \\ 71 & -15 & 587 & -447 \\ -15 & 587 & -447 & 6611 \\ 587 & -447 & 6611 & -10455 \end{bmatrix}.$$

Clusters of roots

We consider systems for which the common roots form clusters of roots.

Clusters of roots

We consider systems for which the common roots form clusters of roots.

Definition

Let $\mathbf{z}_i \in \mathbb{C}^m$ for $i = 1, \dots, k$, and consider k clusters C_1, \dots, C_k of size $|C_i| = n_i$ such that $\sum_{i=1}^k n_i = n$, each of radius proportional to the parameter ε around $\mathbf{z}_1, \dots, \mathbf{z}_k$:

$$C_i = \{\mathbf{z}_i + \delta_{i,1}\varepsilon, \dots, \mathbf{z}_i + \delta_{i,n_i}\varepsilon\},$$

where all the coordinates of $\delta_{i,j}$ are less than 1 for all i, j .

Clusters of roots

We consider systems for which the common roots form clusters of roots.

Definition

Let $\mathbf{z}_i \in \mathbb{C}^m$ for $i = 1, \dots, k$, and consider k clusters C_1, \dots, C_k of size $|C_i| = n_i$ such that $\sum_{i=1}^k n_i = n$, each of radius proportional to the parameter ε around $\mathbf{z}_1, \dots, \mathbf{z}_k$:

$$C_i = \{\mathbf{z}_i + \delta_{i,1}\varepsilon, \dots, \mathbf{z}_i + \delta_{i,n_i}\varepsilon\},$$

where all the coordinates of $\delta_{i,j}$ are less than 1 for all i, j .

In this setting we will use trace matrices to define an **approximate radical**.

GECP and SVD for the matrix of traces

GECP and SVD for the matrix of traces

Proposition

The U_k be the matrix obtained after k steps of the Gaussian Elimination with Complete Pivoting (GECP) on R for a system with k clusters is of the form

$$\begin{bmatrix} [U_k]_{1,1} & \cdots & \cdots & \cdots & [U_k]_{1,n} \\ 0 & \ddots & \cdots & \cdots & \vdots \\ & & [U_k]_{k,k} & \cdots & [U_k]_{k,n} \\ \vdots & & 0 & c_{k+1,k+1}\varepsilon^2 & \cdots & c_{k+1,n}\varepsilon^2 \\ & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & c_{n,k+1}\varepsilon^2 & \cdots & c_{n,n}\varepsilon^2 \end{bmatrix} + h.o.t.(\varepsilon).$$

GECP and SVD for the matrix of traces

Proposition

The U_k be the matrix obtained after k steps of the Gaussian Elimination with Complete Pivoting (GECP) on R for a system with k clusters is of the form

$$\begin{bmatrix} [U_k]_{1,1} & \cdots & \cdots & \cdots & [U_k]_{1,n} \\ 0 & \ddots & \cdots & \cdots & \vdots \\ & & [U_k]_{k,k} & \cdots & [U_k]_{k,n} \\ \vdots & & 0 & c_{k+1,k+1}\varepsilon^2 & \cdots & c_{k+1,n}\varepsilon^2 \\ & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & c_{n,k+1}\varepsilon^2 & \cdots & c_{n,n}\varepsilon^2 \end{bmatrix} + h.o.t.(\varepsilon).$$

Proposition

Let $\sigma_1 \geq \cdots \geq \sigma_n$ be the singular values of R . Then

$$\sigma_{k+1} = C \varepsilon^2 + h.o.t.(\varepsilon).$$

Multiplication matrices for the approximate radical

Definition

Let \tilde{R} be a maximal numerically non-singular submatrix of R , and \tilde{R}_{x_i} is the submatrix of R_{x_i} with the same row and column indices as in \tilde{R} . Then the solution M_{x_i} of the linear matrix equation

$$\tilde{R}M_{x_i} = \tilde{R}_{x_i}$$

is the multiplication matrix of x_i defining the **approximate radical**.

Multiplication matrices for the approximate radical

Definition

Let \tilde{R} be a maximal numerically non-singular submatrix of R , and \tilde{R}_{x_i} is the submatrix of R_{x_i} with the same row and column indices as in \tilde{R} . Then the solution M_{x_i} of the linear matrix equation

$$\tilde{R}M_{x_i} = \tilde{R}_{x_i}$$

is the multiplication matrix of x_i defining the **approximate radical**.

Theorem

Modulo ε^2 the multiplication matrices M_{x_1}, \dots, M_{x_m} form a pairwise commuting system of matrices for the roots ξ_1, \dots, ξ_k satisfying

$$\xi_s = \mathbf{z}_s + \frac{\sum_{r=1}^{n_s} \delta_{s,r} \varepsilon}{n_s} \pmod{\varepsilon^2}.$$

Computation of Matrices of Traces

Computation of Matrices of Traces

Let $f_1, \dots, f_m \in \mathbb{C}[x_1, \dots, x_m]$, and J their Jacobian.

Computation of Matrices of Traces

Let $f_1, \dots, f_m \in \mathbb{C}[x_1, \dots, x_m]$, and J their Jacobian.
Compute the Macaulay matrix:

f_1
f_2
\vdots
f_m
J

Computation of Matrices of Traces

Let $f_1, \dots, f_m \in \mathbb{C}[x_1, \dots, x_m]$, and J their Jacobian.

Compute a *moment matrix* in the nullspace of the top part:

$$\begin{array}{|c|} \hline f_1 \\ \hline f_2 \\ \hline \vdots \\ \hline f_m \\ \hline \end{array} \cdot \begin{array}{|c|c|c|} \hline & 0 & \\ \hline 0 & & y \\ \hline & y & \\ \hline y & & \\ \hline \end{array} = 0$$

Computation of Matrices of Traces

Let $f_1, \dots, f_m \in \mathbb{C}[x_1, \dots, x_m]$, and J their Jacobian.
 Multiply the *moment matrix* with the Jacobian part:

$$\boxed{J} \cdot \begin{array}{|c|c|c|} \hline & 0 & y \\ \hline 0 & & \\ \hline y & y & \\ \hline \end{array} = \boxed{R}$$

Univariate example

$$f = x^3 + a_1x^2 + a_2x + a_3.$$

Univariate example

$f = x^3 + a_1x^2 + a_2x + a_3$. The Sylvester matrix of f and f' :

$$\text{Syl}(f, f') = \begin{bmatrix} a_3 & a_2 & a_1 & 1 & 0 \\ 0 & a_3 & a_2 & a_1 & 1 \\ a_2 & 2a_1 & 3 & 0 & 0 \\ 0 & a_2 & 2a_1 & 3 & 0 \\ 0 & 0 & a_2 & 2a_1 & 3 \end{bmatrix},$$

Univariate example

$f = x^3 + a_1x^2 + a_2x + a_3$. A moment matrix of the top part:

$$\begin{bmatrix} a_3 & a_2 & a_1 & 1 & 0 \\ 0 & a_3 & a_2 & a_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 1 & & \\ 0 & 1 & & & \\ 1 & & & & \\ -a_1 & & & & \\ a_1^2 - a_2 & & & & \end{bmatrix} = 0$$

$$\begin{bmatrix} 0 & 0 & 1 & & \\ 0 & 1 & & & \\ 1 & & & & \\ -a_1 & & & & \\ a_1^2 - a_2 & & & & \\ -a_1 & & & & \\ a_1^2 - a_2 & & & & \\ -a_1^3 + 2a_2a_1 - a_3 & & & & \\ a_1^2 - a_2 & -a_1^3 + 2a_2a_1 - a_3 & a_1^4 - 3a_2a_1^2 + 2a_3a_1 + a_2^2 & & \end{bmatrix}$$

Univariate example

$f = x^3 + a_1x^2 + a_2x + a_3$. The bottom part multiplied by the moment matrix:

$$\begin{bmatrix} a_2 & 2a_1 & 3 & 0 & 0 \\ 0 & a_2 & 2a_1 & 3 & 0 \\ 0 & 0 & a_2 & 2a_1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -a_1 \\ 1 & -a_1 & a_1^2 - a_2 \\ -a_1 & a_1^2 - a_2 & -a_1^3 + 2a_2a_1 - a_3 \\ a_1^2 - a_2 & -a_1^3 + 2a_2a_1 - a_3 & a_1^4 - 3a_2a_1^2 + 2a_3a_1 + a_2^2 \end{bmatrix}$$

Univariate example

$f = x^3 + a_1x^2 + a_2x + a_3$. The bottom part multiplied by the moment matrix:

$$\begin{bmatrix} a_2 & 2a_1 & 3 & 0 & 0 \\ 0 & a_2 & 2a_1 & 3 & 0 \\ 0 & 0 & a_2 & 2a_1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -a_1 \\ 1 & -a_1 & a_1^2 - a_2 \\ -a_1 & a_1^2 - a_2 & -a_1^3 + 2a_2a_1 - a_3 \\ a_1^2 - a_2 & -a_1^3 + 2a_2a_1 - a_3 & a_1^4 - 3a_2a_1^2 + 2a_3a_1 + a_2^2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -a_1 & -2a_2 + a_1^2 \\ -a_1 & -2a_2 + a_1^2 & -3a_3 + 3a_2a_1 - a_1^3 \\ -2a_2 + a_1^2 & -3a_3 + 3a_2a_1 - a_1^3 & -4a_2a_1^2 + 2a_2^2 + a_1^4 + 4a_3a_1 \end{bmatrix}.$$

Over-constrained systems

Let $f_1, \dots, f_s \in \mathbb{C}[x_1, \dots, x_m]$, and assume that $s > m$.

- The method works using a generalized notion of the Jacobian.
- The size of the output matrices are much smaller than in the well-constrained case.
- The bottleneck of the method is the computation of a moment matrix:

$$\begin{array}{|c|} \hline f_1 \\ \hline f_2 \\ \hline \vdots \\ \hline f_s \\ \hline \end{array} \cdot \begin{array}{|c|c|c|} \hline & 0 & \\ \hline 0 & & y \\ \hline & y & \\ \hline y & & \\ \hline \end{array} = 0.$$

- There is a worst case degree bound

$$D = \sum_{i=1}^{m+1} d_i - m$$

where $d_1 := \deg(f_1) \geq \dots \geq d_{m+1} := \deg(f_{m+1})$ (Lazard, 1981).

Future directions

- For fixed s and d_1, \dots, d_s , there exists a unique **generic degree bound**. If $s > m + 1$ then the generic degree bound is lower than Lazard's worst case degree bound. **Goal:** Determine the generic case degree bound. (Ongoing collaboration with Mireille Boutin and Stephan Ritscher.)
- The entries of the matrix of traces are polynomials in the coefficients of f_1, \dots, f_s . **Goal:** Design a fraction free algorithm for computing the matrix of traces.
- Traces are symmetric functions of the roots. **Goal:** Design iterative methods that computes trace functions of the roots, instead of the coordinates of the roots.
- Computation of **positive definite** moment matrices are used in **solution over the reals**. **Goal:** Apply these improved techniques in real solution.

THANK YOU!