Homotopy Methods for Solving Polynomial Systems

Jan Verschelde
Department of Math, Stat & CS
University of Illinois at Chicago
Chicago, IL 60607-7045, USA

Email: jan@math.uic.edu
URL: http://www.math.uic.edu/~jan

tutorial at ISSAC’05, Beijing, China, 24 July 2005
Structure of the Tutorial

Lecture I: Root Counting and Homotopy Methods
Lecture II: Numerical Irreducible Decomposition
Lecture III: Software and Applications

Three lectures of 40 minutes each. Two breaks of 15 minutes.


This material is based upon work supported by the National Science Foundation under Grant No. 0134611 and Grant No. 0410036. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Lecture I: Root Counting and Homotopy Methods

1. Deformation Methods avoid the Discriminant Variety
2. Product Structures generalize Bézout’s Theorem
3. Newton Polytopes model Sparse Structures
4. Isolated Singularities computed by Deflation

A homotopy is \textit{optimal} if every path leads to one isolated solution. Bézout’s theorem leads to optimal homotopies for \textit{dense} systems, Bernshteïn’s theorem gives optimal homotopies for \textit{sparse} systems, \textit{if the coefficients are random enough...}
Consider for example $f(x) = 0$, $x = (x_1, x_2)$:

$$f(x) = \begin{cases} x_1^2 + x_2 - 3 = 0, \\ x_1 + 0.125x_2^2 - 1.5 = 0 \end{cases}$$

$$\frac{\partial f}{\partial x} = \begin{bmatrix} 2x_1 & 1 \\ 1 & 0.125 \end{bmatrix}.$$

Start at $x^0$, solve $\frac{\partial f}{\partial x}(x^0)\Delta x = -f(x^0)$, update $x^1 = x^0 + \Delta x$.

**Quadratic convergence if sufficiently close and regular Jacobian.**

Residual $\|f(x^0)\|$ measures **backward error.**

Condition number $C$ of $\frac{\partial f}{\partial x}(x^0)$ measures the **forward error,**

i.e.: if the coefficients are given with $m$ digits precision,

then the error on $x^0$ can be as large as $C10^{-m}$. 
1. deformation

**Homotopy with gamma trick**

A natural choice of a homotopy to solve our example is $h(x, t) = \gamma \begin{pmatrix} x_1^2 - 1 = 0 \\ x_2^2 - 1 = 0 \end{pmatrix} (1 - t) + \begin{pmatrix} x_1^2 + x_2 - 3 = 0 \\ x_1 + 0.125x_2^2 - 1.5 = 0 \end{pmatrix} t = 0,$

where $t$ goes from 0 to 1, and $\gamma \in \mathbb{C}$.

For almost all choices of $\gamma \in \mathbb{C}$, every isolated solution of multiplicity $m$ is reached by exactly $m$ solution paths.

If we take $\gamma = 1$, then at $t \approx 0.92$ singular solutions occur.
Avoiding the Discriminant Variety

With Maple 9.5 we can compute the discriminant

\[
\begin{align*}
&f := [x^2 + y - 3, x + 1/8*y^2 - 3/2]; \\
&g := [x^2 - 1, y^2 - 1]; \\
&h := [g[1]*(1-t) + f[1]*t, g[2]*(1-t) + f[2]*t]; \\
&dh := \text{Matrix}(2,2,[[\text{diff}(h[1],x),\text{diff}(h[1],y)], \\
&\quad \text{diff}(h[2],x),\text{diff}(h[2],y)]]); \\
&d := \text{LinearAlgebra}[\text{Determinant}](dh); \\
&gb := \text{Groebner}[\text{gbasis}](\{h[1],h[2],d\},\text{plex}(x,y,t)); \\
&\text{fsolve}(gb[1],t,\text{complex});
\end{align*}
\]

Elimination of \(x\) and \(y\) gives a nonzero polynomial in \(t\).

Elimination shows: only finitely many critical values for \(t\).

Random choice of \(\gamma \Rightarrow \) no critical values in \([0, 1]\).
Coefficient-Parameter Polynomial Continuation

Consider this “nontrivial example”:

\[
\begin{align*}
    f(x) &= \begin{cases} 
    0.5(x_2^2 + 4x_2x_3 + x_3^2) + a(x_2^2x_3^2 - 1) = 0, \\
    0.5(x_3^2 + 4x_3x_1 + x_1^2) + a(x_3^2x_1^2 - 1) = 0, \\
    0.5(x_1^2 + 4x_1x_2 + x_2^2) + a(x_1^2x_2^2 - 1) = 0.
    \end{cases}
\end{align*}
\]

from molecular chemistry, with parameter \( a \).

For generic choices of the parameter \( a \), there are 16 regular isolated solutions.

*Cheater’s homotopy*: substituting the parameters by random complex values gives a start system for an optimal homotopy to compute all isolated solutions.
Multihomogenous Bézout numbers

Partition \( \mathbf{x} = (x_1, x_2, x_3) \) into \((x_1), (x_2), (x_3)\) and form degree matrix \( A, A = (a_{ij}) \), \( a_{ij} = \deg(f_i, x_j) \):

\[
\begin{align*}
0.5(x_1^2 + 4x_1x_2 + x_2^2) + x_1^2x_2^2 - 1 &= 0 \\
0.5(x_2^2 + 4x_2x_3 + x_3^2) + x_2^2x_3^2 - 1 &= 0 \\
0.5(x_3^2 + 4x_3x_1 + x_1^2) + x_3^2x_1^2 - 1 &= 0 \\
\end{align*}
\]

\[
A = \begin{bmatrix} 2 & 2 & 0 \\ 0 & 2 & 2 \\ 2 & 0 & 2 \end{bmatrix}
\]

**Permanent** of \( A \) via row expansion of \( A \) as in \( \det(A) \) only with +:

\[
\text{per}(A, ((x_1), (x_2), (x_3))) = 2_{(1,1)} \times 2_{(2,2)} \times 2_{(3,3)} \\
+ 2_{(1,2)} \times 2_{(2,3)} \times 2_{(3,1)} = 16.
\]

The system \( f(\mathbf{x}) = 0 \) can have at most 16 isolated solutions.

Notice: \( 16 < 64 = 4^3 \), the 1-homogeneous Bézout number.
Linear-Product Start Systems

Computing a permanent corresponds to solving a start system:

\[ g(x) = \begin{cases} 
(c_{31} + c_{32}x_1)(c_{33} + c_{34}x_1) \cdot (c_{35} + c_{36}x_2)(c_{37} + c_{38}x_2) \cdot 1 = 0 \\
1 \cdot (c_{11} + c_{12}x_2)(c_{13} + c_{14}x_2) \cdot (c_{15} + c_{16}x_3)(c_{17} + c_{18}x_3) = 0 \\
(c_{21} + c_{22}x_1)(c_{23} + c_{24}x_1) \cdot 1 \cdot (c_{25} + c_{26}x_3)(c_{27} + c_{28}x_3) = 0 
\end{cases} \]

where the coefficients \( c_{ij} \) are random complex numbers.

For almost all choices of the coefficients \( c_{ij} \in \mathbb{C} \),
\[ g(x) = 0 \] has exactly 16 regular isolated solutions.
Recommended Background Literature


Solving Binomial Systems

\[ f(x) = \begin{cases} 
    x_1^3 x_2^2 = 1 \\
    x_1^5 x_2^1 = 1
\end{cases} \quad \Leftrightarrow \quad x^A = c : [x_1 \ x_2] \begin{bmatrix} 3 & 5 \\ 2 & 1 \end{bmatrix} = [1 \ 1]. \]

Compute \( \gcd(3, 2) = 1 = (+1) \cdot 3 + (-1) \cdot 2 \) to define \( M \):

\[
M = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}, \det(M) = 1 \quad MA = \begin{bmatrix} 1 & 4 \\ 0 & -7 \end{bmatrix} = U.
\]

\( M \) defines a coordinate change \( y^M = x \Rightarrow x^A = y^{MA} = y^U \).

\[
\begin{align*}
    x_1^3 x_2^2 &= (y_1 y_2^{-2})^3 (y_1^{-1} y_2^3)^2 = y_1^{1 \cdot 3 + (-2) \cdot 3} y_2^{(-2) \cdot 3 + 3 \cdot 2} = y_1 \\
    x_1^5 x_2^1 &= (y_1 y_2^{-2})^5 (y_1^{-1} y_2^3)^1 = y_1^{1 \cdot 5 + (-1) \cdot 1} y_2^{-2 \cdot 5 + 3 \cdot 1} = y_1^4 y_2^{-7}
\end{align*}
\]

\( f(x) = 0 \) has exactly \( 7(= |\det(A)|) \) isolated regular solutions
3. polytopes

Supports and Newton Polytopes

\[ f = (f_1, f_2) \]
\[ = \begin{cases} 
   x_1^3x_2 + x_1x_2^2 + 1 = 0 \\
   x_1^4 + x_1x_2 + 1 = 0 
\end{cases} \]

\[ \mathcal{A} = (A_1, A_2), \text{ supports of } f \]

\[ A_1 = \{(3, 1), (1, 2), (0, 0)\} \]

\[ A_2 = \{(4, 0), (1, 1), (0, 0)\} \]

The \textit{Newton polytopes}:

\[ P_1 \]
\[ (0,0) \rightarrow (1,2) \rightarrow (3,1) \rightarrow (0,0) \]

\[ P_2 \]
\[ (0,0) \rightarrow (1,1) \rightarrow (0,0) \rightarrow (4,0) \]
3. polytopes

The Cayley Trick

\[ f = (f_1, f_2) \]
\[ \begin{cases} 
    x_1^3x_2 + x_1x_2^2 + 1 = 0 \\
    x_1^4 + x_1x_2 + 1 = 0 
\end{cases} \]

\[ A = (A_1, A_2), \text{ supports of } f \]
\[ A_1 = \{(3, 1), (1, 2), (0, 0)\} \]
\[ A_2 = \{(4, 0), (1, 1), (0, 0)\} \]

A triangulation of the Cayley polytope has three cells:

The cross section contains a \textit{mixed subdivision} of \( A \).
Minkowski’s Theorem and Mixed volumes

Supports $A_1 = \{(3,1), (1,2), (0,0)\}$ and $A_2 = \{(4,0), (1,1), (0,0)\}$.

$P_1 = \text{conv}(A_1)$ and $P_2 = \text{conv}(A_2)$ are Newton polygons.

Minkowski’s theorem applied: $V(P_1, P_2)$ is \textit{mixed volume}

\[
\text{area}(\lambda_1 P_1 + \lambda_2 P_2) = V(P_1, P_1)\lambda_1^2 + V(P_1, P_2)\lambda_1 \lambda_2 + V(P_2, P_2)\lambda_2^2 \\
= 5\lambda_1^2 + 8\lambda_1 \lambda_2 + 4\lambda_2^2.
\]

Note: the area of the unit triangle equals one.
Assign an integer **lifting** to the points in $\mathcal{A}$ to obtain $\hat{\mathcal{A}} = (\hat{\mathcal{A}}_1, \hat{\mathcal{A}}_2)$:

$\hat{\mathcal{A}}_1 = \{(3,1,1), (1,2,1), (0,0,0)\}$, $\hat{\mathcal{A}}_2 = \{(4,0,0), (1,1,1), (0,0,1)\}$.

A mixed cell is spanned by two edges. Since it corresponds to a facet of the lower hull of the sum of the polytopes, it is defined by one **inner normal** $\mathbf{v} \neq \mathbf{0}$. One mixed cell is $\hat{\mathcal{C}}^{(1)} = (\hat{\mathcal{C}}_1, \hat{\mathcal{C}}_2)$:

$\hat{\mathcal{C}}_1 = \{(1,2,1), (0,0,0)\}$, $\hat{\mathcal{C}}_2 = \{(4,0,0), (1,1,1)\}$, $\langle (1,2,1), \mathbf{v} \rangle = \langle (0,0,0), \mathbf{v} \rangle < \langle (3,1,1), \mathbf{v} \rangle$

$\langle (4,0,0), \mathbf{v} \rangle = \langle (1,1,1), \mathbf{v} \rangle < \langle (0,0,1), \mathbf{v} \rangle$

$\mathbf{v} = (1, -4, 7)$, $V(\mathcal{C}^{(1)}) = 7$ **$\mathbf{v}$ satisfies linear inequality system**

Linear programming **prunes** the tree of all edge-edge combinations.
Polyhedral Homotopies induced by Lifting

\( \hat{A}_1 = \{(3, 1, 1), (1, 2, 1), (0, 0, 0)\} \quad \hat{A}_2 = \{(4, 0, 0), (1, 1, 1), (0, 0, 1)\} \)

are lifted supports, defining exponents of \( t \):

\[
\hat{g}(x, t) = \begin{cases} 
    c^{(1)}_{31} x_1^3 x_2 t^1 + c^{(1)}_{12} x_1 x_2^2 t^1 + c^{(1)}_{00} t^0 = 0 \\
    c^{(2)}_{40} x_1^4 t^0 + c^{(2)}_{11} x_1 x_2 t^1 + c^{(2)}_{00} t^1 = 0 
\end{cases}
\]

Inner normal \( \mathbf{v} = (1, -4, 7) \) defines \( x_1 = y_1 s^1, x_2 = y_2 s^{-4}, t = s^7 \):

\[
g(y, s) = \begin{cases} 
    c^{(1)}_{31} y_1^3 y_2 s^6 + c^{(1)}_{12} y_1 y_2^2 + c^{(1)}_{00} = 0 \\
    (c^{(2)}_{40} y_1^4 + c^{(2)}_{11} y_1 y_2 + c^{(2)}_{00} s^3) s^4 = 0 
\end{cases}
\]

After clearing the \( s^4 \) from the second equation, we observe:

\textbf{exponent vectors of those terms which occur with \( s^0 \) belong to the support of the mixed cell with inner normal \( \mathbf{v} \).}
Bernshtein’s first theorem

Let \( g(x) = 0 \) have the same Newton polytopes \( \mathcal{P} \) as \( f(x) = 0 \), but with randomly chosen complex coefficients.

I. Compute \( V_n(\mathcal{P}) \):
   I.1 lift polytopes ⇔ II.1 introduce parameter \( t \)
   I.2 mixed cells ⇔ II.2 start systems
   I.3 volume of mixed cell ⇔ II.3 path following

III. Coefficient-parameter continuation to solve \( f(x) = 0 \):

\[
h(x, t) = (1 - t)g(x) + tf(x) = 0, \quad \text{for } t \text{ from } 0 \text{ to } 1.
\]

#isolated solutions in \((\mathbb{C}^*)^n\), \( \mathbb{C}^* = \mathbb{C} \setminus \{0\} \), of \( f(x) = 0 \) is bounded by the mixed volume of the Newton polytopes of \( f \).
Bernshtein’s second theorem

- Face \( \partial_v f = (\partial_v f_1, \partial_v f_2, \ldots, \partial_v f_n) \) of system \( f = (f_1, f_2, \ldots, f_n) \) with Newton polytopes \( \mathcal{P} = (P_1, P_2, \ldots, P_n) \) and mixed volume \( V(\mathcal{P}) \).

\[
\partial_v f_i(x) = \sum_{a \in \partial_v A_i} c_{ia} x^a \quad \Rightarrow \quad \partial_v P_i = \text{conv}(\partial_v A_i) \\
\text{face of Newton polytope}
\]

**Theorem:** If \( \forall v \neq 0, \partial_v f(x) = 0 \) has no solutions in \((\mathbb{C}^*)^n\), then \( V(\mathcal{P}) \) is exact and all solutions are isolated.

Otherwise, for \( V(\mathcal{P}) \neq 0: V(\mathcal{P}) > \#\text{isolated solutions} \).

- Newton polytopes in general position:
  \( V(\mathcal{P}) \) is **exact** for every nonzero choice of the coefficients.
3. polytopes

**Newton polytopes in general position**

Consider $f(x) = \begin{cases} c_{111}x_1x_2 + c_{110}x_1 + c_{101}x_2 + c_{100} = 0 \\ c_{222}x_1^2x_2^2 + c_{210}x_1 + c_{201}x_2 = 0 \end{cases}$

The Newton polytopes:

\[ \forall \mathbf{v} \neq \mathbf{0} : \partial_{\mathbf{v}} A_1 + \partial_{\mathbf{v}} A_2 \leq 3 \Rightarrow V(P_1, P_2) = 4 \text{ always exact} \]

for all nonzero coefficients

*Polyhedral Endgames*: the direction $\mathbf{v}$ of a diverging solution path is normal to the faces of the Newton polytopes defining $\partial_{\mathbf{v}} f(x) = 0$ and solution at $\infty$. 
3. polytopes

**Recommended Background Literature**


Deflation Operator \( \text{Dfl} \) reduces to Corank One

Consider \( f(x) = 0 \), \( N \) equations in \( n \) unknowns, \( N \geq n \).

Suppose \( \operatorname{Rank}(A(z_0)) = R < n \) for \( z_0 \) an isolated zero of \( f(x) = 0 \).

Choose \( h \in \mathbb{C}^{R+1} \) and \( B \in \mathbb{C}^{n \times (R+1)} \) at random.

Introduce \( R + 1 \) new multiplier variables \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{R+1}) \).

\[
\text{Dfl}(f)(x, \lambda) := \begin{cases} 
  f(x) &= 0 \\
  A(x)B\lambda &= 0 \\
  h\lambda &= 1 
\end{cases} \quad \Downarrow \quad \operatorname{corank}(A(x)B) = 1
\]

Compared to the deflation of Ojika, Watanabe, and Mitsui:

1. we do not compute a maximal minor of the Jacobian matrix;
2. we only add new equations, we never replace equations.
Newton with Deflation – A Simple Example

\[ f(x, y) = \begin{cases} 
  x^2 = 0 \\
  xy = 0 \\
  y^2 = 0 
\end{cases} \]

\[ A(x, y) = \begin{bmatrix} 
  2x & 0 \\
  y & x \\
  0 & 2y 
\end{bmatrix} \]

\[ z_0 = (0, 0), m = 3 \]

\[ \text{Rank}(A(z_0)) = 0 \]

We use a 2-by-1 random matrix \( B \) and one multiplier \( \lambda_1 \):

\[
\text{Dfl}(f)(x, y, \lambda_1) = \begin{cases} 
  f(x, y) = 0 \\
  \begin{bmatrix} 
  2x & 0 \\
  y & x \\
  0 & 2y 
\end{bmatrix} \begin{bmatrix} b_{11} \\
  b_{21} \end{bmatrix} \lambda_1 = \begin{bmatrix} 0 \\
  0 
\end{bmatrix} \\
  h_1 \lambda_1 = 1, \quad \text{random } h_1 \in \mathbb{C} 
\end{cases}
\]

\( \text{Dfl}(f)(x, y, \lambda_1) = 0 \) has \((0, 0, \lambda_1^*)\) as \textbf{regular} zero!
Newton’s Method with Deflation

Input: $f(x) = 0$ polynomial system;
   $x_0$ initial approximation for $x^*$;
   $\epsilon$ tolerance for numerical rank.
Newton’s Method with Deflation

**Input:** $f(x) = 0$ polynomial system; 
$x_0$ initial approximation for $x^*$; 
$\epsilon$ tolerance for numerical rank.

\[
\begin{align*}
[A^+, R] &:= \text{SVD}(A(x_k), \epsilon); \\
x_{k+1} &:= x_k - A^+ f(x_k);
\end{align*}
\]
### Newton’s Method with Deflation

**Input:** $f(x) = 0$ polynomial system;  
$x_0$ initial approximation for $x^*$;  
$\epsilon$ tolerance for numerical rank.

\[
\begin{align*}
[A^+, R] &:= \text{SVD}(A(x_k), \epsilon); \\
x_{k+1} &:= x_k - A^+ f(x_k);
\end{align*}
\]

**Gauss-Newton**

- $R = \#\text{columns}(A)$?
  - Yes: **Output:** $f; x_{k+1}$.
Newton’s Method with Deflation

**Input:** \( f(x) = 0 \) polynomial system;
\( x_0 \) initial approximation for \( x^* \);
\( \epsilon \) tolerance for numerical rank.

\[
\begin{align*}
[A^+, R] & := \text{SVD}(A(x_k), \epsilon); \\
x_{k+1} & := x_k - A^+ f(x_k);
\end{align*}
\]

Gauss-Newton

\( R = \#\text{columns}(A) \)?

Yes

Output: \( f; x_{k+1} \).

No

\[
\begin{align*}
f & := \text{Dfl}(f)(x, \lambda) = \begin{cases} 
  f(x) = 0 \\
  G(x, \lambda) = 0
\end{cases} \\
\hat{\lambda} & := \text{LeastSquares}(G(x_{k+1}, \lambda)); \\
k & := k + 1; \quad x_k := (x_k, \hat{\lambda});
\end{align*}
\]

Deflation Step

Page 22 of 23
A Bound on the Number of Deflations

**Theorem** (Anton Leykin, JV, Ailing Zhao):

The number of deflations needed to restore the quadratic convergence of Newton’s method converging to an isolated solution is strictly less than the multiplicity.

Duality Analysis of Barry H. Dayton and Zhonggang Zeng:

(1) tighter bound on number of deflations; and

(2) special case algorithms, for corank = 1.

(Proceedings of ISSAC 2005, pages 116-123.)