Deflating Isolated Singularities by Newton's Method

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`

Joint work with Anton Leykin and Ailing Zhao.

Geometry and Symmetry in Numerical Computation
in honor of Gene Allgower, 8-11 August 2005.

Computing Singular Isolated Roots

(Outline of the Talk)

1. Problem: Newton's method **fails** for singular roots.
   Our goal is to **restore quadratic convergence**.
2. **Deflation Algorithm**: add linear combinations of derivatives.
   We rely on **only one tolerance** to determine the rank.
3. Why it works: **#deflations $<$ multiplicity**.
   The deflation reduces #monomials under the staircase.
4. Implementation and Examples: **Reconditioning**.
   We use a **directed acyclic graph** of derivative operators.
5. Prepares the computation of the **multiplicity structure**.

Singularities are keeping us in business

numerical analysis: bifurcation points and endgames

Rall (1966); Reddien (1978); Decker-Keller-Kelley (1983);
Griewank-Osborne (1981); Hoy (1989);
Deuflard-Friedler-Kunkel (1987); Kunkel (1989, 1996);
Morgan-Sommese-Wampler (1991); Li-Wang (1993, 1994);
Allgower-Schwetlick (1995); Pönisch-Schnabel-Schwetlick (1999);
Allgower-Böhmer-Hoy-Janovský (1999); Govaerts (2000)

computer algebra: standard bases (SINGULAR)

Mora (1982); Greuel-Pfister (1996); Marinari-Möller-Mora (1993)

numerical polynomial algebra: multiplicity structure

Möller-Stetter (1995); Mourrain (1997);
Stetter-Thallinger (1998); Dayton-Zeng (2005)

deflation: Ojika-Watanabe-Mitsui (1983); Lecerf (2003)

A Motivating Example: cyclic 9-roots

The system

$$f(\mathbf{x}) = \begin{cases} f_i = \displaystyle\sum_{j=0}^{8} \prod_{k=1}^{i} x_{(k+j) \bmod 9} = 0, \quad i = 1, 2, \ldots, 8 \\ f_9 = x_0 x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 - 1 = 0 \end{cases}$$

has 333×18 **isolated regular** zeros, 164 **isolated 4-fold** zeros,
and 6 **cubic 2-dimensional** irreducible solution components.

Newton's method with 64 decimal places, tolerance is 10^{-60}:

regular : 4 iterations (**quadratic convergence**)

4-fold : 79 iterations (>1 **step for one correct decimal place**)

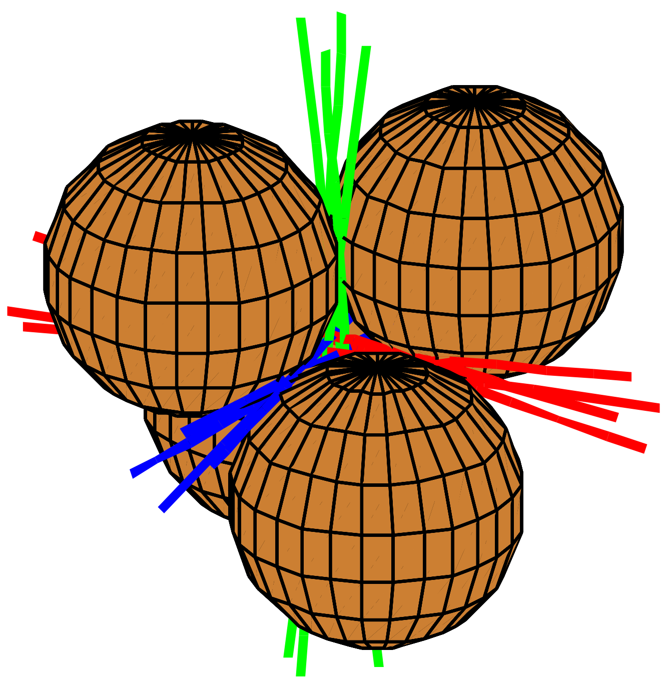
about 20 times slower to reach same magnitude of residual . . .

Twelve lines tangent to four spheres

Frank Sottile and Thorsten Theobald: Lines tangents to $2n-2$ spheres in $\mathbb{R}^n$

Trans. Amer. Math. Soc. 354

pages 4815-4829, 2002.



Problem:

Given 4 spheres,

find all lines tangent

to all 4 given spheres.

Observe:

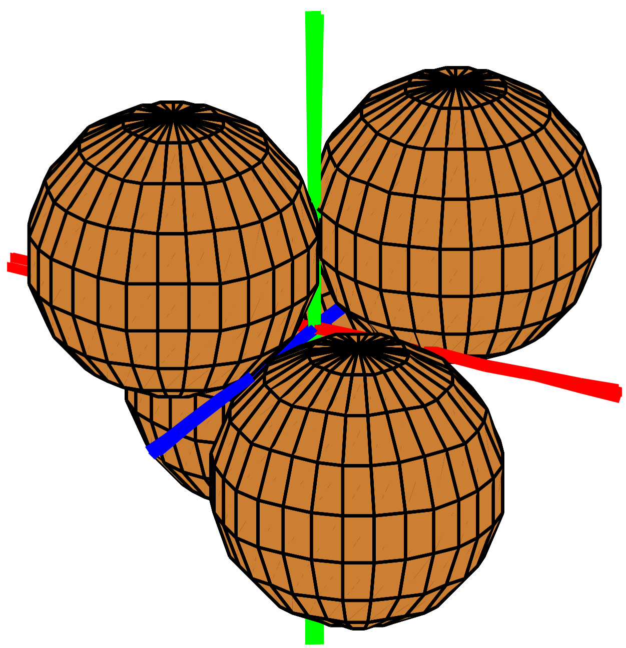
12 solutions in groups of 4.

Twelve lines tangent to four spheres

Frank Sottile and Thorsten Theobald: Lines tangents to $2n-2$ spheres in $\mathbb{R}^n$

Trans. Amer. Math. Soc. 354

pages 4815-4829, 2002.



Problem:

Given 4 spheres,

find all lines tangent

to all 4 given spheres.

Observe:

3 lines of multiplicity 4.

An Input Polynomial System

```
x0**2 + x1**2 + x2**2 - 1;
x0*x3 + x1*x4 + x2*x5;
x3**2 + x4**2 + x5**2 - 0.25;
x3**2 + x4**2 - 2*x2*x4 + x2**2 + x5**2 + 2*x1*x5 + x1**2 - 0.25;
x3**2 + 1.73205080756888*x2*x3 + 0.75*x2**2 + x4**2 - x2*x4 + 0.25*x2**2
+ x5**2 - 1.73205080756888*x0*x5 + x1*x5
+ 0.75*x0**2 - 0.86602540378444*x0*x1 + 0.25*x1**2 - 0.25;
x3**2 - 1.63299316185545*x1*x3 + 0.57735026918963*x2*x3
+ 0.66666666666667*x1**2 - 0.47140452079103*x1*x2 + 0.08333333333333*x2**2
+ x4**2 + 1.63299316185545*x0*x4 - x2*x4 + 0.66666666666667*x0**2
- 0.81649658092773*x0*x2 + 0.25*x2**2
+ x5**2 - 0.57735026918963*x0*x5 + x1*x5 + 0.08333333333333*x0**2
- 0.28867513459481*x0*x1 + 0.25*x1**2 - 0.25;
```

Original formulation as polynomial system: **Cassiano Durand**.
Centers of the spheres at the vertices of a tetrahedron: **Thorsten Theobald**.
Algebraic numbers sqrt(3), sqrt(6), etc. approximated by double floats.

The system has 6 isolated solutions, each of multiplicity 4.

Solutions at the End of Continuation

Two solutions in a **cluster**: (real and imaginary parts)

```
solution 1 :
  x0 :  -7.07106803165780E-01    3.77452918725401E-08
  x1 :  -4.08248430737360E-01   -1.83624917064964E-07
  x2 :   5.77350143082334E-01   -8.36140714113780E-08
  x3 :  -2.50000000000000E-01   -1.57896818458518E-16
  x4 :   4.33012701892221E-01   -9.11600174682333E-17
  x5 :   9.56878363411174E-08    1.54062878745083E-07
solution 2 :
  x0 :  -7.07106794356709E-01   -1.29682370414209E-07
  x1 :  -4.08248217029256E-01    1.11010906008961E-07
  x2 :   5.77350304985648E-01   -8.03312536501087E-08
  x3 :  -2.50000000000001E-01   -1.74789416181029E-16
  x4 :   4.33012701892220E-01   -1.00914936462574E-16
  x5 :  -6.07788020445124E-08   -1.39412292964849E-07
```

this is the **input** to our deflation algorithm

A Simple Example

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

(0,0) is an isolated root of multiplicity 3

Randomization or Embedding:

$$\begin{cases} x^2 + \gamma_1 y^2 = 0 \\ xy + \gamma_2 y^2 = 0 \end{cases} \quad \text{or} \quad \begin{cases} x^2 + \gamma_1 z = 0 \\ xy + \gamma_2 z = 0 \\ y^2 + \gamma_3 z = 0, \end{cases}$$

where $\gamma_1, \gamma_2, \gamma_3 \in \mathbb{C}$ are random numbers and z is a slack variable,

raises the multiplicity from 3 to 4!

Simple to algebraic geometry, but not to numerical homotopies...

Newton's Method for Overdetermined Systems

Singular Value Decomposition of N-by-n Jacobian matrix J_f:

$J_f = U\Sigma V^T$, $\quad$ U and V are orthogonal: $U^T U = I_N, V^T V = I_n$,

and singular values $\boldsymbol{\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n}$ as the only nonzero elements on the diagonal of the N-by-n matrix Σ $(N > n)$.

The **condition number** $\mathrm{cond}(J_f(\mathbf{z})) = \frac{\sigma_1}{\sigma_n}$.

$\mathrm{Rank}(J_f(\mathbf{z})) = R \iff \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_R, 0, \ldots, 0)$.

At a **multiple root $\mathbf{z}_0$**: $\mathrm{Rank}(J_f(\mathbf{z}_0)) = R < n$.

Close to $\mathbf{z}_0$, $\mathbf{z} \approx \mathbf{z}_0 : \sigma_{R+1} \approx 0$, or $\boldsymbol{|\sigma_{R+1}| < \epsilon}$, ϵ is tolerance.

Moore-Penrose inverse: $J_f^+ = V\Sigma^+ U^T$, with $R = \mathrm{Rank}(J_f)$,

and $\Sigma^+ = \mathrm{diag}(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_R}, 0, \ldots, 0)$.

Then $\Delta\mathbf{z} = -J_f(\mathbf{z})^+ f(\mathbf{z})$ is the least squares solution.

Dedieu-Shub (1999); Li-Zeng (2005)

The Simple Example – with deflation

$$f(x,y) = \begin{cases} x^2 = 0 \\ xy = 0 \\ y^2 = 0 \end{cases} \qquad J_f(x,y) = \begin{bmatrix} 2x & 0 \\ y & x \\ 0 & 2y \end{bmatrix}$$

$$\mathbf{z}_0 = (0,0), m = 3$$
$$\mathrm{Rank}(J_f(\mathbf{z}_0)) = 0$$

A nontrivial linear combination of the columns of $J_f(\mathbf{z}_0)$ is zero.

$$F(x,y,\lambda_1,\lambda_2) = \begin{cases} f(x,y) & = & 0 \\ \begin{bmatrix} 2x & 0 \\ y & x \\ 0 & 2y \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} & = & \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ h_1\lambda_1 + h_2\lambda_2 & = & 1, \qquad \text{random } h_1, h_2 \in \mathbb{C} \end{cases}$$

The system $F(x,y,\lambda_1,\lambda_2) = 0$ has $(0,0,\lambda_1^*,\lambda_2^*)$ as **regular** zero!

Deflation Operator Dfl reduces to Corank One

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

Suppose Rank$(A(\mathbf{z}_0)) = R < n$ for $\mathbf{z}_0$ an isolated zero of $f(\mathbf{x}) = 0$.

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

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

$$\mathbf{Dfl}(f)(\mathbf{x}, \boldsymbol{\lambda}) := \left\{ \begin{array}{rcl} f(\mathbf{x}) & = & \mathbf{0} \\ A(\mathbf{x})B\boldsymbol{\lambda} & = & \mathbf{0} \\ \mathbf{h}\boldsymbol{\lambda} & = & 1 \end{array} \right. \qquad \begin{array}{c} \mathrm{Rank}(A(\mathbf{x})) = R \\ \Downarrow \\ \mathrm{corank}(A(\mathbf{x})B) = 1 \end{array}$$

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's Method with Deflation

Input: $f(\mathbf{x}) = \mathbf{0}$ polynomial system;
$\mathbf{x}_0$ initial approximation for $\mathbf{x}^*$;
ϵ tolerance for numerical rank.

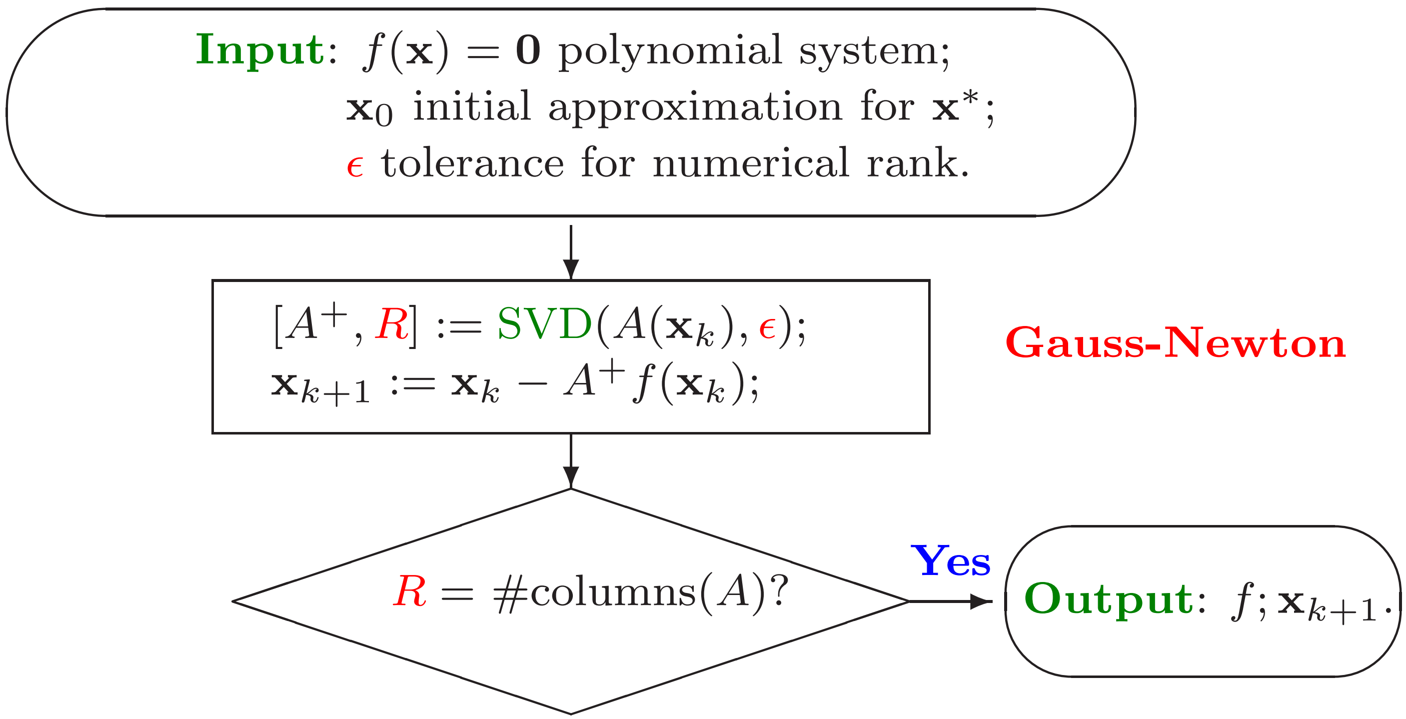
Newton's Method with Deflation

Input: $f(\mathbf{x}) = \mathbf{0}$ polynomial system;
$\mathbf{x}_0$ initial approximation for $\mathbf{x}^*$;
ϵ tolerance for numerical rank.

$[A^+, R] := \mathrm{SVD}(A(\mathbf{x}_k), \epsilon);$
$\mathbf{x}_{k+1} := \mathbf{x}_k - A^+ f(\mathbf{x}_k);$

Gauss-Newton

Newton's Method with Deflation



Newton's Method with Deflation

Input: $f(\mathbf{x}) = \mathbf{0}$ polynomial system;
$\mathbf{x}_0$ initial approximation for $\mathbf{x}^*$;
ϵ tolerance for numerical rank.

Gauss-Newton

$[A^+, R] := \mathrm{SVD}(A(\mathbf{x}_k), \epsilon)$;
$\mathbf{x}_{k+1} := \mathbf{x}_k - A^+ f(\mathbf{x}_k)$;

$R = \#\mathrm{columns}(A)$?

Yes → **Output**: $f; \mathbf{x}_{k+1}$.

No ↓

Deflation Step

$$f := \mathbf{Dfl}(f)(\mathbf{x}, \boldsymbol{\lambda}) = \left\{ \begin{array}{c} f(\mathbf{x}) = \mathbf{0} \\ G(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0} \end{array} \right. ;$$

$\widehat{\boldsymbol{\lambda}} := \mathbf{LeastSquares}(G(\mathbf{x}_{k+1}, \boldsymbol{\lambda}))$;
$k := k + 1; \quad \mathbf{x}_k := (\mathbf{x}_k, \widehat{\boldsymbol{\lambda}})$;

cyclic 9-roots revisited

$$\text{Recall}: \quad f(\mathbf{x}) = \left\{ \begin{array}{l} \displaystyle f_i = \sum_{j=0}^{8} \prod_{k=1}^{i} x_{(k+j) \bmod 9} = 0, \quad i = 1, 2, \ldots, 8 \\ f_9 = x_0 x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 - 1 = 0 \end{array} \right.$$

has 164 solutions of multiplicity 4.

One deflation suffices to restore quadratic convergence.

An average **condition number** drops from 1.8E+9 to 5.6E+2.

$\rightarrow$ **deflation reconditions the problem**

A system with a cluster of solutions is close to a system with a multiple root.

12 Lines Tangent to 4 Spheres revisited

Continuation methods find 24 solutions, clustered in groups of 4.

The rank at all solutions is 4, corank is 2.

One deflation suffices to restore quadratic convergence.

An average **condition number** drops from 3.4E+8 to 1.1E+2.

We can compute the solutions
with **accuracy close to machine precision**,
on a system with approximate coefficients,
given with double float precision.

Numerical Results (double float)

System	n	m	D	corank$(A(\mathbf{x}))$	Inverse Condition#	#Digits
baker1	2	2	1	$1 \rightarrow 0$	1.7e-08 $\rightarrow$ 3.8e-01	$9 \rightarrow 24$
cbms1	3	11	1	$3 \rightarrow 0$	4.2e-05 $\rightarrow$ 5.0e-01	$5 \rightarrow 20$
cbms2	3	8	1	$3 \rightarrow 0$	1.2e-08 $\rightarrow$ 5.0e-01	$8 \rightarrow 18$
mth191	3	4	1	$2 \rightarrow 0$	1.3e-08 $\rightarrow$ 3.5e-02	$7 \rightarrow 13$
decker1	2	3	2	$1 \rightarrow 1 \rightarrow 0$	3.4e-10 $\rightarrow$ 2.6e-02	$6 \rightarrow 11$
decker2	2	4	3	$1 \rightarrow 1 \rightarrow 1 \rightarrow 0$	4.5e-13 $\rightarrow$ 6.9e-03	$5 \rightarrow 16$
decker3	2	2	1	$1 \rightarrow 0$	4.6e-08 $\rightarrow$ 2.5e-02	$8 \rightarrow 17$
ojika1	2	3	2	$1 \rightarrow 1 \rightarrow 0$	9.3e-12 $\rightarrow$ 4.3e-02	$5 \rightarrow 12$
ojika2	3	2	1	$1 \rightarrow 0$	3.3e-08 $\rightarrow$ 7.4e-02	$6 \rightarrow 14$
ojika3	3	2	1	$1 \rightarrow 0$	1.7e-08 $\rightarrow$ 9.2e-03	$7 \rightarrow 15$
		4	1	$2 \rightarrow 0$	6.5e-08 $\rightarrow$ 8.0e-02	$6 \rightarrow 13$
ojika4	3	3	2	$1 \rightarrow 1 \rightarrow 0$	1.9e-13 $\rightarrow$ 2.4e-04	$6 \rightarrow 11$
cyclic9	9	4	1	$2 \rightarrow 0$	5.6e-10 $\rightarrow$ 1.8e-03	$5 \rightarrow 15$

Some Remaining Questions ...

1) Does the deflation algorithm terminate?

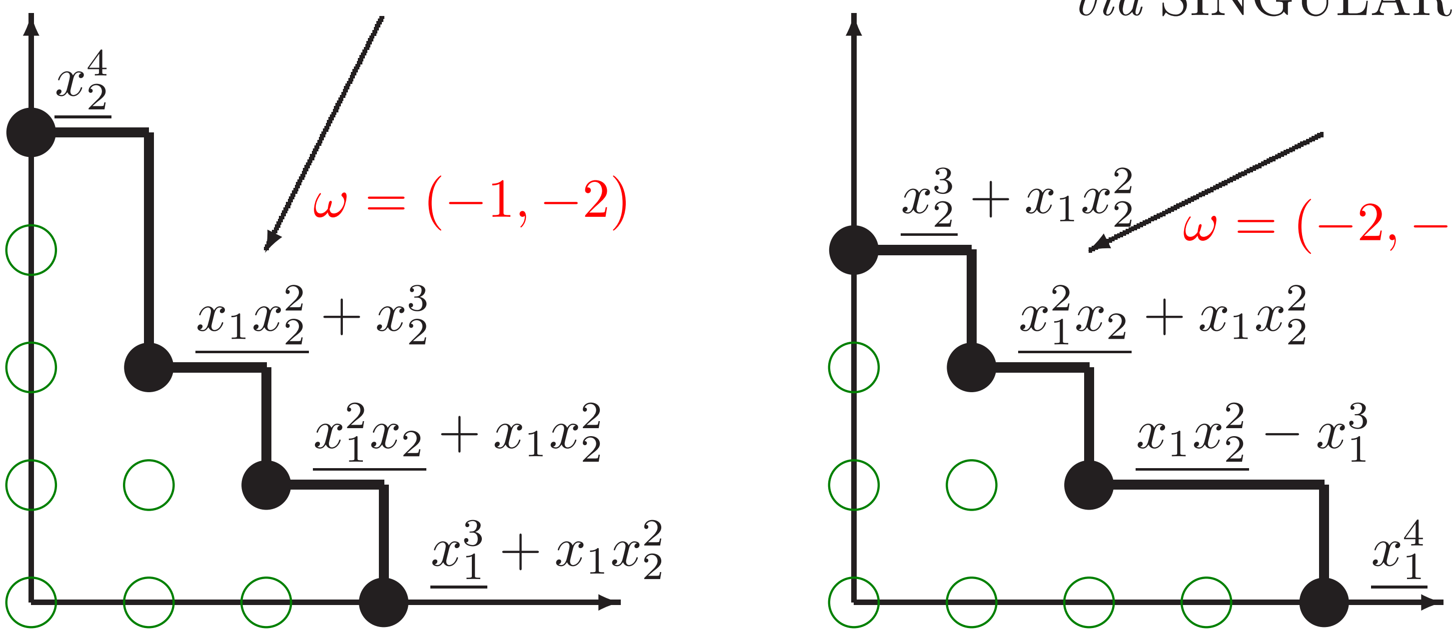
2) Is the deflation algorithm efficient?

3) How to compute the multiplicity of a root?

Two Staircases with Different Local Ordering

Example: $I = \langle x_1^3 + x_1x_2^2, x_1x_2^2 + x_2^3, x_1^2x_2 + x_1x_2^2\rangle$ in the ring $\mathbb{Q}[x_1, x_2]$, $\mathbf{x}^* = \mathbf{0}$, ω defines a local monomial order.

via SINGULAR



●: monomials generating $\mathbf{in}_\omega(I)$ ○: standard monomials

$\boxed{\textbf{\#standard monomials = multiplicity of } \mathbf{x}^* = \mathbf{7}}$

Multiplicity of an Isolated Zero via Duality

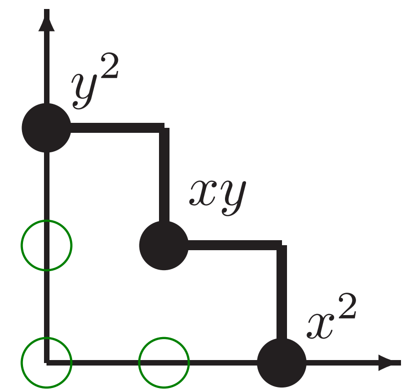
Analogy with Univariate Case: z_0 is $\boldsymbol{m}$-fold zero of $f(x) = 0$:

$$\underbrace{f(z_0) = 0, \frac{\partial f}{\partial x}(z_0) = 0, \frac{\partial^2 f}{\partial x^2}(z_0) = 0, \ldots, \frac{\partial^{m-1} f}{\partial x^{m-1}}(z_0) = 0}\quad .$$

$\boldsymbol{m}$ = number of linearly independent polynomials annihilating z_0

The dual space D_0 at $\mathbf{z}_0$ is spanned by $\boldsymbol{m}$ linear independent differentiation functionals annihilating $\mathbf{z}_0$.

Consider again $$f(x,y) = \begin{cases} x^2 = 0 \\ xy = 0 \\ y^2 = 0 \end{cases}$$



The **multiplicity of $\mathbf{z}_0 = (0,0)$ is 3** because

$$D_0 = \text{span}\{\partial_{00}[\mathbf{z}_0], \partial_{10}[\mathbf{z}_0], \partial_{01}[\mathbf{z}_0]\}, \text{with } \partial_{ij}[\mathbf{z}_0] = \frac{1}{i!j!}\frac{\partial^{i+j}}{\partial x^i \partial y^j} f(\mathbf{z}_0).$$

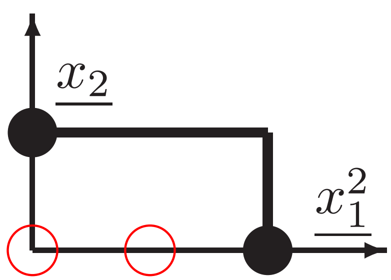
Standard Bases and Dual Space

Consider $\begin{cases} x_1^2 + 2x_2^2 - 2x_2 = 0 \\ x_1 x_2^2 - x_1 x_2 = 0 \\ x_2^3 - 2x_2^2 + x_2 = 0 \end{cases}$ from Möller-Stetter (1995).

$\mathbf{z}_0 = (0,0)$

$\boldsymbol{m_0 = 2}$

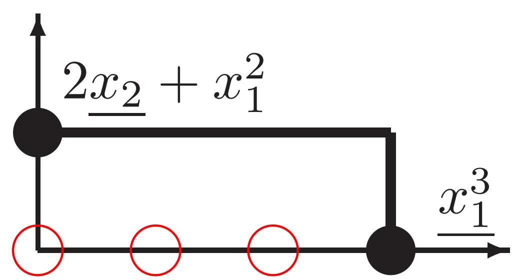
$D_0 = \mathrm{span}\{\partial_{00}, \partial_{10}\}$



$\mathbf{z}_1 = (0,1)$ (shift to (0,0))

$\boldsymbol{m_1 = 3}$

$D_1 = \mathrm{span}\{\partial_{00}, \partial_{10}, 2\partial_{20} - \partial_{01}\}$

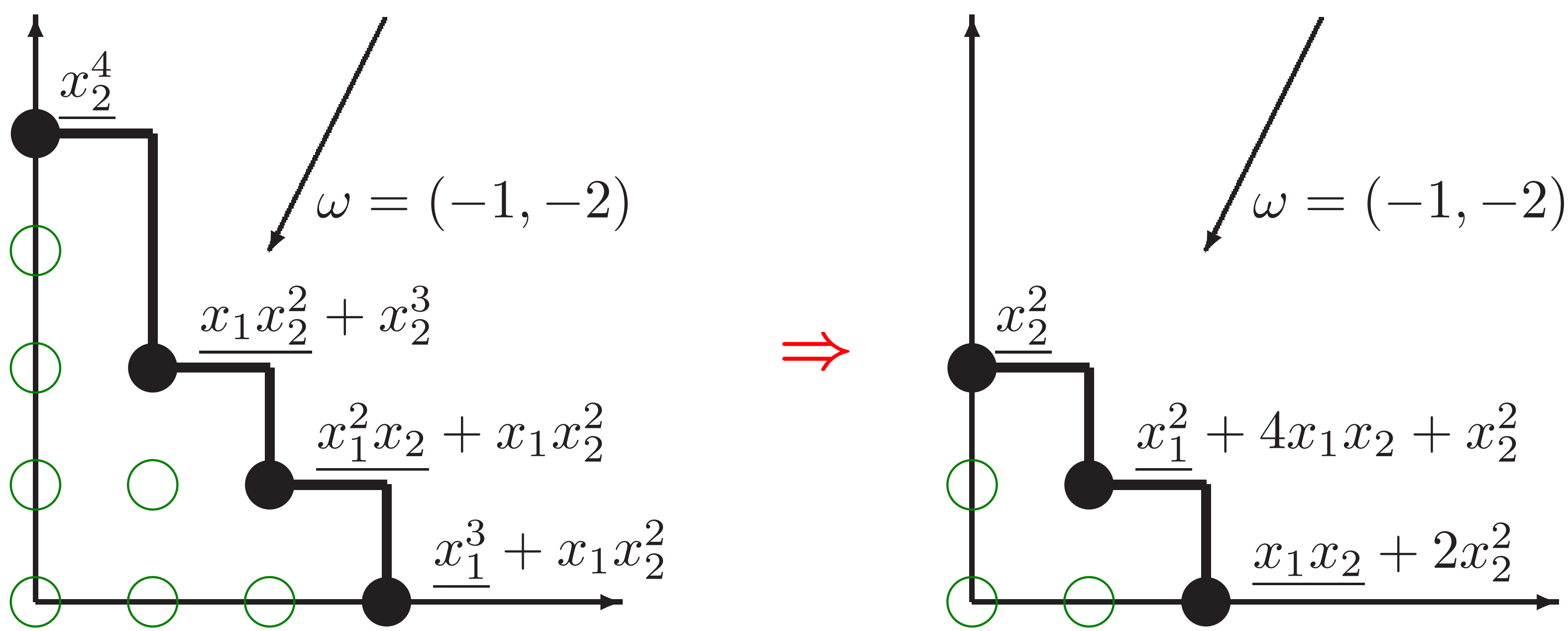


$$D[I] = D_0 \cup D_1$$

Effect of Deflation on the Staircase

$I = \langle f_1 = x_1^3 + x_1 x_2^2, f_2 = x_1 x_2^2 + x_2^3, f_3 = x_1^2 x_2 + x_1 x_2^2 \rangle$, $\boldsymbol{\lambda} = (1, 1)$.

$J = \langle f_1, f_2, f_3, \frac{\partial f_1}{\partial x_1} + \frac{\partial f_1}{\partial x_2}, \frac{\partial f_2}{\partial x_1} + \frac{\partial f_2}{\partial x_2}, \frac{\partial f_3}{\partial x_1} + \frac{\partial f_3}{\partial x_2} \rangle$ is a **deflation** of I.



●: monomials generating $\mathbf{in}_\omega(I)$ ○: standard monomials

$\boldsymbol{m = 7}$ ⟶ $\boldsymbol{m = 3}$

deflation

One Deflation Step with fixed $\boldsymbol{\lambda}$

- Assume $\text{corank}(A(\mathbf{x}^*)) = 1$.
  (reduce to this case with random combinations of columns)
- Let $\boldsymbol{\lambda} \in \ker(A(\mathbf{x}^*))$, $\boldsymbol{\lambda} \neq \mathbf{0}$,
  then for $g_i(\mathbf{x}) = \boldsymbol{\lambda} \cdot \nabla f_i = \displaystyle\sum_{j=1}^{n} \lambda_j \frac{\partial f_i}{\partial x_j}(x)$, we have: $g_i(\mathbf{x}^*) = \mathbf{0}$.

<u>**Theorem**</u>:

The augmented system $\left\{ \begin{array}{l} \boldsymbol{f_1 = f_2 = \cdots = f_N = 0} \\ \boldsymbol{g_1 = g_2 = \cdots = g_N = 0} \end{array} \right.$

has x* as isolated root of lower multiplicity.

Proposition: Suppose $m > 1$ and let $g \in \mathcal{B}$, a reduced standard basis of I with respect to a local monomial ordering $\leq$, such that $g = x_i^d +$ lower order terms, for $i \in \{1, 2, \ldots, n\}$ and $d > 1$. Then $I' = I + \langle \frac{\partial g}{\partial x_i} \rangle$ is a **deflation** of I.

Lemma: Take a nonzero vector $\boldsymbol{\lambda} \in \ker A(\mathbf{0}) \subset \mathbb{C}^n$ and let $\mathbf{x} = T(\mathbf{y})$ be a linear coordinate transformation such that

$$y_i = \lambda_i x_1 + \sum_{j=2}^{n} \mu_{ij} x_j, \qquad \text{for } i = 1, 2, \ldots, n,$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ are the new variables and $[\boldsymbol{\lambda}, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_n]$ is a nonsingular matrix.
Let $T(I) = \{f(T(\mathbf{y})) \mid f \in I\} = \langle f_1(T(\mathbf{y})), \ldots, f_N(T(\mathbf{y}))\rangle$ be the ideal after the change of coordinates.
Then $\partial_1 T(I) = \left\{ \frac{\partial f}{\partial y_1} \mid f \in T(I) \right\}$ leads to a **deflation** of $T(I)$.

One Deflation Step with indeterminate $\boldsymbol{\lambda}$

- Still assuming $\text{corank}(A(\mathbf{x}^*)) = 1$.
- Denote $G(\mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} g_i(\mathbf{x}, \boldsymbol{\lambda}) = \boldsymbol{\lambda} \cdot \nabla f_i(\mathbf{x}) = \mathbf{0} \\ \langle \mathbf{h}, \boldsymbol{\lambda} \rangle = h_1\lambda_1 + h_2\lambda_2 + \cdots + h_n\lambda_n = 1. \end{cases}$

Theorem:

Let $\mathbf{x}^* \in \mathbb{C}^n$ be an isolated singular root of $f(\mathbf{x}) = \mathbf{0}$ with multiplicity m. There exists a unique $\boldsymbol{\lambda}^*$ such that $\begin{cases} f(\mathbf{x}) = \mathbf{0} \\ G(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0} \end{cases}$ has $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ as isolated root of multiplicity strictly less than m.

Proof: Consider $G(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$ in the local ring $R_* = \mathbb{C}[\mathbf{x}, \boldsymbol{\lambda}]_{(\mathbf{x}^*, \boldsymbol{\lambda}^*)}$. Because $G(\mathbf{x}, \boldsymbol{\lambda})$ is linear in $\boldsymbol{\lambda}$, specializing $\mathbf{x} = \mathbf{x}^*$ turns $G(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$ into a linear system with unique solution $\boldsymbol{\lambda}^*$.

Using row operations in R_*, reduce $G(\mathbf{x}, \boldsymbol{\lambda})$ to the form :
$$\left\{ \begin{array}{c} \lambda_1 = a_1(\mathbf{x}) \\ \vdots \\ \lambda_n = a_n(\mathbf{x}) \end{array} \right.$$

where $a_i(\mathbf{x})$ are rational expressions ($a_i(\mathbf{x}^*) = \lambda_i^*$).

$$\begin{array}{c} \text{multiplicity} \\ \text{of } \mathbf{x}^* \text{ in} \end{array} \left\{ \begin{array}{c} f(\mathbf{x}) = 0 \\ G(\mathbf{x}, \boldsymbol{\lambda}) = 0 \end{array} \right. \Leftrightarrow \begin{array}{c} \text{multiplicity} \\ \text{of } \mathbf{x}^* \text{ in} \end{array} \left\{ \begin{array}{c} f(\mathbf{x}) = 0 \\ G(\mathbf{x}, \boldsymbol{\lambda}^*) = 0 \end{array} \right.$$

local ring $\mathbb{C}[\mathbf{x}, \boldsymbol{\lambda}]_{(\mathbf{x}^*, \boldsymbol{\lambda}^*)}$ local ring $\mathbb{C}[\mathbf{x}]_{(\mathbf{x}^*)}$

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)

Avoiding Expression Swell

Evaluation of $A(\mathbf{x})B$: for efficiency we must first replace $\mathbf{x}$ by values ***before*** the matrix multiplication.

Triangular block structure of Jacobian matrix: for example:

$$A_2(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \begin{bmatrix} A & \mathbf{0} & \mathbf{0} \\ \left(\frac{\partial A}{\partial \mathbf{x}}\right) B_1 \boldsymbol{\lambda}_1 & AB_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_1 & \mathbf{0} \\ \left(\frac{\partial A_1}{\partial \mathbf{x}}\right) B_2 \boldsymbol{\lambda}_2 & \left(\frac{\partial A_1}{\partial \boldsymbol{\lambda}_1}\right) B_2 \boldsymbol{\lambda}_2 & A_1 B_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{h}_2 \end{bmatrix}.$$

Multipliers occur linearly: compute derivatives only with respect to $\mathbf{x}$, not with respect to $\boldsymbol{\lambda}$.

Structure of the Deflated Systems

At stage k in the deflation:

$$f_k(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}, \boldsymbol{\lambda}_k) = \left\{ \begin{array}{rcl} f_{k-1}(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}) & = & 0 \\ A_{k-1}(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}) B_k \boldsymbol{\lambda}_k & = & 0 \\ \mathbf{h}_k \boldsymbol{\lambda}_k & = & 1, \end{array} \right.$$

$f_0 = f$, $A_0 = A$, $R_k = \text{rank}(A_{k-1}(\mathbf{z}_0))$, $R_k + 1$ multipliers in $\boldsymbol{\lambda}_k$.

random vector $\mathbf{h}_k \in \mathbb{C}^{R_k+1}$ and n_{k-1}-by-$(R_k + 1)$ matrix B_k

$$\begin{array}{rcl} \#\text{rows in } A_k & : & N_k = 2N_{k-1} + 1, \quad N_0 = N, \\ \#\text{columns in } A_k & : & n_k = n_{k-1} + R_k + 1, \quad n_0 = n. \end{array}$$

Multiplication of the polynomial matrix A_{k-1} with the random matrix B_k and vector of $R_k + 1$ multiplier variables $\boldsymbol{\lambda}_k$ is expensive!

Structure of the Jacobian Matrices

At stage k in the deflation:

$$f_k(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}, \boldsymbol{\lambda}_k) = \begin{cases} f_{k-1}(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}) & = & 0 \\ A_{k-1}(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}) B_k \boldsymbol{\lambda}_k & = & 0 \\ \mathbf{h}_k \boldsymbol{\lambda}_k & = & 1. \end{cases}$$

Jacobian matrix at the kth deflation:

$$A_k(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}, \boldsymbol{\lambda}_k) = \begin{bmatrix} A_{k-1} & \mathbf{0} \\ \left[\frac{\partial A_{k-1}}{\partial \mathbf{x}} \, \frac{\partial A_{k-1}}{\partial \boldsymbol{\lambda}_1} \cdots \frac{\partial A_{k-1}}{\partial \boldsymbol{\lambda}_{k-1}} \right] B_k \boldsymbol{\lambda}_k & A_{k-1} B_k \\ \mathbf{0} & \mathbf{h}_k \end{bmatrix}.$$

The multiplier variables $\boldsymbol{\lambda}_i$, $i = 1, 2, \ldots, k$, occur linearly,

we compute derivatives only with respect to the original variables $\mathbf{x}$.

Derivatives of Jacobian matrices

Define $\frac{\partial A}{\partial \mathbf{x}} = \left[\frac{\partial A}{\partial x_1} \; \frac{\partial A}{\partial x_2} \; \cdots \; \frac{\partial A}{\partial x_n} \right]$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$,

where $\frac{\partial A}{\partial x_k} = \left[\frac{\partial a_{ij}}{\partial x_k} \right]$ for $A = [a_{ij}(\mathbf{x})]_{j,k \in \{1,2,\ldots,n\}}^{i \in \{1,2,\ldots,N\}}$.

Natural tree structure:
(to compute $\frac{\partial^2 A}{\partial \mathbf{x}^2}$)

$$A(x_1, x_2)$$

$$\frac{\partial A}{\partial x_1} \qquad\qquad \frac{\partial A}{\partial x_2}$$

$$\frac{\partial^2 A}{\partial x_1^2} \qquad \frac{\partial^2 A}{\partial x_2 \partial x_1} = \frac{\partial^2 A}{\partial x_1 \partial x_2} \qquad \frac{\partial^2 A}{\partial x_2^2}$$

Complexity of $\frac{\partial^k A}{\partial \mathbf{x}^k} \neq O(n^k)$,

but $O(\#\text{monomials of degree } k \text{ in } n \text{ variables})$.

e.g.: $k = 10$, $n = 3$:
$3^{10} = 59049 \gg 66$

Column Format of Jacobian Matrices

Jacobian matrix at the kth deflation:

$$A_k(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}, \boldsymbol{\lambda}_k) = \left[\begin{array}{cc} A_{k-1} & \mathbf{0} \\ \left[\frac{\partial A_{k-1}}{\partial \mathbf{x}} \; \frac{\partial A_{k-1}}{\partial \boldsymbol{\lambda}_1} \; \cdots \; \frac{\partial A_{k-1}}{\partial \boldsymbol{\lambda}_{k-1}} \right] B_k \boldsymbol{\lambda}_k & A_{k-1} B_k \\ \mathbf{0} & \mathbf{h}_k \end{array} \right].$$

Tree with k children:

$$\begin{tikzcd} & & A_k(\mathbf{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}, \boldsymbol{\lambda}_k) \arrow[dll] \arrow[dl] \arrow[d] \arrow[drr] & & \\ A_{k-1} & \frac{\partial A_{k-1}}{\partial \mathbf{x}} & \frac{\partial A_{k-1}}{\partial \boldsymbol{\lambda}_1} & \cdots & \frac{\partial A_{k-1}}{\partial \boldsymbol{\lambda}_{k-1}} \end{tikzcd}$$

Unwinding the Multipliers

$$A_2(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \begin{bmatrix} A_1 & \mathbf{0} \\ \left[\frac{\partial A_1}{\partial \mathbf{x}} \;\; \frac{\partial A_1}{\partial \boldsymbol{\lambda}_1} \right] B_2 \boldsymbol{\lambda}_2 & A_1 B_2 \\ \mathbf{0} & \mathbf{h}_2 \end{bmatrix}$$

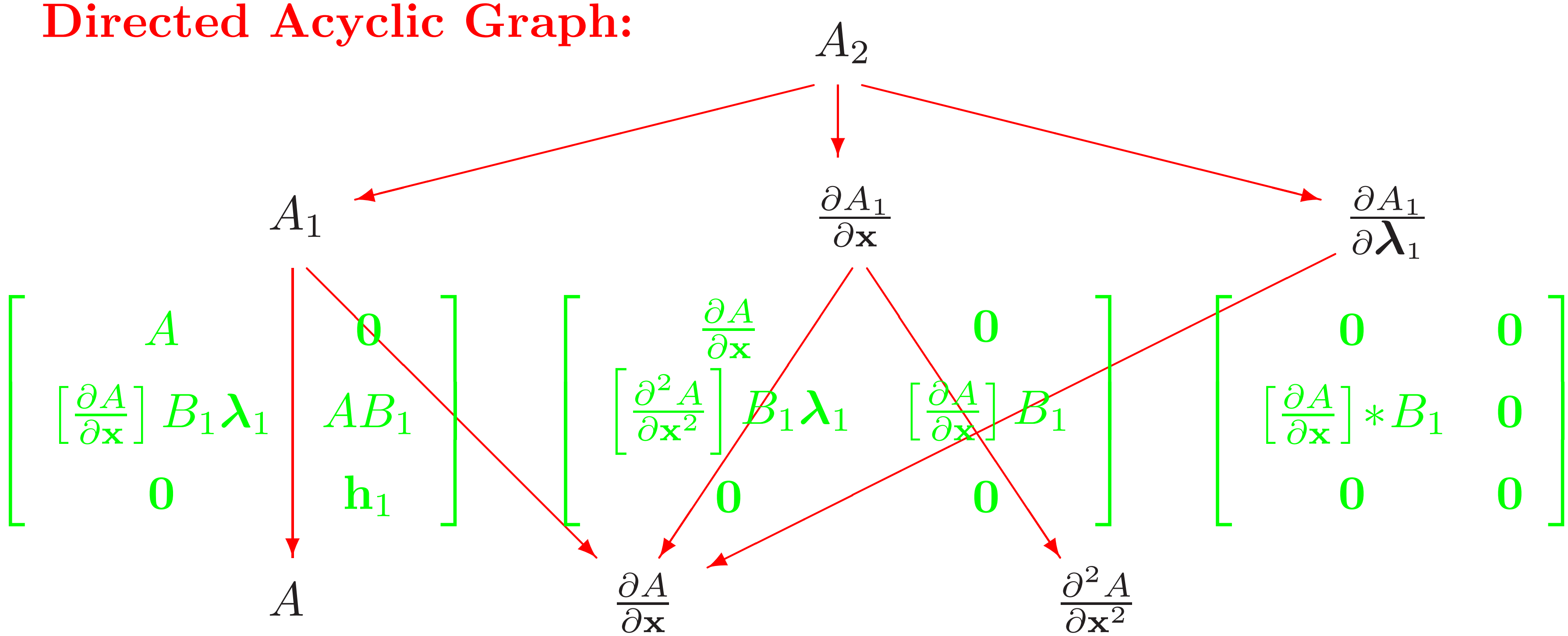
$$\begin{tikzcd} & A_2 \arrow[dl] \arrow[d] \arrow[dr] & \\ A_1 & \frac{\partial A_1}{\partial \mathbf{x}} & \frac{\partial A_1}{\partial \boldsymbol{\lambda}_1} \end{tikzcd}$$

$$\begin{bmatrix} A & \mathbf{0} \\ \left[\frac{\partial A}{\partial \mathbf{x}} \right] B_1 \boldsymbol{\lambda}_1 & A B_1 \\ \mathbf{0} & \mathbf{h}_1 \end{bmatrix} \qquad \begin{bmatrix} \frac{\partial A}{\partial \mathbf{x}} & \mathbf{0} \\ \left[\frac{\partial^2 A}{\partial \mathbf{x}^2} \right] B_1 \boldsymbol{\lambda}_1 & \left[\frac{\partial A}{\partial \mathbf{x}} \right] B_1 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \left[\frac{\partial A}{\partial \mathbf{x}} \right] * B_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Unwinding the Multipliers

$$A_2(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \begin{bmatrix} A_1 & \mathbf{0} \\ \left[\frac{\partial A_1}{\partial \mathbf{x}} \;\; \frac{\partial A_1}{\partial \boldsymbol{\lambda}_1}\right] B_2 \boldsymbol{\lambda}_2 & A_1 B_2 \\ \mathbf{0} & \mathbf{h}_2 \end{bmatrix}$$

Directed Acyclic Graph:



The Operator $*$

For a matrix B: $\left[\frac{\partial A}{\partial \mathbf{x}}\right] B = \left[\frac{\partial A}{\partial x_1} B \;\; \frac{\partial A}{\partial x_2} B \;\; \cdots \;\; \frac{\partial A}{\partial x_n} B\right]$.

However, $\frac{\partial}{\partial \boldsymbol{\lambda}}\left(\left[\frac{\partial A}{\partial \mathbf{x}}\right] B \boldsymbol{\lambda}\right) = \left[\underbrace{\frac{\partial A}{\partial \mathbf{x}} \mathbf{b}_1}_{\partial \lambda_1} \;\; \underbrace{\frac{\partial A}{\partial \mathbf{x}} \mathbf{b}_2}_{\partial \lambda_2} \;\; \cdots \;\; \underbrace{\frac{\partial A}{\partial \mathbf{x}} \mathbf{b}_m}_{\partial \lambda_m}\right]$,

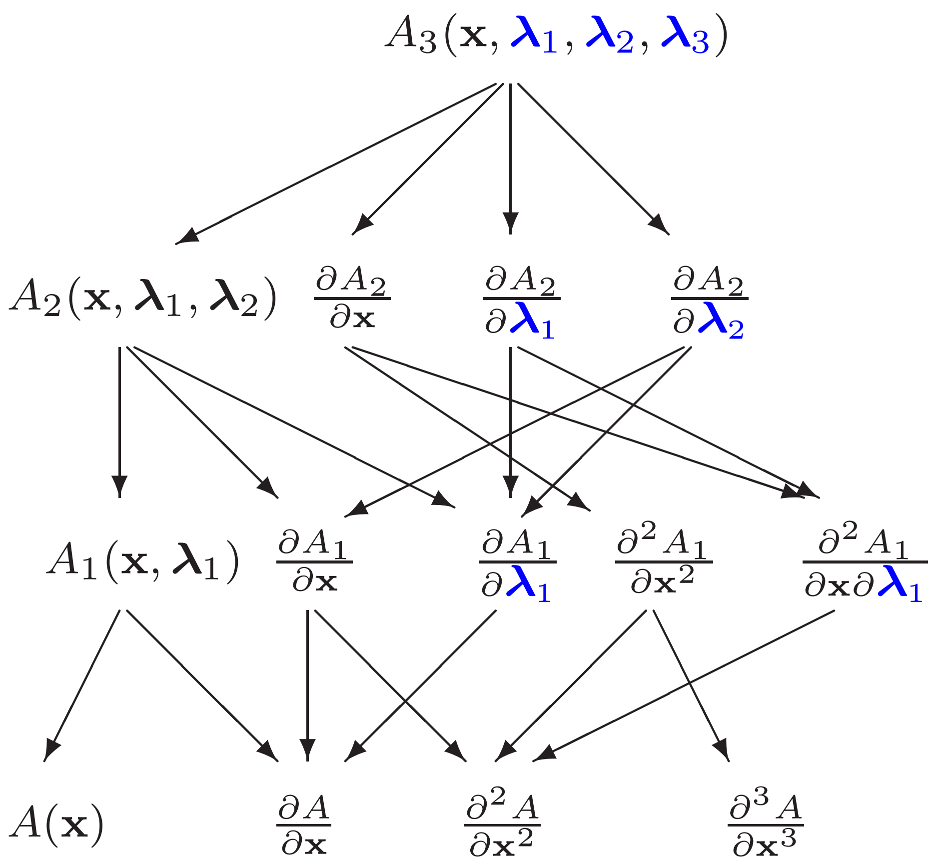
$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m), \; B = [\mathbf{b}_1 \;\; \mathbf{b}_2 \;\; \cdots \mathbf{b}_m].$$

Unlike scalar differentiation: $\frac{\partial}{\partial \boldsymbol{\lambda}}\left(\left[\frac{\partial A}{\partial \mathbf{x}}\right] B \boldsymbol{\lambda}\right) \neq \left[\frac{\partial A}{\partial \mathbf{x}}\right] B$.

With the operator $*$ we permute $\left[\frac{\partial A}{\partial \mathbf{x}}\right] B$ into $\frac{\partial}{\partial \boldsymbol{\lambda}}\left(\left[\frac{\partial A}{\partial \mathbf{x}}\right] B \boldsymbol{\lambda}\right)$.

So we have: $\frac{\partial}{\partial \boldsymbol{\lambda}}\left(\left[\frac{\partial A}{\partial \mathbf{x}}\right] B \boldsymbol{\lambda}\right) = \left[\frac{\partial A}{\partial \mathbf{x}}\right] * B$.

A Directed Acyclic Graph of Derivative Operators



Growth of Number of Nodes in DAG

The growth of the number of nodes in the directed acyclic graph, for increasing deflations k:

k	1	2	3	4	5	6	7	8	9	10
#nodes	3	7	14	26	46	79	133	221	364	596

#nodes ranges between $O(k^2)$ and $O(k^3)$

Computing the Multiplicity Structure

following B.H. Dayton and Z. Zeng

Looking for differentiation functionals $d[\mathbf{z}_0] = \sum_{\mathbf{a}} c_{\mathbf{a}} \partial_{\mathbf{a}}[\mathbf{z}_0]$,

$$\text{with} \quad \partial_{\mathbf{a}}[\mathbf{z}_0](p) = \frac{1}{a_1!a_2!\cdots a_n!} \left(\frac{\partial^{a_1+a_2+\cdots+a_n}}{\partial x_1^{a_1} \partial x_2^{a_2} \cdots \partial x_n^{a_n}} p \right) (\mathbf{z}_0).$$

Membership criterium for $d[\mathbf{z}_0]$:

$$\boldsymbol{d[\mathbf{z}_0] \in D_0 \Leftrightarrow d[\mathbf{z}_0](pf_i) = 0, \forall p \in \mathbb{C}[\mathbf{x}], i = 1, 2, \ldots, N.}$$

To turning this criterium into an **algorithm**, observe:

1. since $d[\mathbf{z}_0]$ is linear, restrict p to $\mathbf{x}^{\mathbf{k}} = x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}$; and
2. limit degrees $k_1 + k_2 + \cdots + k_n \leq a_1 + a_2 + \cdots + a_n$,
   as $\mathbf{z}_0 = \mathbf{0}$ vanishes trivially if not annihilated by $\partial_{\mathbf{a}}$.

Computing the Multiplicity Structure – An Example

$f_1 = x_1 - x_2 + x_1^2,\ f_2 = x_1 - x_2 + x_2^2$ following B.H. Dayton and Z. Zeng

		$\lvert a\rvert=0$	$\lvert a\rvert=1$		$\lvert a\rvert=2$			$\lvert a\rvert=3$			
		∂_{00}	∂_{10}	∂_{01}	∂_{20}	∂_{11}	∂_{02}	∂_{30}	∂_{21}	∂_{12}	∂_{03}
	f_1	0	1	–1	1	0	0	0	0	0	0
S_1	f_2	0	1	–1	0	0	1	0	0	0	0
	$x_1 f_1$	0	0	0	1	–1	0	1	0	0	0
	$x_1 f_2$	0	0	0	1	–1	0	0	0	1	0
	$x_2 f_1$	0	0	0	0	1	–1	0	1	0	0
S_2	$x_2 f_2$	0	0	0	0	1	–1	0	0	0	1
	$x_1^2 f_1$	0	0	0	0	0	0	1	–1	0	0
	$x_1^2 f_2$	0	0	0	0	0	0	1	–1	0	0
	$x_1 x_2 f_1$	0	0	0	0	0	0	0	1	–1	0
	$x_1 x_2 f_2$	0	0	0	0	0	0	0	1	–1	0
	$x_2^2 f_1$	0	0	0	0	0	0	0	0	1	–1
S_3	$x_2^2 f_2$	0	0	0	0	0	0	0	0	1	–1

$\text{Nullity}(S_2) = \text{Nullity}(S_3) \Rightarrow$ stop algorithm

$D_0 = \text{span}\{\ \partial_{00}, \partial_{10} + \partial_{01}, -\partial_{10} + \partial_{20} + \partial_{11} + \partial_{02}\ \} \Rightarrow$ multiplicity $= 3$

cyclic 9-roots once more

Recall:

$$f(\mathbf{x}) = \left\{ \begin{array}{l} f_i = \displaystyle\sum_{j=0}^{8} \prod_{k=1}^{i} x_{(k+j) \bmod 9} = 0, \quad i = 1, 2, \ldots, 8 \\ f_9 = x_0 x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 - 1 = 0 \end{array} \right.$$

has 164 solutions of **multiplicity 4**.

Running the algorithm of Dayton and Zeng:

$$\boldsymbol{H[1] = 1, H[2] = 2, H[3] = 1, H[4] = 0,}$$

$$\textbf{with } \boldsymbol{H[i] = \textbf{Nullity}(S_i) - \textbf{Nullity}(S_{i-1}), i > 0,}$$

so we compute **locally** the **multiplicity as 4**.

12 Lines Tangent to 4 Spheres once more

With deflation 6 solutions of **multiplicity 4** are computed accurately, i.e.: the cluster radius is close to machine precision.

Running the algorithm of Dayton and Zeng:

$$\boldsymbol{H[1] = 1, H[2] = 2, H[3] = 1, H[4] = 0,}$$
$$\textbf{with } \boldsymbol{H[i] = \mathbf{Nullity}(S_i) - \mathbf{Nullity}(S_{i-1}), i > 0,}$$

so we compute **locally** the **multiplicity as 4**.

Concluding Remarks

software at `http://www.math.uic.edu/~jan/download.html`

Work in Progress

- higher order deflations
- combine Stetter-Thallinger with Dayton-Zeng
- integrate into endgame of solver
- singular positive dimensional solution sets
- local dimension test