The Calculation Of The Degree Of An Approximate Greatest Common Divisor Of Two Polynomials

Joab R. Winkler

Department of Computer Science
The University of Sheffield
United Kingdom

Charles University, Prague. May 2009
CONTENTS:

• Greatest common divisor and approximate greatest common divisors

• The Sylvester resultant matrix
  – Properties of the Sylvester resultant matrix
  – Subresultant matrices of the Sylvester resultant matrix

• Polynomial scaling
  – Scaling by the geometric mean
  – Relative scaling of the polynomials
• The calculation of the degree of an approximate greatest common divisor
  – The principle of maximum likelihood
  – The angle between subspaces
  – The polynomial form of a common divisor

• Results

• Summary
1. GREATEST COMMON DIVISOR AND APPROXIMATE GREATEST COMMON DIVISORS

- The greatest common divisor (GCD) of two exact polynomials is defined.

Two or more inexact polynomials have an approximate greatest common divisor (AGCD), but not a GCD:

- AGCDs arise in many practical applications, including computing theory, image processing, control theory, and the computation of the roots of a polynomial.

- The given inexact polynomials are with high probability coprime.

- If there is prior information that the theoretically exact forms of the inexact polynomials have a non-constant GCD, it is necessary to perturb their inexact forms such that these perturbed forms have a non-constant common divisor.

- The common divisor of maximum degree is called an AGCD because it is approximate with respect to the given inexact and coprime polynomials.
If $\hat{f}(y)$ and $\hat{g}(y)$ are exact and all computations are performed exactly, their GCD can be computed by the Sylvester resultant matrix $S(\hat{f}, \hat{g})$.

Polynomials are rarely known exactly, and so the given data is

\[
f(y) = \hat{f}(y) + \delta \hat{f}(y) \quad \text{and} \quad g(y) = \hat{g}(y) + \delta \hat{g}(y)
\]

**Example 1.1** The GCD of the polynomials

\[
\hat{f}(y) = (y - 1)^2 \quad \text{and} \quad \hat{g}(y) = (y - 1)(y - 2)
\]

is $(y - 1)$, but if $\delta \hat{f}(y)$ and $\delta \hat{g}(y)$ are random polynomials, the inexact polynomials

\[
f(y) = \hat{f}(y) + \delta \hat{f}(y) \quad \text{and} \quad g(y) = \hat{g}(y) + \delta \hat{g}(y)
\]

are, with high probability, coprime.
There may not be any AGCDs that satisfy a given criterion, or there may be several AGCDs, all of the same degree, for different realisations of the perturbations $\delta \hat{f}(y)$ and $\delta \hat{g}(y)$.

Criteria for the calculation of an AGCD include:

- Calculate the minimum amount that the given inexact polynomials must be perturbed, such that the perturbed polynomials have an AGCD of degree greater than or equal to one.
- Calculate the family of AGCDs that are within a specified distance of the given inexact polynomials.
- Calculate the nearest AGCD that is of given degree.
Example 1.2 Consider the pairs of perturbed polynomials

\[ f_1(y) = \hat{f}_1(y) + \delta \hat{f}_1(y) \quad \text{and} \quad g_1(y) = \hat{g}_1(y) + \delta \hat{g}_1(y) \]

and

\[ f_2(y) = \hat{f}_2(y) + \delta \hat{f}_2(y) \quad \text{and} \quad g_2(y) = \hat{g}_2(y) + \delta \hat{g}_2(y) \]

The perturbations may be such that

\[ \text{degree AGCD} (f_1(y), g_1(y)) = \text{degree AGCD} (f_2(y), g_2(y)) \]

but

\[ \text{AGCD} (f_1(y), g_1(y)) \neq \text{AGCD} (f_2(y), g_2(y)) \]

- Different noisy realisations of two theoretically exact polynomials yield different AGCDs, of possibly different degrees.
- An AGCD is not unique - compare with the GCD.
2. THE SYLVESTER RESULTANT MATRIX

Consider the polynomials

\[ f(y) = \sum_{i=0}^{m} a_i y^i \quad \text{and} \quad g(y) = \sum_{i=0}^{n} b_i y^i \]

where \( a_m, b_n \neq 0 \).

- Each entry of the Sylvester resultant matrix \( S(f, g) \in \mathbb{R}^{(m+n) \times (m+n)} \) contains either a coefficient \( a_i \) or a coefficient \( b_i \).
- \( S(f, g) \) is used extensively in GCD computations.
\[ S(f, g) = \begin{bmatrix}
  a_m \\
  a_{m-1} & a_m \\
  \vdots & \ddots & \ddots \\
  a_0 & \ddots & a_m \\
  a_0 & \ddots & a_{m-1} \\
  \vdots & \ddots & \ddots \\
  a_0 & & \ddots & \ddots \\
  \end{bmatrix} \]

- The coefficients \( a_i \) of \( f(y) \) occupy the first \( n \) columns
- The coefficients \( b_i \) of \( g(y) \) occupy the last \( m \) columns
The Sylvester resultant matrix is linear, which makes it easy to use structure preserving matrix algorithms:

- If \( p(y) \) and \( q(y) \) are polynomials of degrees \( m \) and \( n \) respectively, then

\[
S(f + p, g + q) = S(f, g) + S(p, q)
\]

Theorem 2.1  Let the degree of the GCD of \( f(y) \) and \( g(y) \) be \( d \), then

1. The rank of \( S'(f, g) \) is equal to \( m + n - d \).

2. The coefficients of their GCD are given in the last non-zero row of \( S'(f, g)^T \) after it has been reduced to upper triangular form by an LU or QR decomposition.
Example 2.1 Consider the polynomials $\hat{f}(y)$ and $\hat{g}(y)$

$$\hat{f}(y) = -3y^3 + \frac{25}{2}y^2 - \frac{23}{2}y + 3 \quad \text{and} \quad \hat{g}(y) = 6y^2 - 7y + 2$$

whose GCD is $\hat{g}(y)$

$$\hat{f}(y) = -\frac{1}{2}(y - 3)\hat{g}(y)$$

The transpose $S(\hat{f}, \hat{g})^T$ of $S(\hat{f}, \hat{g})$ is

$$S(\hat{f}, \hat{g})^T = \begin{bmatrix}
-3 & \frac{25}{2} & -\frac{23}{2} & 3 & 0 \\
0 & -3 & \frac{25}{2} & -\frac{23}{2} & 3 \\
6 & -7 & 2 & 0 & 0 \\
0 & 6 & -7 & 2 & 0 \\
0 & 0 & 6 & -7 & 2
\end{bmatrix}$$
Its reduction to row echelon (upper triangular) form yields the matrix

\[
\begin{bmatrix}
-3 & \frac{25}{2} & -\frac{23}{2} & 3 & 0 \\
0 & -3 & \frac{25}{2} & -\frac{23}{2} & 3 \\
0 & 0 & 6 & -7 & 2 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

- The rank loss of this matrix is two, which is equal to the degree of the GCD of \( \hat{f}(y) \) and \( \hat{g}(y) \).
- The non-zero coefficients in its last non-zero row are 6, \(-7\), 2, and thus the GCD is \( 6y^2 - 7y + 2 \), which is equal to \( \hat{g}(y) \), as required.
2.1 Properties of the Sylvester resultant matrix

The Sylvester matrix satisfies

\[ S(f, \alpha g) \neq \alpha S(f, g), \quad \alpha \neq 1 \]

and the singular values \( \sigma_i \left( S(f, \alpha g) \right) \) of \( S(f, \alpha g) \) satisfy

\[ \sigma_i (S(f, \alpha g)) \neq \alpha \sigma_i (S(f, g)), \quad i = 1, \ldots, m + n, \quad \alpha \neq 1 \]

Compare

\[ \text{rank } S(f, \alpha g) \neq \text{rank } S(f, g), \quad \alpha \neq 1 \]

with

\[ \text{GCD}(f, \alpha g) \sim \text{GCD}(f, g), \quad \alpha \neq 0 \]
Compare with the Bézout resultant matrix $B(f, g) \in \mathbb{R}^{p \times p}, p = \max(m, n)$

$$B(\alpha f, \beta g) = \alpha \beta B(f, g), \quad \alpha, \beta \in \mathbb{R} \setminus 0$$

Every element $b_{i,j}$ of $B(f, g)$ is a bilinear function of the coefficients $a_i$ of $f(y)$ and $b_i$ of $g(y)$:

$$b_{i,j} = \sum_{k=0}^{\min(i-1,j-1)} (a_{p-i-j+1+k} b_{p-k} - a_{p-k} b_{p-i-j+1+k}), \quad i, j = 1, \ldots, p$$

Summary:

- The Bézout resultant matrix is smaller than the Sylvester resultant matrix
- The Bézout resultant matrix is scale invariant
Example 2.2 Consider the exact polynomials

$$\hat{f}(y) = (y - 1)^2(y - 2)^3(y - 4) \quad \text{and} \quad \hat{g}(y) = (y - 1)(y - 2)^2(y - 6)^9$$

Figure 1: The normalised singular values $\sigma_i/\sigma_1$ of (a) the Sylvester matrix $S(\hat{f}, \alpha\hat{g})$ for $\alpha = \alpha_1 = 1$, $\alpha = \alpha_2 = 10$ and $\alpha = \alpha_3 = 50$, and (b) the Bézout matrix $B(\hat{f}, \hat{g})$, in the absence of noise.
2.2 Subresultant matrices of the Sylvester resultant matrix

The product of two polynomials is equal to the convolution of their coefficients:

\[
\begin{bmatrix}
   r_{m+n} \\
   r_{m+n-1} \\
   \vdots \\
   r_1 \\
   r_0
\end{bmatrix}
= \begin{bmatrix}
   p_m \\
   p_{m-1} & p_m \\
   \vdots & \vdots & \ddots \\
   p_0 & \vdots & \vdots & p_m \\
   \vdots & \vdots & \ddots & \vdots \\
   p_0 & & & & \ddots & \vdots \\
   & & & & \vdots & \ddots & \vdots \\
   & & & \vdots & \ddots & \ddots & \vdots \\
   & & \vdots & \ddots & \ddots & \ddots & \vdots \\
   & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
   & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
   & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
   & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\end{bmatrix}
\begin{bmatrix}
   q_n \\
   q_{n-1} \\
   \vdots \\
   q_1 \\
   q_0
\end{bmatrix}
\]

\[
\mathbf{r} = S_{n+1}(\mathbf{p})\mathbf{q} = \mathbf{p} \otimes \mathbf{q}
\]

\(\mathbf{r} \in \mathbb{R}^{m+n+1}, \mathbf{p} \in \mathbb{R}^{m+1}, \mathbf{q} \in \mathbb{R}^{n+1}\) and \(S_{n+1}(\mathbf{p}) \in \mathbb{R}^{(m+n+1) \times (n+1)}\)
Let:

- $d_k(y)$ be a common divisor of degree $k$ of the exact polynomials $\hat{f}(y)$ and $\hat{g}(y)$
- The degree of the GCD of $\hat{f}(y)$ and $\hat{g}(y)$ be $d$
- $u_k(y)$ and $v_k(y)$ be the quotient polynomials

\[
\hat{f}(y) = u_k(y)d_k(y) \quad \text{and} \quad \hat{g}(y) = v_k(y)d_k(y)
\]

Thus

\[
\hat{f}(y)v_k(y) - \hat{g}(y)u_k(y) = 0 \iff S_{n-k+1}(\hat{f})v_k - S_{m-k+1}(\hat{g})u_k = 0
\]

where

\[
S_{n-k+1}(\hat{f}) \in \mathbb{R}^{(m+n-k+1) \times (n-k+1)} \quad \text{and} \quad v_k \in \mathbb{R}^{n-k+1}
\]

\[
S_{m-k+1}(\hat{g}) \in \mathbb{R}^{(m+n-k+1) \times (m-k+1)} \quad \text{and} \quad u_k \in \mathbb{R}^{m-k+1}
\]
\[
\begin{bmatrix}
S_{n-k+1}(\hat{f}) & S_{m-k+1}(\hat{g})
\end{bmatrix}
\begin{bmatrix}
v_k \\
-u_k
\end{bmatrix} = 0 \quad \text{or} \quad
S_k(\hat{f}, \hat{g})
\begin{bmatrix}
v_k \\
-u_k
\end{bmatrix} = 0
\]

- \(S_k(\hat{f}, \hat{g}) \in \mathbb{R}^{(m+n-k+1) \times (m+n-2k+2)}\) and it is rank deficient
- The nullspace vectors yield the coefficients of the quotient polynomials
- Since the degree of the GCD of \(\hat{f}(y)\) and \(\hat{g}(y)\) is \(\hat{d}\), these polynomials possess common divisors of degrees 1, 2, \ldots, \(\hat{d}\), but not a divisor of degree \(\hat{d} + 1\):
  \[
  \text{rank } S_k(\hat{f}, \hat{g}) < m + n - 2k + 2, \quad k = 1, \ldots, \hat{d}
  \]
  \[
  \text{rank } S_k(\hat{f}, \hat{g}) = m + n - 2k + 2, \quad k = \hat{d} + 1, \ldots, \min(m, n)
  \]

Calculating the degree of the GCD reduces to estimating the rank of a matrix.
Example 2.3 Consider $S_k(\hat{f}, \hat{f}^{(1)})$, for $k = 1, 2, 3$, where

\[
\hat{f}(y) = (y - 1)^2(y - 2)(y - 3) = y^4 - 7y^3 + 17y^2 - 17y + 6
\]
\[
\hat{f}^{(1)}(y) = 4y^3 - 21y^2 + 34y - 17
\]

Hence $S_1(\hat{f}, \hat{f}^{(1)}) = S(\hat{f}, \hat{f}^{(1)})$ is equal to

\[
\begin{bmatrix}
1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
-7 & 1 & 0 & -21 & 4 & 0 & 0 & 0 \\
17 & -7 & 1 & 34 & -21 & 4 & 0 & 0 \\
-17 & 17 & -7 & -17 & 34 & -21 & 4 & 0 \\
6 & -17 & 17 & 0 & -17 & 34 & -21 & 0 \\
0 & 6 & -17 & 0 & 0 & -17 & 34 & 0 \\
0 & 0 & 6 & 0 & 0 & 0 & -17 & 0 \\
\end{bmatrix}
\]

and this matrix has unit loss of rank.
The subresultant matrix $S_2(\hat{f}, \hat{f}^{(1)})$ is

$$
S_2(\hat{f}, \hat{f}^{(1)}) = \begin{bmatrix}
1 & 0 & 4 & 0 & 0 \\
-7 & 1 & -21 & 4 & 0 \\
17 & -7 & 34 & -21 & 4 \\
-17 & 17 & -17 & 34 & -21 \\
6 & -17 & 0 & -17 & 34 \\
0 & 6 & 0 & 0 & -17
\end{bmatrix}
$$

and this matrix has full column rank.
The subresultant matrix $S_3(\hat{f}, \hat{f}^{(1)})$ is

$$S_3(\hat{f}, \hat{f}^{(1)}) = \begin{bmatrix} 4 & 0 & 1 \\ -21 & 4 & -7 \\ 34 & -21 & 17 \\ -17 & 34 & -17 \\ 0 & -17 & 6 \end{bmatrix}$$

and this matrix has full column rank.

It follows that the first rank deficient matrix in the sequence

$$S_3(\hat{f}, \hat{f}^{(1)}), S_2(\hat{f}, \hat{f}^{(1)}), S_1(\hat{f}, \hat{f}^{(1)})$$

is $S_1(\hat{f}, \hat{f}^{(1)})$, and thus the degree of the GCD of $\hat{f}(y)$ and $\hat{f}^{(1)}(y)$ is one. □
3. POLYNOMIAL SCALING

Computations on polynomials whose coefficients vary widely in magnitude are problematic.

- Perform two preprocessing operations on the polynomials before the AGCD computations. These operations minimise the ratio

\[
\frac{\max \{\max |\text{coeff. of } f(y)|, \max |\text{coeff. of } g(y)|\}}{\min \{\min |\text{coeff. of } f(y)|, \min |\text{coeff. of } g(y)|\}}
\]

The operations are:

- Scale the coefficients of each polynomial by the geometric mean of its coefficients
- Introduce a parameter substitution
3.1 Scaling by the geometric mean

- It is numerically advantageous for the coefficients of $S(f, g)$ to have the same magnitude.
- Many researchers scale the coefficients of $f(y)$ and $g(y)$ to have unit 2-norm because it yields a Sylvester matrix that is better conditioned.
- Normalisation by the geometric mean of the coefficients is better for polynomials whose coefficients vary greatly in magnitude because it yields a ‘better average’ than does normalisation by the 2-norm of the coefficients.
The polynomials are therefore redefined as

\[ f(y) = \sum_{i=0}^{\bar{m}} \bar{a}_i y^{m-i}, \quad \bar{a}_i = \frac{a_i}{\left(\prod_{j=0}^{m} |a_j|\right)^{1/(m+1)}} \neq 0 \]

and

\[ g(y) = \sum_{i=0}^{\bar{n}} \bar{b}_i y^{n-i}, \quad \bar{b}_i = \frac{b_i}{\left(\prod_{j=0}^{n} |b_j|\right)^{1/(n+1)}} \neq 0 \]

- Normalisation by the geometric mean is not required for the Bézout resultant matrix.
- The parameter \( \alpha \) in \( S(f, \alpha g) \) is interpreted as the weight of \( g(y) \) relative to the weight of \( f(y) \).
3.2 Relative scaling of the polynomials

Scale the independent variable $y$ by

$$y = \theta w$$

and thus the transformed polynomials $\bar{f}(w)$ and $\bar{g}(w)$ are

$$\bar{f}(w) = \sum_{i=0}^{m} (\bar{a}_i \theta^{m-i}) w^{m-i} \quad \text{and} \quad \bar{g}(w) = \sum_{i=0}^{n} (\bar{b}_i \theta^{n-i}) w^{n-i}$$

The ratio of the maximum coefficient of $\{\bar{f}(w), \alpha \bar{g}(w)\}$ to the minimum coefficient of $\{\bar{f}(w), \alpha \bar{g}(w)\}$, that is, the arguments of $S(\bar{f}, \alpha \bar{g})$, is

$$\frac{\max \left\{ \max_{i=0,\ldots,m} |\bar{a}_i \theta^{m-i}|, \max_{j=0,\ldots,n} |\alpha \bar{b}_j \theta^{n-j}| \right\}}{\min \left\{ \min_{i=0,\ldots,m} |\bar{a}_i \theta^{m-i}|, \min_{j=0,\ldots,n} |\alpha \bar{b}_j \theta^{n-j}| \right\}}$$
The values $\alpha_0$ and $\theta_0$ that minimise this ratio are the optimal values of $\alpha$ and $\theta$, respectively.

Pose this minimisation as a linear programming (LP) problem:

Minimise $\frac{t}{s}$

Subject to

\[
\begin{align*}
  t &\geq \frac{\bar{a}_i}{\theta}^{m-i}, & i = 0, \ldots, m \\
  t &\geq \alpha \frac{\bar{b}_j}{\theta}^{n-j}, & j = 0, \ldots, n \\
  s &\leq \frac{\bar{a}_i}{\theta}^{m-i}, & i = 0, \ldots, m \\
  s &\leq \alpha \frac{\bar{b}_j}{\theta}^{n-j}, & j = 0, \ldots, n \\
  s &> 0 \\
  \theta &> 0 \\
  \alpha &> 0
\end{align*}
\]
The transformations

\[ T = \log t \quad S = \log s \quad \phi = \log \theta \]
\[ \mu = \log \alpha \quad \bar{\alpha}_i = \log |\bar{a}_i| \quad \bar{\beta}_j = \log |\bar{b}_j| \]

transform this constrained minimisation problem to

Minimise \( T - S \)

Subject to

\[ T - (m - i)\phi \geq \bar{\alpha}_i, \quad i = 0, \ldots, m \]
\[ T - (n - j)\phi - \mu \geq \bar{\beta}_j, \quad j = 0, \ldots, n \]
\[ -S + (m - i)\phi \geq -\bar{\alpha}_i, \quad i = 0, \ldots, m \]
\[ -S + (n - j)\phi + \mu \geq -\bar{\beta}_j, \quad j = 0, \ldots, n \]
which is a standard linear programming (LP) problem, whose objective function is

\[
T - S = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} T \\ S \\ \phi \\ \mu \end{bmatrix}
\]

The polynomials on which computations are performed are

\[
\tilde{f}(w) = \sum_{i=0}^{m} \tilde{a}_i w^{m-i} \quad \text{and} \quad \tilde{g}(w) = \sum_{i=0}^{n} \tilde{b}_i w^{n-i}
\]

where

\[
\tilde{a}_i = \bar{a}_i \theta_0^{m-i} \quad \text{and} \quad \tilde{b}_i = \alpha_0 \bar{b}_i \theta_0^{n-i}
\]
4. THE CALCULATION OF THE DEGREE OF AN APPROXIMATE GREATEST COMMON DIVISOR

Three methods are used to calculate the degree of an AGCD:

- The principle of maximum likelihood
- The angle between subspaces
- The polynomial form of an AGCD
Let the singular values of $B(\hat{f}, \hat{g})$, where $\hat{f}(w)$ and $\hat{g}(w)$ are the theoretically exact forms of $\tilde{f}(w)$ and $\tilde{g}(w)$ respectively, be

$$\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \cdots \geq \hat{\sigma}_r > 0, \hat{\sigma}_j = 0, \quad j = r + 1, \ldots, p$$

where $p = \max(m, n)$ and the degree of the GCD of $\hat{f}(w)$ and $\hat{g}(w)$ is $p - r$. In practical examples, only estimates $\sigma_i$ of the exact singular values $\hat{\sigma}_i$ are available, and thus

$$\sigma_i = \begin{cases} 
\hat{\sigma}_i + e_i & i = 1, \ldots, r \\
 e_i & i = r + 1, \ldots, p 
\end{cases}$$
Assumptions:

- The errors $e_i$ are statistically independent
- The errors associated with the non-zero and zero singular values have Gaussian and exponential probability distributions respectively,

\[
p(e_i) = \begin{cases} 
\frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{e_i^2}{2s^2}\right) & i = 1, \ldots, r \\
\alpha \exp(-\alpha e_i) & i = r + 1, \ldots, p
\end{cases}
\]

where $s, \alpha > 0$.

The value of $r$ that maximises $L(r)$

\[
L(r) = (p - r) \ln \left( \frac{p - r}{\sum_{j=r+1}^{p} \sigma_j} \right) - \frac{r}{2} \ln \left( \frac{2\pi \rho_r}{r} \right) - \left( p - \frac{r}{2} \right)
\]

is equal to the rank of $B(\tilde{f}, \tilde{g})$. 
4.2 The angle between subspaces

Consider the Sylvester matrix

\[
S(f, g) = \begin{bmatrix}
    a_m & & \\
    a_m-1 & a_m & \\
    & \ddots & \ddots \\
    a_0 & \ddots & a_m \\
    & \ddots & \ddots \\
    a_0 & & \\
    & \ddots & \\
    & & a_0 \\
\end{bmatrix}
\]
Let

- Let $\mathcal{F}_k$ be the space spanned by the first $n$ columns of $S_k(f, g)$
- Let $\mathcal{G}_k$ be the space spanned by the last $m$ columns of $S_k(f, g)$

where $k = 1, \ldots, \min(m, n)$.

The cosine of the angle $\phi_k$ between an arbitrary unit vector $u_k \in \mathcal{F}_k$ and an arbitrary unit vector $v_k \in \mathcal{G}_k$ is

$$\cos \phi_k = u_k^T v_k, \quad \|u_k\| = \|v_k\| = 1$$

where $\|\cdot\| = \|\cdot\|_2$. 
Definition: First principal angle

The first principal angle $\phi_{1,k}$ between $F_k$ and $G_k$ is the smallest angle that can be formed between $u_k$ and $v_k$.

Theorem 4.1 The first principal angle $\phi_{1,k}$ between $F_k$ and $G_k$ is zero if and only if the exact polynomials $\hat{f}(y)$ and $\hat{g}(y)$ have a common divisor of degree $k \geq 1$.

- If the exact polynomials $\hat{f}(y)$ and $\hat{g}(y)$ are coprime, then $\phi_{1,k} > 0$.
- If inexact polynomials $\tilde{f}(w)$ and $\tilde{g}(w)$ are specified, the value of $d$ is given by the index $k$ for which the change in the first principal angle between two successive values of $k$ is a maximum

$$d = \{k : |\phi_{1,k+1} - \phi_{1,k}| \rightarrow \text{max}\}, \quad k = 1, \ldots, \min(m, n) - 1$$
4.3 The polynomial form of a common divisor

Recall:

- $d_k(y)$ is a common divisor of degree $k$ of the exact polynomials $\hat{f}(y)$ and $\hat{g}(y)$
- The degree of the GCD of $\hat{f}(y)$ and $\hat{g}(y)$ is $\hat{d}$
- $u_k(y)$ and $v_k(y)$ are the quotient polynomials

\[
\hat{f}(y) = u_k(y)d_k(y) \quad \text{and} \quad \hat{g}(y) = v_k(y)d_k(y)
\]

In exact arithmetic

\[
\frac{\hat{f}(y)}{u_k(y)} = \frac{\hat{g}(y)}{v_k(y)}, \quad k = 1, \ldots, \hat{d}
\]
In finite precision arithmetic with inexact data

\[
\frac{\tilde{f}(w)}{u_k(w)} \approx \frac{\tilde{g}(w)}{v_k(w)}, \quad k = 1, \ldots, \min(m, n)
\]

and each division yields a rational function

A common divisor is a polynomial, not a rational function

Solution:

- Obtain initial estimates of \( u_k(w) \) and \( v_k(w) \) for \( k = 1, \ldots, \min(m, n) \)

- Express the equations

\[
\tilde{f}(w) \approx u_k(w)d_k^{(1)}(w) \quad \text{and} \quad \tilde{g}(w) \approx v_k(w)d_k^{(2)}(w)
\]

where \( d_k^{(1)}(w) \neq d_k^{(2)}(w) \), in matrix form

\[
F_k(u_k)d_k^{(1)} \approx \tilde{f} \quad \text{and} \quad G_k(v_k)d_k^{(2)} \approx \tilde{g}
\]
Solve each of these approximate equations in the least squares sense

\[ d_k^{(1)} = F_k (u_k)^\dagger \tilde{f} \quad \text{and} \quad d_k^{(2)} = G_k (v_k)^\dagger \tilde{g}, \quad k = 1, \ldots, \min(m, n) \]

and the degree \( d \) of an AGCD is equal to the index \( k = 1, \ldots, \min(m, n) \), for which the error measure

\[ e_k = \frac{\| d_k^{(1)} - d_k^{(2)} \|}{\| d_k^{(1)} \| + \| d_k^{(2)} \|} \]

is a minimum.
5. RESULTS

All polynomials in the examples were preprocessed by normalising by the geometric mean of their coefficients, scaling \( g(y) \) by \( \alpha_0 \) and making the parameter substitution \( y = \theta_0 w \), thereby yielding the polynomials \( \tilde{f}(w) \) and \( \tilde{g}(w) \).

Example 5.1 One thousand random pairs of the polynomials \( \{ \hat{f}_1(y), \hat{g}_1(y) \} \)

\[
\hat{f}_1(y) = \prod_{i=1}^{r_1} (y - \alpha_i)^{m_{1,i}} \quad \text{and} \quad \hat{g}_1(y) = \prod_{i=1}^{s_1} (y - \beta_i)^{n_{1,i}}
\]

were generated, where \( r_1, s_1 = 2, 3, 4, \alpha_1 = \beta_1, \alpha_2 = \beta_2 \), the roots \( \alpha_3, \ldots, \alpha_{r_1} \) and \( \beta_1, \ldots, \beta_{s_1} \) are arbitrary, and

\[
1 \leq m_{1,i}, n_{1,i} \leq 6 \\
5 \leq \sum_{i=1}^{r_1} m_{1,i} \leq 20 \\
-10 \leq \alpha_i, \beta_i \leq 10 \\
5 \leq \sum_{i=1}^{s_1} n_{1,i} \leq 20
\]
One thousand random pairs of a similar set \( \{ \hat{f}_2(y), \hat{g}_2(y) \} \) of polynomials, but with roots of higher maximum multiplicities

\[
\hat{f}_2(y) = \prod_{i=1}^{r_2} (y - \lambda_i)^{m_{2,i}} \quad \text{and} \quad \hat{g}_2(y) = \prod_{i=1}^{s_2} (y - \mu_i)^{n_{2,i}},
\]

were generated, where \( r_2, s_2 = 2, 3, 4, \lambda_1 = \mu_1, \lambda_2 = \mu_2 \), the roots \( \lambda_3, \ldots, \lambda_{r_2} \) and \( \mu_1, \ldots, \mu_{s_2} \) are arbitrary, and

\[
1 \leq m_{2,i}, n_{2,i} \leq 11 \quad -10 \leq \lambda_i, \mu_i \leq 10
\]
\[
5 \leq \sum_{i=1}^{r_2} m_{2,i} \leq 35 \quad 5 \leq \sum_{i=1}^{s_2} n_{2,i} \leq 35
\]

Noise with a componentwise signal-to-noise ratio of \( 10^8 \) was added to each of the 4000 polynomials.
Figure 2: Histograms of the results for the polynomials (a) \( \{ \hat{f}_1(y), \hat{g}_1(y) \} \) and (b) \( \{ \hat{f}_2(y), \hat{g}_2(y) \} \).
Example 5.2  Consider the exact polynomials

\[ \hat{f}(y) = (y - 6.7974)(y - 0.5903)^4(y - 3.3634)^3(y + 1.1265)^6 \]

and

\[ \hat{g}(y) = (y - 6.7974)^8(y - 0.5903)^9(y + 4.8572)^5(y + 6.8740)^5 \]

whose GCD is of degree five. Noise with a componentwise signal-to-noise ratio of $10^8$ was added to each polynomial.
Figure 3: The variation of (a) the likelihood function $L(r)$ with the rank $r$, (b) the first principal angle $\log \phi_{1,k}$ and (c) the error measure $\log e_k$, with the degree $k$ of an approximate common divisor.
Example 5.3

One hundred random pairs of polynomials \( \{ \hat{f}(y), \hat{g}(y) \} \), where each polynomial is of degree 20, were chosen such that the degree \( \hat{d} \) of their GCD is equal to one. The roots of each polynomial were randomly distributed in the interval \([-10, \ldots, 10]\), and the number of distinct roots of each polynomial was a random integer in the interval \([2, \ldots, 6]\). The multiplicity of each distinct root was chosen randomly, such that the constraint \( \hat{d} = 1 \) was satisfied. Noise of componentwise signal-to-noise ratio \( 10^8 \) was added to each polynomial, and the degree \( d \) of an AGCD was computed.

The experiment was repeated for \( \hat{d} = 2, 3, \ldots, 19 \), and the results are shown in Figure 4. It is seen that the methods yield similar results, and that the probability that \( \hat{d} = d \) decreases initially and then increases as \( \hat{d} \) increases. This figure provides more detail than the histograms in Figure 2 because it shows that the success of the methods is dependent upon the value of \( \hat{d} \).
Figure 4: The number of successful computations of the calculation of $\hat{d}$, the degree of the theoretically exact GCD, against $\hat{d}$. 

$\varepsilon_c = 10^{-8}$
5. SUMMARY

• This paper has compared three methods for the estimation of the degree of an AGCD of two inexact polynomials.

• The results suggest that the principle of maximum likelihood yields unsatisfactory results. The methods based on the angle between two subspaces and the polynomial form of an AGCD yield the best results.

• A possible explanation for this difference in results is that the principle of maximum likelihood is a general method, that is, it does not exploit the properties of a resultant matrix. By contrast, the method based on the angle exploits the partitioned structure of $S(f, g)$, and the error measure $e_k$ forces the polynomial form of an AGCD on the computations.

• The methods based on the angle and the error measure $e_k$ must be developed so that good results are obtained for small and large values of the GCD.