Article:
Winkler, J.R. and Allan, J.D. (2008) Structured total least norm and approximate GCDs of inexact polynomials. Journal of Computational and Applied Mathematics, 215 (1). pp. 1-13. ISSN 0377-0427

https://doi.org/10.1016/j.cam.2007.03.018

Structured total least norm and approximate GCDs of inexact polynomials

Joab R. Winkler, [a] John D. Allan [a]

[a] *Department of Computer Science, The University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, United Kingdom*

j.winkler@dcs.shef.ac.uk, j.allan@dcs.shef.ac.uk

Abstract

The determination of an approximate greatest common divisor (GCD) of two inexact polynomials $f = f(y)$ and $g = g(y)$ arises in several applications, including signal processing and control. This approximate GCD can be obtained by computing a structured low rank approximation $S^*(f, g)$ of the Sylvester resultant matrix $S(f, g)$. In this paper, the method of structured total least norm (STLN) is used to compute a low rank approximation of $S(f, g)$, and it is shown that important issues that have a considerable effect on the approximate GCD have not been considered. For example, the established works only yield one matrix $S^*(f, g)$, and therefore one approximate GCD, but it is shown in this paper that a family of structured low rank approximations can be computed, each member of which yields a different approximate GCD. Examples that illustrate the importance of these and other issues are presented.

Key words: Sylvester matrix, structured total least norm, approximate greatest common divisor

1 Introduction

The determination of the greatest common divisor (GCD) of two polynomials arises in several applications of signal processing and control. If the polynomials are known exactly and computations are performed symbolically, then Euclid's algorithm may be used, but this algorithm should not be executed in a floating point environment because it is numerically unstable. If the data is inexact, it is only possible to define an approximate GCD because the input data now consists of a (potentially infinite) family of polynomials that lie within the specified error bounds of their coefficients. In particular, if the inexact polynomials $f = f(y)$ and $g = g(y)$,

$$f(y) = \sum_{i=0}^{m} a_i y^{m-i} \qquad \text{and} \qquad g(y) = \sum_{j=0}^{n} b_j y^{n-j}, \tag{1}$$

are coprime, and the maximum normwise error in their coefficients is e, then a minor structured perturbation of the coefficients to $a_i + \delta a_i$ and $b_j + \delta b_j$, where $\delta a = \{\delta a_i\}_{i=0}^{m}$ and $\delta b = \{\delta b_j\}_{j=0}^{n}$, and

$$\|\delta a\| \leq e \ll \|a\|, \qquad \|\delta b\| \leq e \ll \|b\|, \qquad \|\cdot\| = \|\cdot\|_2, \tag{2}$$

may cause the perturbed inexact polynomials

$$\tilde{f}(y) = \sum_{i=0}^{m} (a_i + \delta a_i) y^{m-i} \qquad \text{and} \qquad \tilde{g}(y) = \sum_{j=0}^{n} (b_j + \delta b_j) y^{n-j}, \tag{3}$$

to have a non-constant GCD. Even if it is required to compute the smallest perturbations δa_i and δb_j such that the polynomials (3) have a non-constant GCD, different noisy realisations $f(y)$ and $g(y)$ of their theoretically exact forms yield different approximate GCDs, all of which are valid if (2) is satisfied.

The computation of an approximate GCD of the inexact polynomials (1) has been considered by several authors. For example, Corless *et. al.* [5], and Zarowski *et. al.* [13], use the QR decomposition of the Sylvester resultant matrix $S(f, g)$ [3], which will henceforth be called the Sylvester matrix. Similarly, the singular value decomposition of $S(f, g)$ is used in [4] in order to compute an approximate GCD, but both these decompositions do not preserve its structure. In particular, the smallest non-zero singular value of $S(f, g)$ is a measure of its distance to singularity, but this is the distance to an arbitrary rank deficient matrix, and not the distance to the nearest rank deficient Sylvester matrix. Karmarkar and Lakshman [8] use optimisation techniques in order to compute the smallest perturbations that must be applied to the coefficients of two polynomials such that they have a non-constant GCD, and Pan [11] uses Padé approximations to compute an approximate GCD. Zeng [14] uses partial singular value decompositions of Sylvester subresultant matrices, and an iterative algorithm is then used to calculate the factors of the GCD.

In this paper, the perturbations $\delta a_i, i = 0, \ldots, m$, and $\delta b_j, j = 0, \ldots, n$, in (3) are calculated by applying the method of structured total least norm (STLN) [12] to $S(f, g)$, thereby yielding the polynomials $\tilde{f} = \tilde{f}(y)$ and $\tilde{g} = \tilde{g}(y)$. The Sylvester matrix $S(f, g)$ is considered in Section 2, and it is shown that these perturbations allow the computation of an approximate GCD of the inexact polynomials $f(y)$ and $g(y)$. Although the computation of $S(\tilde{f}, \tilde{g})$ has been considered previously [7,9,15], there exist several important issues that must be considered when this matrix is used for the computation of an approximate GCD of $f(y)$ and $g(y)$. In particular:

- Since the GCD of $f(y)$ and $g(y)$ is equal to, up to an arbitrary scalar multiplier, the GCD of $f(y)$ and $\alpha g(y)$ where α is an arbitrary non-zero constant, it follows that the Sylvester resultant matrix $S(f, \alpha g)$ should be used when it is desired to compute an approximate GCD of $f(y)$ and $g(y)$. Since $S(f, \alpha g) \neq \alpha S(f, g)$, the inclusion of α permits a family of approximate GCDs, rather than only one approximate GCD, to be computed. In particular, it is shown in the examples in Section 4 that the restriction $\alpha = 1$ yields unsatisfactory solutions, but that the inclusion of α allows significantly improved solutions to be obtained.
- The method of STLN yields a non-linear least squares problem with an equality constraint, and the minimisation of the residual leads to a non-linear algebraic equation that is solved iteratively. It is shown that a stopping criterion that is based on a small normalised residual may lead to a poor or incorrect solution, and that an additional stopping criterion that is based on the singular values of $S(\tilde{f}, \tilde{g})$ must be included.
- The perturbed inexact polynomials $\tilde{f}(y)$ and $\tilde{g}(y)$ are obtained by computing the perturbations $\delta a_i, i = 0, \ldots, m$, and $\delta b_j, j = 0, \ldots, n$, and a computed approximate GCD is valid if (2) is satisfied. This condition requires that the norm of these perturbations be less than the maximum permissible normwise error in the coefficients because it guarantees that the perturbed inexact polynomials $\tilde{f}(y)$ and $\tilde{g}(y)$ are legitimate realisations of the theoretically exact forms of $f(y)$ and $g(y)$, respectively. This criterion has not been considered in previous work as a condition for the acceptance or rejection of an approximate GCD.

It is important to note that if interest is restricted to the computation of a structured low rank approximation of $S(f, g)$ and an approximate GCD computation is not required, then the default value $\alpha = 1$ must be used. In this paper, however, it is desired to compute a family of approximate GCDs from structured low rank approximations of $S(f, \alpha g)$, and thus the introduction of α is required. Different values of α yield different structured low rank approximations, and each of these approximations is a candidate for the computation of an approximate GCD of $f(y)$ and $g(y)$. It will be shown, however, that it is not possible to construct an approximate GCD for all values of α because a necessary equality constraint is not satisfied exactly, and the numerical rank of $S(\tilde{f}, \tilde{g})$ is not defined, for all values α.

Section 3 considers the application of the method of STLN to the computation of a structured low rank approximation of $S(f, \alpha g)$, and methods for the solution of the resulting equation are discussed. Section 4 contains examples that illustrate the importance of the three points that are discuused above. All the examples are non-trivial because polynomials of high degree are considered, and the roots have high multiplicity. A summary of the paper is contained in Section 5.

2 The Sylvester matrix

The Sylvester matrix $S(f, \alpha g) \in \mathbb{R}^{(m+n)\times(m+n)}$ is equal to

$$S(f,\alpha g) = \left[\begin{array}{cccc|cccc} a_0 & & & & \alpha b_0 & & & \\ a_1 & a_0 & & & \alpha b_1 & \alpha b_0 & & \\ \vdots & a_1 & \ddots & & \vdots & \alpha b_1 & \ddots & \\ a_{m-1} & \vdots & \ddots & a_0 & \alpha b_{n-1} & \vdots & \ddots & \alpha b_0 \\ a_m & a_{m-1} & \ddots & a_1 & \alpha b_n & \alpha b_{n-1} & \ddots & \alpha b_1 \\ & a_m & \ddots & \vdots & & \alpha b_n & \ddots & \vdots \\ & & \ddots & a_{m-1} & & & \ddots & \alpha b_{n-1} \\ & & & a_m & & & & \alpha b_n \end{array}\right],$$

where the coefficients a_i of $f(y)$ occupy the first n columns, and the coefficients αb_i of $\alpha g(y)$ occupy the last m columns. It is shown in [3] that if the degree of the GCD of $f(y)$ and $\alpha g(y)$ is equal to d, then the rank of $S(f, \alpha g)$ is equal to $(m + n - d)$, that is, the rank loss of the Sylvester matrix is equal to the degree of the GCD of $f(y)$ and $\alpha g(y)$. The condition number of $S(f, \alpha g)$ is a function of α, and it therefore follows that the accuracy and stability of the numerical computation of the GCD of $f(y)$ and $\alpha g(y)$ is dependent on α. This will be confirmed in Section 4, where several examples are considered, and it will be shown that an incorrect value of α leads to poor results.

The k'th Sylvester matrix, or subresultant, $S_k \in \mathbb{R}^{(m+n-k+1)\times(m+n-2k+2)}$ is a submatrix of $S(f, \alpha g)$ that is formed by deleting the last $(k - 1)$ rows of $S(f, \alpha g)$, the last $(k-1)$ columns of the coefficients of $f(y)$, and the last $(k-1)$ columns of the coefficients of $\alpha g(y)$.

Example 2.1

If $m = 4$ and $n = 3$, then

$$S_1 = S(f, \alpha g) = \begin{bmatrix} a_0 & & \alpha b_0 & & & \\ a_1 & a_0 & \alpha b_1 & \alpha b_0 & & \\ a_2 & a_1 & a_0 & \alpha b_2 & \alpha b_1 & \alpha b_0 \\ a_3 & a_2 & a_1 & \alpha b_3 & \alpha b_2 & \alpha b_1 & \alpha b_0 \\ a_4 & a_3 & a_2 & & \alpha b_3 & \alpha b_2 & \alpha b_1 \\ & a_4 & a_3 & & & \alpha b_3 & \alpha b_2 \\ & & a_4 & & & & \alpha b_3 \end{bmatrix},$$

$$S_2 = \begin{bmatrix} a_0 & & \alpha b_0 & & \\ a_1 & a_0 & \alpha b_1 & \alpha b_0 & \\ a_2 & a_1 & \alpha b_2 & \alpha b_1 & \alpha b_0 \\ a_3 & a_2 & \alpha b_3 & \alpha b_2 & \alpha b_1 \\ a_4 & a_3 & & \alpha b_3 & \alpha b_2 \\ & a_4 & & & \alpha b_3 \end{bmatrix}, \qquad S_3 = \begin{bmatrix} a_0 & \alpha b_0 & \\ a_1 & \alpha b_1 & \alpha b_0 \\ a_2 & \alpha b_2 & \alpha b_1 \\ a_3 & \alpha b_3 & \alpha b_2 \\ a_4 & & \alpha b_3 \end{bmatrix}.$$

□

Each matrix S_k is partitioned into a vector $c_k \in \mathbb{R}^{m+n-k+1}$ and a matrix $A_k = A_k(\alpha) \in \mathbb{R}^{(m+n-k+1)\times(m+n-2k+1)}$, where c_k is the first column of S_k, and A_k is the matrix formed from the remaining columns of S_k,

$$S_k = \Big[c_k \Big| \ A_k \Big] = \Big[c_k \Big| \text{ coeffs. of } f(y) \Big| \text{ coeffs. of } \alpha g(y) \Big].$$

The application of the method of STLN to the computation of an approximate GCD of the inexact polynomials $f(y)$ and $\alpha g(y)$ requires that the equation

$$A_k x = c_k, \qquad x \in \mathbb{R}^{m+n-2k+1}, \tag{4}$$

be considered. The following theorem is established in [7,9,15].

Theorem 2.1 *Consider the polynomials $f(y)$ and $\alpha g(y)$, where $f(y)$ and $g(y)$ are defined in (1), and let k be a positive integer, where $1 \leq k \leq \min(m, n)$. Then*

(1) The dimension of the null space of S_k is greater than or equal to one if and only if (4) possesses a solution.

(2) A necessary and sufficient condition for the polynomials $f(y)$ and $\alpha g(y)$ to have a common divisor of degree greater than or equal to k is that the rank of S_k is less than or equal to $(m+n-2k+1)$.

It is recalled that $f(y)$ and $\alpha g(y)$ are inexact and coprime, and that their theoretically exact forms have a non-constant GCD. There therefore exist perturbations $\delta f(y)$ and $\alpha\delta g(y)$ such that $f(y)+\delta f(y)$ and $\alpha\left(g(y)+\delta g(y)\right)$ have a non-constant common divisor, that is, if $h_k \in \mathbb{R}^{m+n-k+1}$ and $E_k \in \mathbb{R}^{(m+n-k+1)\times(m+n-2k+1)}$ are structured perturbations of c_k and A_k respectively, it follows from Theorem 2.1 that the equation

$$\left(A_k + E_k\right) x = c_k + h_k, \tag{5}$$

which is the perturbed form of (4), has an exact solution.

It follows from Theorem 2.1 that (5) has a solution if and only if $f(y)+\delta f(y)$ and $g(y)+\delta g(y)$ have a common divisor of degree greater than or equal to k. The computation of a structured low rank approximation of $S(f,\alpha g)$ therefore requires the determination of a structured matrix E_k and a structured vector h_k such that (5) possesses a solution for which A_k and E_k have the same structure, and c_k and h_k have the same structure. This is an overdetermined equation, and k is initially set equal to its maximum value, $k=k_0=\min(m,n)$. If a solution exists, then the degree of the GCD of $f(y)+\delta f(y)$ and $g(y)+\delta g(y)$ is equal to k_0. If this equation does not possess a solution, then k is reduced to k_0-1, and if a solution exists for this value of k, then the degree of the GCD of $f(y)+\delta f(y)$ and $g(y)+\delta g(y)$ is equal to k_0-1. If a solution does not exist, then k is reduced to k_0-2, and this process is repeated until (5) possesses a solution. This result is used in the next section in order to compute a structured low rank approximation of $S(f,\alpha g)$.

3 The method of STLN for a Sylvester matrix

It is shown in this section that the method of structured total least norm (STLN) can be used to compute a solution of (5), subject to the constraints on E_k and h_k that are stated in the previous paragraph.

Let $z=\{z_i\}_{i=0}^{m+n+1}$ be the vector of perturbations of the coefficients of $f(y)$ and $\alpha g(y)$ such that their perturbed forms have a non-constant common divisor. In particular, let z_i be the perturbation of the coefficient $a_i, i=0,\ldots,m$, of $f(y)$, and let z_{m+1+j} be the perturbation of the coefficient $\alpha b_j, j=0,\ldots,n$, of $\alpha g(y)$. The perturbed form $B_k \in \mathbb{R}^{(m+n-k+1)\times(m+n-2k+2)}$ of S_k is, therefore,

$$B_k := \left[\begin{array}{c|c} h_k & E_k \end{array}\right]$$

$$= \begin{bmatrix} z_0 & & & & z_{m+1} & & \\ z_1 & z_0 & & & z_{m+2} & & \\ \vdots & z_1 & \ddots & & \vdots & \ddots & \\ z_{m-1} & \vdots & \ddots & z_0 & z_{m+n} & \ddots & z_{m+1} \\ z_m & z_{m-1} & \ddots & z_1 & z_{m+n+1} & \ddots & z_{m+2} \\ & z_m & \ddots & \vdots & & \ddots & \vdots \\ & & \ddots & z_{m-1} & & \ddots & z_{m+n} \\ & & & z_m & & & z_{m+n+1} \end{bmatrix},$$

where h_k is equal to the first column of B_k, and E_k is equal to the last $(m + n - 2k + 1)$ columns of B_k.

Equation (5), which is non-linear, is solved iteratively for the perturbations $z_i, i = 0, \ldots, m+n+1$, such that the structures of E_k and h_k are retained. In particular, the residual $r(z, x)$ that is associated with an approximate solution of (5) due to the approximate perturbations h_k and E_k is

$$r(z, x) = c_k + h_k - (A_k + E_k)x, \qquad h_k = P_k z, \qquad E_k = E_k(z), \tag{6}$$

and it is required to minimise $\|Dz\|$ subject to the constraint $r(z, x) = 0$. The matrix $D \in \mathbb{R}^{(m+n+2)\times(m+n+2)}$ is diagonal and accounts for the repetition of the elements of z in B_k. In particular, each of the perturbations $z_i, i = 0, \ldots, m$, occurs $(n - k + 1)$ times in B_k, and each of the perturbations $z_i, i = m + 1, \ldots, m + n + 1$, occurs $(m - k + 1)$ times in B_k, and thus

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} = \begin{bmatrix} (n-k+1)I_{m+1} & 0 \\ 0 & (m-k+1)I_{n+1} \end{bmatrix}.$$

It is shown in [7,9,15] if $r(z, x)$ is linearised and the 2-norm is used, this constrained minimisation leads to a least squares problem with an equality constraint (the LSE problem),

$$\min_{\delta z} \left\| \begin{bmatrix} D & 0 \end{bmatrix} \begin{bmatrix} \delta z \\ \delta x \end{bmatrix} - (-Dz) \right\| \qquad \text{subject to} \qquad C \begin{bmatrix} \delta z \\ \delta x \end{bmatrix} = q, \tag{7}$$

where $\|\cdot\| = \|\cdot\|_2$, $C \in \mathbb{R}^{(m+n-k+1)\times(2m+2n-2k+3)}$ is a function of A_k, E_k and x,

and $q \in \mathbb{R}^{m+n-k+1}$ is a function of the residual of (6) due to an approximate solution of this equation.

If $E \in \mathbb{R}^{(m+n+2)\times(2m+2n-2k+3)}$, $\omega \in \mathbb{R}^{2m+2n-2k+3}$ and $p \in \mathbb{R}^{m+n+2}$ are defined as

$$E = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \end{bmatrix}, \qquad \omega = \begin{bmatrix} \delta z \\ \delta x \end{bmatrix} \qquad \text{and} \qquad p = -Dz,$$

respectively, where $\delta z \in \mathbb{R}^{m+n+2}$ and $\delta x \in \mathbb{R}^{m+n-2k+1}$, then the LSE problem (7) can be written as

$$\min_{w} \|E\omega - p\| \qquad \text{subject to} \qquad C\omega = q.$$

Algorithm 3.1 implements the LSE problem using the QR decomposition [6], and the initial value of x in this iterative algorithm is given by setting $r(z, x) = h_k = E_k = 0$ in (6),

$$x = \arg\min_{t} \|A_k t - c_k\|. \tag{8}$$

3.1 Solution methods for the LSE problem

The LSE problem (7) is usually solved by the method of weights [1,2,10]. Although this method transforms the LSE problem into an unconstrained least squares problem that can be solved by standard methods, it is necessary to introduce a weight parameter τ whose value is specified by heuristic methods. Van Loan [10] recommends that $\tau = \mu^{-\frac{1}{2}}$, but Barlow [1], and Barlow and Vemulapati [2], recommend that $\tau = \mu^{-\frac{1}{3}}$, where μ is the machine precision. The heuristic nature of τ is a disadvantage of this method because the convergence of the algorithm for the method of weights is critically dependent on the value of τ. In particular, if τ is too large or too small, the algorithm may converge slowly, or it may converge to an inaccurate solution, or it may not converge at all [1]. Furthermore, it is noted in [1] that the algorithm for the method of weights converges quickly for all but ill-conditioned problems. The QR decomposition does not suffer from these disadvantages of the method of weights, and it was therefore used for the solution of the LSE problem.

Algorithm 3.1: STLN for a Sylvester matrix

Input The polynomials $f(y)$ and $g(y)$, the scalar α, a value for k, where $1 \leq k \leq \min(m, n)$, and the tolerances ϵ_x and ϵ_z.

Output Polynomials $\tilde{f}(y) = f(y) + \delta f(y)$ and $\tilde{g}(y) = g(y) + \delta g(y)$ such that the degree of the GCD of $\tilde{f}(y)$ and $\tilde{g}(y)$ is greater than or equal to k.

Begin

(1) Form the k'th Sylvester matrix S_k from $f(y), g(y)$ and α.

(2) Set $E_k = 0$ and $h_k = 0$, and compute the initial value of x from (8). Construct the residual $r(z, x) = c_k - A_k x$, the matrix Y_k from x, and the matrix P_k.

(3) **Repeat**

(a) Compute the QR decomposition of C^T,

$$C^T = QR = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix}.$$

(b) Set $w_1 = R_1^{-T} q$.

(c) Partition EQ as

$$EQ = \begin{bmatrix} E_1 & E_2 \end{bmatrix},$$

where $E_1 \in \mathbb{R}^{(m+n+2)\times(m+n-k+1)}$ and $E_2 \in \mathbb{R}^{(m+n+2)\times(m+n-k+2)}$.

(d) Compute

$$z_1 = E_2^{\dagger}(p - E_1 w_1).$$

(e) Compute the solution

$$y = Q \begin{bmatrix} w_1 \\ z_1 \end{bmatrix}.$$

(f) Set $x := x + \delta x$ and $z := z + \delta z$.

(g) Update E_k and h_k from z, and Y_k from x. Compute the residual $r(z, x) = (c_k + h_k) - (A_k + E_k)x$.

Until $\frac{\|\delta x\|}{\|x\|} \leq \epsilon_x$ AND $\frac{\|\delta z\|}{\|z\|} \leq \epsilon_z$.

End

4 Examples

This section contains several examples that illustrate the method of STLN for the computation of a structured low rank approximation of $S(f, \alpha g)$ when it is required to compute an approximate GCD of $f(y)$ and $g(y)$ from this low rank approximation. It is shown that Algorithm 3.1 does not always yield a valid solution, and refinements to it are therefore required. Polynomials of high degree with roots of high multiplicity are considered in the examples in order to establish the robustness of the proposed algorithm.

The given inexact polynomials $f(y)$ and $g(y)$ are constructed by perturbing their theoretically exact forms $\hat{f} = \hat{f}(y)$ and $\hat{g} = \hat{g}(y)$. In particular, let $\mu = 1/\varepsilon$ be the signal-to-noise ratio,

$$\left\|\delta\hat{f}\right\| = \varepsilon \left\|\hat{f}\right\| \qquad \text{and} \qquad \|\delta\hat{g}\| = \varepsilon \|\hat{g}\|,$$

where the norm of a polynomial is equal to the 2-norm of its coefficients. If $c_f \in \mathbb{R}^{m+1}$ and $c_g \in \mathbb{R}^{n+1}$ are vectors of random variables, all of which are uniformly distributed in the interval $[-1, \ldots, +1]$, then the perturbations $\delta\hat{f}$ and $\delta\hat{g}$ are given by

$$\delta\hat{f} = \varepsilon \frac{\left\|\hat{f}\right\| c_f}{\|c_f\|} \qquad \text{and} \qquad \delta\hat{g} = \varepsilon \frac{\|\hat{g}\| c_g}{\|c_g\|},$$

respectively, and thus the inexact polynomials $f(y)$ and $g(y)$ are

$$f = \hat{f} + \varepsilon \frac{\left\|\hat{f}\right\| c_f}{\|c_f\|} \qquad \text{and} \qquad g = \hat{g} + \varepsilon \frac{\|\hat{g}\| c_g}{\|c_g\|}, \tag{9}$$

respectively. If the polynomials $f(y)$ and $g(y)$ are normalised, then α can be interpreted as the relative magnitude of $g(y)$ with respect to $f(y)$. Common normalisations include the 1, 2, and infinity norms of the coefficients, but normalisation by the geometric mean of the coefficients is used in this work because it is more suitable if the coefficients vary by several orders of magnitude. It therefore follows from (1) that the polynomials $f(y)$ and $g(y)$ in (9) are redefined as

$$f(y) := \frac{1}{\left(\prod_{i=0}^{m} |a_i|\right)^{\frac{1}{m+1}}} \sum_{i=0}^{m} a_i y^{m-i}, \tag{10}$$

and

$$g(y) := \frac{1}{\left(\prod_{j=0}^{n} |b_j|\right)^{\frac{1}{n+1}}} \sum_{j=0}^{n} b_j y^{n-j}, \tag{11}$$

respectively, where $a_i, i = 0, \ldots, m$, and $b_j, j = 0, \ldots, n$, are the perturbed coefficients, and thus the Sylvester matrix $S(f, \alpha g)$ is constructed from these polynomials. It is clear that if one or more of the coefficients of a polynomial is equal to zero, then the normalisation by the geometric mean of its coefficients, as shown in (10) and (11), requires modification.

The method of STLN allows the best vector z, the vector of perturbations of the coefficients of $f(y)$ and $\alpha g(y)$, that satisfies the LSE problem to be calculated, but the maximum permissible value of $\|z\|$ is related to the signal-to-noise ratio μ. In particular, the smaller the value of μ, the larger the maximum permissible value of $\|z\|$. This consideration leads to the definition of the *legitimate solution space*.

Definition 4.1 (Legitimate solution space) *The legitimate solution space of $\hat{f}(y)$ is the region that contains all perturbations of its coefficients that are allowed by the signal-to-noise ratio μ. The maximum allowable magnitude of these perturbations is ρ, where*

$$\rho = \frac{\|\hat{f}\|}{\mu}, \tag{12}$$

and all perturbations that are smaller than ρ lie in the legitimate solution space.

Since the errors consist of the data errors $\left\|f - \hat{f}\right\|$ and the structured perturbations from the method of STLN, it follows that (12) yields

$$\left\|f - \hat{f}\right\| + \|z_f\| \leq \frac{\left\|\hat{f}\right\|}{\mu}, \tag{13}$$

where $z_f \in \mathbb{R}^{m+1}$ denotes the structured perturbations of $f(y)$. This equation requires modification because $\hat{f}(y)$ is not known, and thus if it is assumed that

$$\left\|f - \hat{f}\right\| \ll \|z_f\| \qquad \text{and} \qquad \left\|\hat{f}\right\| \approx \|f\|,$$

then (13) can be approximated by

$$\|z_f\| \leq \frac{\|f\|}{\mu}. \tag{14}$$

This definition of the legitimate solution space is expressed in terms of $f(y)$, and it is clear that (14) is also satisfied by $g(y)$, but with a slight modification.

Specifically, since $z_{m+i+1}, i = 0, \ldots, n$, are the perturbations of the coefficients αb_i, it follows that

$$\frac{\|z_g\|}{\alpha} \leq \frac{\|g\|}{\mu}, \tag{15}$$

where $z_g \in \mathbb{R}^{n+1}$ stores the structured perturbations of the polynomial $\alpha g(y)$. It is clear from the definitions of z_f and z_g that

$$z = \begin{bmatrix} z_f \\ z_g \end{bmatrix},$$

and that acceptable structured perturbations require that the conditions (14) and (15) be satisfied.

Algorithm 4.1 is an extension of Algorithm 3.1 that performs a sequence of tests in order to eliminate values of α, and therefore polynomials $\tilde{f} = \tilde{f}(y)$ and $\tilde{g} = \tilde{g}(y)$, from Algorithm 3.1 that do not satisfy error criteria with regard to the legitimate solution space, the magnitude of the normalised residual, and the rank of the structured low rank approximation. In particular, Algorithm 3.1 is executed for a range of values of α, and the results are stored. Each value of α yields a different pair of polynomials $\tilde{f}$ and $\tilde{g}$, and Step 2 of Algorithm 4.1 is used to eliminate the values of α for which the magnitude of the structured perturbations is greater than the error in the polynomials, that is, polynomials that lie outside the legitimate solution space are discarded. Values of α for which the normalised residual $\|r_{norm}\|$ is too large are eliminated in Step 3 of Algorithm 4.1, which is therefore performed on a reduced set of solutions. Step 4 of Algorithm 4.1 calculates, for each of the remaining values of α, the singular values of the Sylvester matrix $S(\tilde{f}, \tilde{g})$ in order to determine its numerical rank. The value of α for which this quantity is most clearly defined is the optimal value α_0 of α, and a low rank approximation of $S(f, \alpha g)$ is constructed from the polynomials $\tilde{f}_0(y)$ and $\tilde{g}_0(y)$, which are the polynomials that are associated with α_0. An approximate GCD of $f(y)$ and $g(y)$ can be calculated by performing an LU or QR decomposition on $S(\tilde{f}_0, \tilde{g}_0)$.

Algorithm 4.1: Extended STLN for a Sylvester matrix

Input The polynomials $f(y)$ and $g(y)$, the scalar α, a value for k where $1 \leq k \leq \min(m, n)$, the tolerances ϵ_x and ϵ_z, the signal-to-noise ratio μ, and a range of values of α, $\alpha_1 \leq \alpha \leq \alpha_2$.

Output Polynomials $\tilde{f}_0(y)$ and $\tilde{g}_0(y)$ such that the degree of the GCD of $\tilde{f}_0(y)$ and $\tilde{g}_0(y)$ is greater than or equal to k.

Begin

(1) Apply Algorithm 3.1 with the given values of ϵ_x, ϵ_z and all values of α in the specified range. For each value of α, store the values of $\|z_f\|, \|z_g\|$ and r_{norm},

$$r_{norm} = \frac{r}{\|c_k + h_k\|},$$

where $r = r(z, x)$ is calculated in Step 3g of Algorithm 3.1 and r_{norm} is the normalised form of r.

(2) Retain the values of α for the values of $\|z_f\|$ and $\|z_g\|$ that satisfy (14) and (15), respectively.

(3) Retain the values of α for which the normalised residual $\|r_{norm}\|$ satisfies the error criterion

$$\|r_{norm}\| \leq 10^{-13}. \tag{16}$$

(4) For each acceptable value of α, compute the singular values σ_i of $S(\tilde{f}, \tilde{g})$, where $\tilde{f}$ and $\tilde{g}$ are the polynomials that are computed by Algorithm 3.1 and are normalised by the geometric mean of their coefficients, as shown in (10) and (11) for f and g, respectively. Arrange the singular values in non-increasing order, and choose the value α_0 of α for which the numerical rank of $S(\tilde{f}, \tilde{g})$ is equal to $(m + n - k)$, that is, the ratio

$$\frac{\sigma_{m+n-k}}{\sigma_{m+n-(k-1)}}, \tag{17}$$

is a maximum. The polynomials that correspond to the value α_0 are $\tilde{f}_0(y)$ and $\tilde{g}_0(y)$.

End

Example 4.1 Consider the exact polynomials

$$\hat{f}_1(y) = (y - 0.25)^8 (y - 0.5)^9 (y - 0.75)^{10} (y - 1)^{11} (y - 1.25)^{12}, \tag{18}$$

and

$$\hat{g}_1(y) = (y + 0.25)^4 (y - 0.25)^5 (y - 0.5)^6, \tag{19}$$

which have 11 common roots, from which it follows that rank $S(\hat{f}_1, \hat{g}_1) = 54$. The termination constants ϵ_x and ϵ_z, which are defined in Algorithm 3.1, were set equal to 10^{-6} and 10^{-8}, respectively.

Case 1: Signal-to-noise ratio $\mu = 10^8$, 11'th subresultant $k = 11$. The computation of a family of approximate GCDs from a given structured low rank approximation of $S(f_1, \alpha g_1)$.

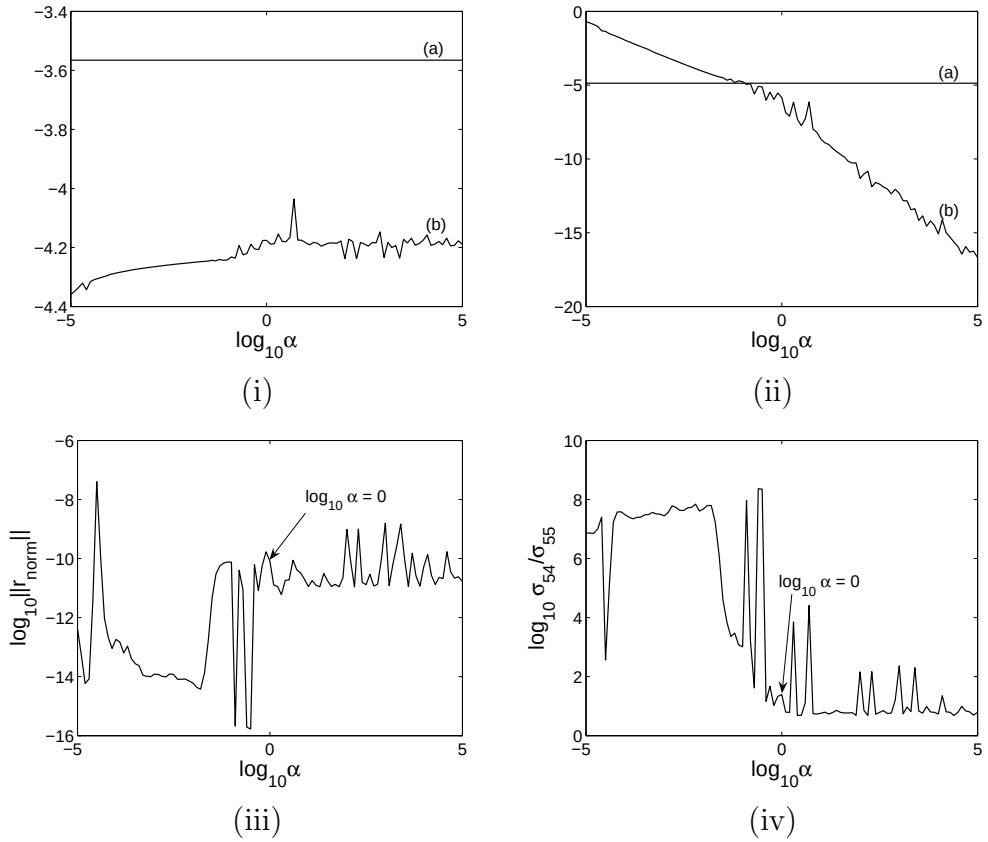

Fig. 1. (i)(a) The maximum allowable value of $\|z_{f_1}\|$, which is equal to $\|f_1\|/\mu$, (b) the computed value of $\|z_{f_1}\|$; (ii)(a) the maximum allowable value of $\|z_{g_1}\|/\alpha$, which is equal to $\|g_1\|/\mu$, (b) the computed value of $\|z_{g_1}\|/\alpha$; (iii) the normalised residual $\|r_{norm}\|$; (iv) the singular value ratio σ_{54}/σ_{55}.

The exact polynomials (18) and (19) were perturbed by noise such that $\mu = 10^8$ and then normalised by the geometric mean of their coefficients, yielding the polynomials $f_1(y)$ and $g_1(y)$. Figure 1 shows the results of applying the criteria in Steps 2, 3 and 4 in Algorithm 4.1. In particular, Figure 1(i) shows the ratio $\|f_1\|/\mu$, which is the maximum allowable perturbation of $f_1(y)$, and the variation with α of the computed value of $\|z_{f_1}\|$, which is calculated by the method of STLN. Figure 1(ii) is the same as Figure 1(i), but for the polynomial $g_1(y)$, and it is seen from (14) and (15) that valid solutions are obtained for $\log_{10}\alpha > -0.9$. Figure 1(iii) shows the variation of the normalised residual $\|r_{norm}\|$ with α, and it is seen that it ranges from $O(10^{-16})$ to $O(10^{-8})$ in the specified range of α.

Figure 1(iv) shows the variation with α of the ratio σ_{54}/σ_{55} that is defined

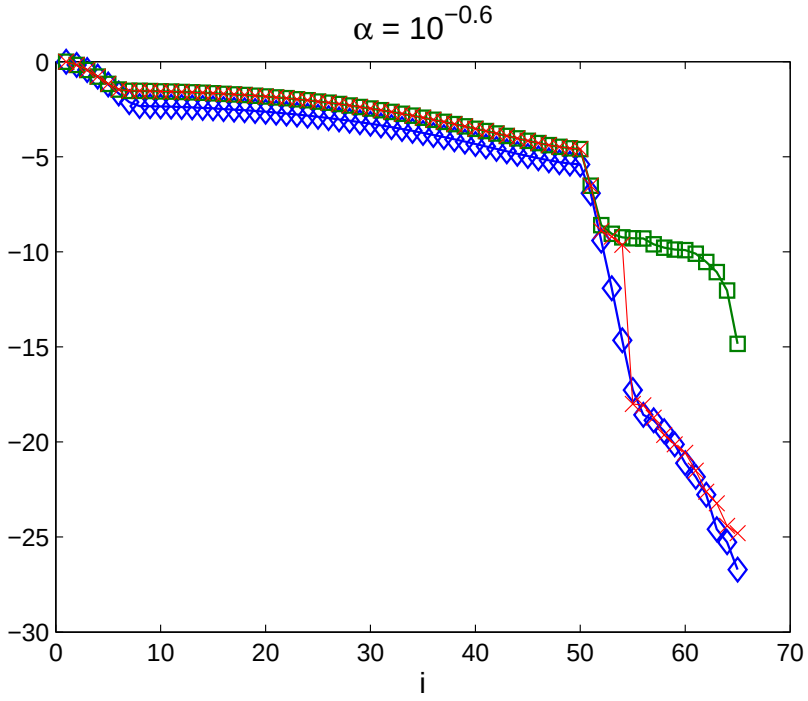

Fig. 2. The normalised singular values of the Sylvester matrix, on a logarithmic scale, for (i) the theoretically exact data $S(\hat{f}_1, \hat{g}_1)$, $\diamondsuit$; (ii) the given inexact data $S(f_1, g_1)$, $\square$; (iii) the computed data $S(\tilde{f}_{1,0}, \tilde{g}_{1,0})$, $\times$, for $\alpha = 10^{-0.6}$. All the polynomials are normalised by the geometric mean of their coefficients.

in (17), and it is seen that the profile of this curve could be produced (approximately) by calculating the reciprocal (to within a scale factor) of the normalised residual shown in Figure 1(iii). This result, which has been observed frequently, suggests that small values of the normalised residual are associated with large values of the ratio (17). It is noted, however, that a small value of the residual does not necessarily imply an accurate solution of a linear algebraic equation, and thus the use of the residual as the criterion for the acceptance or rejection of a solution is not recommended [6]. Rather, the results suggest that the residual should be used as one of several criteria for the acceptance or rejection of a solution.

This example clearly shows the importance of including α in the analysis because there exist, in general, many values of α for which the normalised residual is sufficiently small and the ratio σ_{54}/σ_{55} is sufficiently large. The small value of the normalised residual implies that the perturbed equation (6) is satisfied to high accuracy, and the large value of σ_{54}/σ_{55} implies that the numerical rank of the structured low rank approximation $S(\tilde{f}_1, \tilde{g}_1)$ is well defined. Each of these values of α yields a different structured low rank approximation of $S(f_1, \alpha g_1)$, and therefore a different approximate GCD of $f_1(y)$ and $g_1(y)$.

It is shown in Figure 1(iv) that in the absence of scaling, that is, $\log_{10} \alpha = 0$,

a poor solution is obtained because the ratio of the singular values (17) is approximately equal to $10^{1.5}$, which is about 7 orders of magnitude smaller than the ratio obtained for $\log_{10}\alpha = -0.6$, which is the optimal value of α. Figure 1(iii) shows that if $\log_{10}\alpha = 0$, the normalised residual is about 6 orders of magnitude larger than the value obtained for $\log_{10}\alpha = -0.6$. These observations show that an arbitrary choice of α can yield severely suboptimal results when it required to compute an approximate GCD of $f(y)$ and $g(y)$ from $S(f, \alpha g)$.

Figure 2 shows the normalised singular values of the Sylvester resultant matrices $S(\hat{f}_1, \hat{g}_1)$, $S(f_1, g_1)$, and $S(\tilde{f}_{1,0}, \tilde{g}_{1,0})$ for the optimal value of α, where all the polynomials are normalised by the geometric mean of their coefficients. The polynomials $\tilde{f}_{1,0}(y)$ and $\tilde{g}_{1,0}(y)$ are the polynomials from Algorithm 4.1 that form the structured low rank approximation of $S(f_1, \alpha g_1), \alpha = 10^{-0.6}$. It is seen that the computed singular values of $S(\hat{f}_1, \hat{g}_1)$ do not show a sharp cut off, which would suggest that the polynomials (18) and (19) are coprime. The profile of the singular values of $S(f_1, g_1)$ shows that the noise affects the small singular values severely, but significantly improved results are obtained when the Sylvester matrix $S(\tilde{f}_{1,0}, \tilde{g}_{1,0})$ is considered. In particular, it is clear that the numerical rank of this matrix is equal to 54 because σ_{54} is about 7 orders of magnitude larger than σ_{55}. Since the Sylvester matrix is of order 65×65 and $k = 11$, it is seen that the method of STLN has yielded an excellent result. Convergence of the algorithm was achieved in 45 iterations. It is clear that $S(\tilde{f}_{1,0}, \tilde{g}_{1,0})$ can be used to compute an approximate GCD of $f_1(y)$ and $g_1(y)$.

This example has considered the situation in which the correct subresultant has been selected because the degree of the GCD of $\hat{f}_1(y)$ and $\hat{g}_1(y)$ is 11, which is the chosen value of k, but this information is not, in general, known *a priori*. It is therefore necessary to consider how the solution changes as a function of k, and this is investigated in Case 2.

Case 2: Signal-to-noise ratio $\mu = 10^8$. The effects of different subresultants.

It follows from Theorem 2.1 that the lower bound on the degree of the GCD of $\hat{f}_1(y)$ and $\hat{g}_1(y)$ decreases as k decreases, and the next set of experiments investigates the performance of the method of STLN as k changes.

Computational experiments showed that the method of STLN is able to compute structured low rank approximations for $k = 10, \ldots, 1$. Figure 3 shows the results for $k = 8$, and it is seen that the numerical rank of $S(\hat{f}_1, \hat{g}_1)$ is not defined, but the numerical rank of its structured low rank approximation $S(\tilde{f}_{1,0}, \tilde{g}_{1,0})$ is equal to 57, corresponding to a loss in rank of 8. Convergence was achieved in 26 iterations.

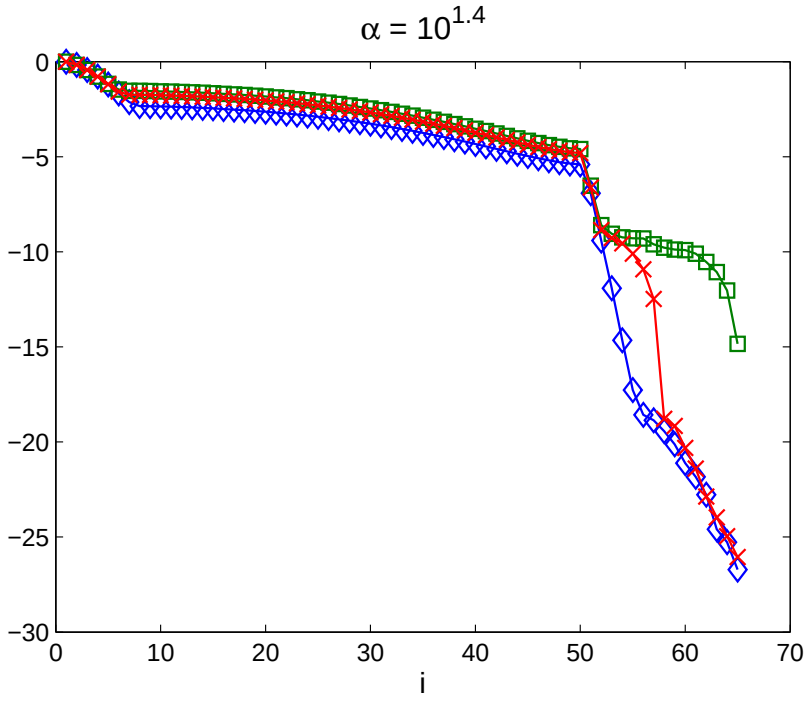

Fig. 3. The normalised singular values of the Sylvester matrix, on a logarithmic scale, for (i) the theoretically exact data $S(\hat{f}_1, \hat{g}_1)$, $\diamondsuit$; (ii) the given inexact data $S(f_1, g_1)$, $\square$; (iii) the computed data $S(\tilde{f}_{1,0}, \tilde{g}_{1,0})$, $\times$, for $\alpha = 10^{1.4}$. All the polynomials are normalised by the geometric mean of their coefficients.

Consider now the situation that occurs for $k = 12, 13$ and 14. In particular, successful results were obtained for $k = 12$ and $k = 13$, but the computed solution for $k \geq 14$ was not acceptable. This can be seen for $k = 14$ in Figures 4(i) and (ii), which show that although valid solutions exist for *either* $f_1(y)$ *or* $g_1(y)$, they do not exist for *both* $f_1(y)$ *and* $g_1(y)$. It is noted that if it is not required that the solution lie in the legitimate solution space, it is possible to construct structured low rank approximations matrices that can be used for the computation of approximate GCDs of $f_1(y)$ and $g_1(y)$, such that the ratio (17) is large and the normalised residual is small. $\square$

The next example is only considered briefly because the important points have been discussed in the previous examples.

Example 4.2 Consider the polynomials

$$\hat{f}_2(y) = (y-1)^8(y-2)^{16}(y-3)^{24},$$

and

$$\hat{g}_2(y) = (y-1)^{12}(y+2)^4(y-3)^8(y+4)^2,$$

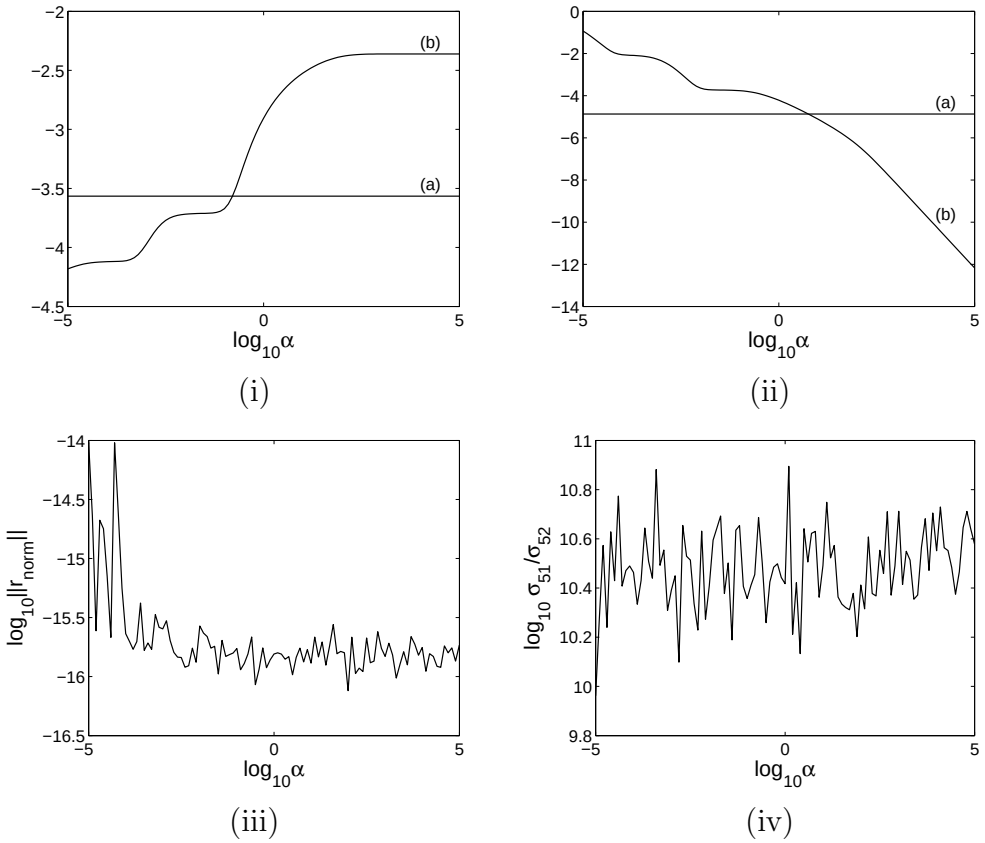

Fig. 4. (i)(a) The maximum allowable value of $\|z_{f_1}\|$, which is equal to $\|f_1\|/\mu$, (b) the computed value of $\|z_{f_1}\|$; (ii)(a) the maximum allowable value of $\|z_{g_1}\|/\alpha$, which is equal to $\|g_1\|/\mu$, (b) the computed value of $\|z_{g_1}\|/\alpha$; (iii) the normalised residual $\|r_{norm}\|$; (iv) the singular value ratio σ_{51}/σ_{52}.

which have 16 common roots, and thus the rank of $S(\hat{f}_2, \hat{g}_2)$ is 58. The polynomials were perturbed by noise such that $\mu = 10^8$, and the result for $k = 16$ is shown in Figure 5. It is seen that although the numerical rank of $S(\hat{f}_2, \hat{g}_2)$ is not well defined, the rank of the structured low rank approximation $S(\tilde{f}_{2,0}, \tilde{g}_{2,0})$ is 58, which is the correct value. Convergence was achieved in 22 iterations. □

5 Summary

This paper has considered the use of the method of STLN applied to the Sylvester resultant matrix for the computation of approximate GCDs of inexact polynomials. It has been shown that it is necessary to introduce a parameter α in order to obtain satisfactory solutions, and that there exist several values of α that satisfy tight tolerances on the normalised residual and the numerical rank of $S(\tilde{f}, \tilde{g})$. Each of these values of α yields a different structured low rank approximation of $S(f, \alpha g)$, and therefore a different approximate

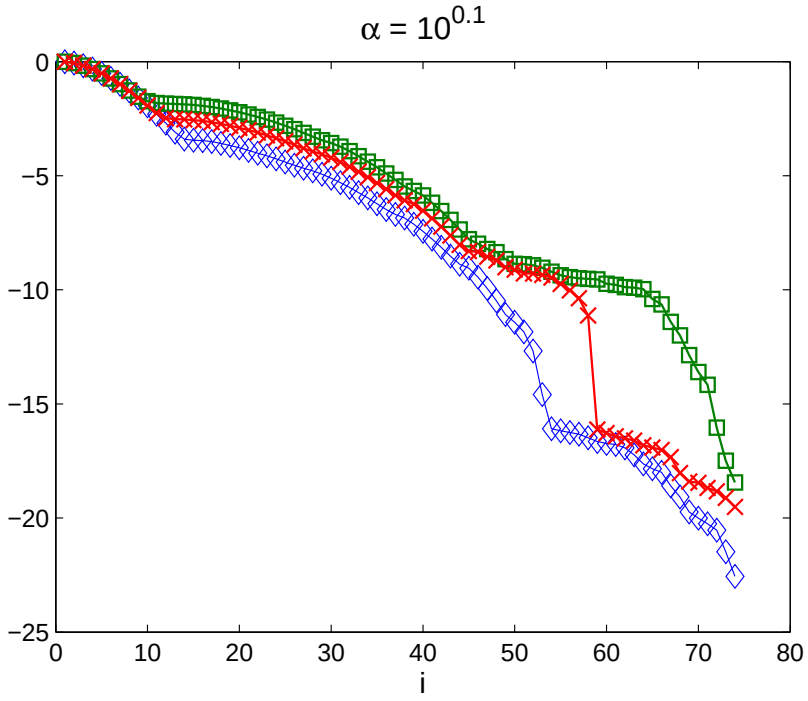

Fig. 5. The normalised singular values of the Sylvester matrix, on a logarithmic scale, for (i) the theoretically exact data $S(\hat{f}_2, \hat{g}_2)$, $\diamondsuit$; (ii) the given inexact data $S(f_2, g_2)$, $\square$; (iii) the computed data $S(\tilde{f}_{2,0}, \tilde{g}_{2,0})$, $\times$, for $\alpha = 10^{0.1}$. All the polynomials are normalised by the geometric mean of their coefficients.

GCD of $f(y)$ and $g(y)$. Additional constraints can be incorporated into the method in order to reduce further the range of acceptable values of α. Scaling the polynomials may affect the computed results, and it must therefore be chosen carefully. In this paper, scaling by the geometric mean of the coefficients of the polynomials was used, and very good results were obtained. It was shown that valid approximate GCDs of two inexact polynomials must satisfy a bound on the magnitude of the perturbations calculated by the method of STLN, and that this bound is related to the signal-to-noise ratio of the coefficients of the inexact polynomials.

References

[1] J. Barlow. Error analysis and implementation aspects of deferred correction for equality constrained least squares problems. *SIAM J. Numer. Anal.*, 25(6):1340–1358, 1988.

[2] J. Barlow and U. Vemulapati. A note on deferred correction for equality constrained least squares problems. *SIAM J. Numer. Anal.*, 29(1):249–256, 1992.

[3] S. Barnett. *Polynomials and Linear Control Systems*. Marcel Dekker, New York, USA, 1983.

[4] R. M. Corless, P. M. Gianni, B. M. Trager, and S. M. Watt. The singular value decomposition for polynomial systems. In *Proc. Int. Symp. Symbolic and Algebraic Computation*, pages 195–207. ACM Press, New York, 1995.

[5] R. M. Corless, S. M. Watt, and L. Zhi. QR factoring to compute the GCD of univariate inexact polynomials. *IEEE Trans. Signal Processing*, 52(12):3394–3402, 2004.

[6] G. H. Golub and C. F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, USA, 1996.

[7] E. Kaltofen, Z. Yang, and L. Zhi. Structured low rank approximation of a Sylvester matrix, 2005. Preprint.

[8] N. Karmarkar and Y. N. Lakshman. Approximate polynomial greatest common divsior and nearest singular polynomials. In *Proc. Int. Symp. Symbolic and Algebraic Computation*, pages 35–39. ACM Press, New York, 1996.

[9] B. Li, Z. Yang, and L. Zhi. Fast low rank approximation of a Sylvester matrix by structured total least norm. *J. Japan Soc. Symbolic and Algebraic Comp.*, 11:165–174, 2005.

[10] C. Van Loan. On the method of weighting for equality-constrained least squares problems. *SIAM J. Numer. Anal.*, 22(5):851–864, 1985.

[11] V. Y. Pan. Computation of approximate polynomial GCDs and an extension. *Information and Computation*, 167:71–85, 2001.

[12] J. Ben Rosen, H. Park, and J. Glick. Total least norm formulation and solution for structured problems. *SIAM J. Mat. Anal. Appl.*, 17(1):110–128, 1996.

[13] C. J. Zarowski, X. Ma, and F. W. Fairman. QR-factorization method for computing the greatest common divisor of polynomials with inexact coefficients. *IEEE Trans. Signal Processing*, 48(11):3042–3051, 2000.

[14] Z. Zeng. The approximate GCD of inexact polynomials. Part 1: A univariate algorithm, 2004. Preprint.

[15] L. Zhi and Z. Yang. Computing approximate GCD of univariate polynomials by structured total least norm. Technical report, Institute of Systems Science, AMSS, Academia Sinica, Beijing, China, 2004.