Electronic Transactions on Numerical Analysis.
Volume 31, pp. 141-155, 2008.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

STRUCTURED LOW RANK APPROXIMATIONS OF THE SYLVESTER RESULTANT MATRIX FOR APPROXIMATE GCDS OF BERNSTEIN BASIS POLYNOMIALS*

JOAB R. WINKLER† AND JOHN D. ALLAN†

Abstract. A structured low rank approximation of the Sylvester resultant matrix $S(f, g)$ of the Bernstein basis polynomials $f = f(y)$ and $g = g(y)$, for the determination of their approximate greatest common divisors (GCDs), is computed using the method of structured total least norm. Since the GCD of $f(y)$ and $g(y)$ is equal to the GCD of $f(y)$ and $\alpha g(y)$, where α is an arbitrary non-zero constant, it is more appropriate to consider a structured low rank approximation $S(\tilde{f}, \tilde{g})$ of $S(f, \alpha g)$, where the polynomials $\tilde{f} = \tilde{f}(y)$ and $\tilde{g} = \tilde{g}(y)$ approximate the polynomials $f(y)$ and $\alpha g(y)$, respectively. Different values of α yield different structured low rank approximations $S(\tilde{f}, \tilde{g})$, and therefore different approximate GCDs. It is shown that the inclusion of α allows to obtain considerably improved approximations, as measured by the decrease of the singular values σ_i of $S(\tilde{f}, \tilde{g})$, with respect to the approximation obtained when the default value $\alpha = 1$ is used. An example that illustrates the theory is presented and future work is discussed.

Key words. Bernstein polynomials, structured low rank approximation, Sylvester resultant matrix.

AMS subject classifications. 15A12, 65F35.

1. Introduction.

Resultant matrices are used in several disciplines, including robot motion planning [5], computer graphics [11], computer vision [17] and computer aided geometric design (CAGD), where they are used for the analysis and processing of curves and surfaces. For example, they allow the transformation of a curve between its parametric and implicit forms, and they can be used to calculate intersection points of curves and surfaces. The application of resultant matrices to CAGD requires that they be expressed in the Bernstein basis, rather than the monomial (power) basis, because the transformation of resultant matrices between these bases may be ill-conditioned [22, 23]. Although resultant matrices were originally developed for power basis polynomials, they have recently been extended to Bernstein basis polynomials [4, 21, 25], which has increased their applicability to CAGD.

The rank deficiency of the Sylvester resultant matrix $S(f, g)$ [3] is equal to the degree of the greatest common divisor (GCD) of the polynomials $f = f(y)$ and $g = g(y)$, and thus the computation of a structured low rank approximation of $S(f, g)$ is closely related to the computation of an approximate GCD of $f(y)$ and $g(y)$. This problem has been studied by Corless et al. [6], Emiris et al. [8] and Rupprecht [19], all of whom use the singular value decomposition of $S(f, g)$, and Corless et al. [7] and Zarowski [27], who use the QR decomposition of $S(f, g)$. These decompositions do not preserve the structure of the Sylvester matrix, which is a disadvantage. Padé approximations have also been used to compute an approximate GCD of two polynomials [16], and Karmarkar and Lakshman [13] use techniques from optimisation to compute the smallest perturbations that must be applied to two polynomials in order that they have a non-constant GCD.

The method of structured total least norm (STLN) [18] has been used to compute a structured low rank approximation of $S(f, g)$ [12, 14, 29], and we extend the work described in these three papers to consider the situation that occurs when low rank approximations of $S(f, g)$ are used to obtain approximate GCDs of $f(y)$ and $g(y)$. This computation is required

*Received November 28, 2007. Accepted August 14, 2008. Published online on February 12, 2009. Recommended by Reinhard Nabben.

†Department of Computer Science, The University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, United Kingdom ({j.winkler, j.allan}@dcs.shef.ac.uk). This work was supported by the Engineering and Physical Sciences Research Council, grant number GR/T24036/01.

ETNA
Kent State University
http://etna.math.kent.edu

in several problems, including the determination of multiple roots of a polynomial [20] and the cancellation of near pole-zero combinations in control systems.

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 a non-zero constant. Also, since the rank of $S(f,g)$ is equal to the rank of $S(f,\alpha g)$, and $S(f,\alpha g) \neq \alpha S(f,g)$, it is more appropriate to consider $S(f,\alpha g)$ when low rank approximations of the Sylvester resultant matrix are used for the computation of approximate GCDs. In particular, the inclusion of α allows significantly improved answers to be obtained, even after $f(y)$ and $g(y)$ have been normalised in the same manner [24]. Previous work on the computation of approximate GCDs has not included α, with the consequence that unsatisfactory results may have been computed.

The method of STLN has been used to compute a structured low rank approximation of the Sylvester resultant matrix for power basis polynomials [24], and it is extended in this paper to Bernstein basis polynomials. It is important to note that a low rank approximation of $S(f,g)$ cannot be computed by the deletion of its small singular values, because the low rank matrix is not a resultant matrix, that is, the structure of $S(f,g)$ is not preserved in this low rank approximation.

This paper contains two contributions that improve the quality of the computed results and have not been considered previously:

1. *The importance of* α. Previous work on the computation of approximate GCDs has not considered the scaling parameter α, that is, the value $\alpha = 1$ is used. The theoretical justification for including α is stated above, and the example in Section 5 clearly shows its computational importance. Furthermore, it is shown in the example, and confirmed by other examples that are not included in the paper, that although the optimal value α^* of α is associated with a small perturbation of the given inexact coprime polynomials in order that they have a non-constant approximate GCD, a small change in α^* yields an approximate GCD that is obtained by a significantly larger perturbation of the given inexact polynomials.
2. *The algorithm for the solution of the equation that results from the method of STLN.* The method of STLN yields a non-linear least squares problem with an equality constraint (the LSE problem), which is defined as

$$\min_{Bx=d} \|Ax-b\|_2 = \min_{x} \left\| \begin{bmatrix} A \\ \tau B \end{bmatrix} x - \begin{bmatrix} b \\ \tau d \end{bmatrix} \right\|_2, \qquad \tau \gg 1,$$

where $Ax = b$ is an over-determined set of linear algebraic equations and A has full column rank, and the constraint $Bx = d$ is consistent and therefore has at least one solution. The weight τ must be large in order that the constraint is satisfied exactly, and it is therefore required to normalise the constraint by a constant κ such that

$$\left\| \begin{bmatrix} A & b \end{bmatrix} \right\| = \kappa \left\| \begin{bmatrix} B & d \end{bmatrix} \right\|,$$

that is, the objective function and constraint are of the same magnitude. A large value of τ thus implies that the normalised constraint is enforced strongly, but this normalisation by κ is not considered in previous work that has used the method of weights [12, 14, 29].

The method of weights is problematic, because it requires that τ be specified heuristically. Van Loan [15] recommends that $\tau = \epsilon^{-\frac{1}{2}}$, but Barlow [1], and Barlow and Vermulapati [2], recommend that $\tau = \epsilon^{-\frac{1}{3}}$, where ϵ is the machine precision. The heuristic nature of τ is a disadvantage, because the convergence of the method of weights is critically dependent on the value of τ. Specifically, the algorithm may not converge, or it may converge to an inaccurate solution, or it may converge slowly,

ETNA
Kent State University
http://etna.math.kent.edu

if τ is too large or too small. The QR decomposition does not suffer from these disadvantages, and it is therefore used for the solution of the LSE problem in this paper. The method of weights is used, however, in [12, 14, 29] for the solution of the LSE problem, and thus the solution algorithm, based on the QR decomposition, that is used in this paper is superior.

The Sylvester resultant matrix $S(f, \alpha g)$ and its subresultant matrices are considered in Section 2, and the method of STLN is considered in Section 3. The standard form of the LSE problem does not impose constraints on the magnitudes of the perturbations that must be applied in order to induce a non-constant GCD in $f(y)$ and $g(y)$. It can be argued, however, that if the signal-to-noise ratio of the coefficients of $f(y)$ and $g(y)$ is μ, then the maximum allowable magnitudes of the perturbations of $f(y)$ and $g(y)$ are functions of μ, and, in particular, the smaller the value of μ, the larger the magnitude of the maximum allowable perturbations of $f(y)$ and $g(y)$. This topic is discussed in Section 4, and Section 5 contains an example of the method of STLN for the construction of a structured low rank approximation of $S(f, \alpha g)$. Sections 6 and 7 contain, respectively, a discussion of future work and a summary of the paper.

2. The Sylvester resultant matrix. This section considers the Sylvester resultant matrix and its subresultant matrices for Bernstein basis polynomials and scaled Bernstein basis polynomials. It is shown in [23, 25] that the Sylvester resultant matrix $S(f, g)$ of dimension $m + n$ of the polynomials

$$f(y) = \sum_{i=0}^{m} a_i \binom{m}{i} (1-y)^{m-i} y^i \qquad \text{and} \qquad g(y) = \sum_{i=0}^{n} b_i \binom{n}{i} (1-y)^{n-i} y^i$$

is

$$S(f, g) = D^{-1} T(f, g), \tag{2.1}$$

where $D, T(f, g) \in \mathbb{R}^{(m+n)\times(m+n)}$,

$$D^{-1} = \text{diag} \begin{bmatrix} \frac{1}{\binom{m+n-1}{0}} & \frac{1}{\binom{m+n-1}{1}} & \cdots & \frac{1}{\binom{m+n-1}{m+n-2}} & \frac{1}{\binom{m+n-1}{m+n-1}} \end{bmatrix},$$

and

$$T(f,g) = \begin{bmatrix} a_0\binom{m}{0} & & & b_0\binom{n}{0} & & \\ a_1\binom{m}{1} & \ddots & & b_1\binom{n}{1} & \ddots & \\ \vdots & \ddots & a_0\binom{m}{0} & \vdots & \ddots & b_0\binom{n}{0} \\ a_{m-1}\binom{m}{m-1} & \ddots & a_1\binom{m}{1} & b_{n-1}\binom{n}{n-1} & \ddots & b_1\binom{n}{1} \\ a_m\binom{m}{m} & \ddots & \vdots & b_n\binom{n}{n} & \ddots & \vdots \\ & \ddots & a_{m-1}\binom{m}{m-1} & & \ddots & b_{n-1}\binom{n}{n-1} \\ & & a_m\binom{m}{m} & & & b_n\binom{n}{n} \end{bmatrix}.$$

The matrix $T(f, g)$ is the Sylvester resultant matrix of $f(y)$ and $g(y)$ when these polynomials are expressed in the scaled Bernstein basis, whose basis functions for a polynomial of degree n are

$$\phi_i(y) = (1-y)^{n-i} y^i, \qquad i = 0, \ldots, n. \tag{2.2}$$

ETNA
Kent State University
http://etna.math.kent.edu

It is clear that $T(f,g)$ displays the strong diagonal pattern of its power basis equivalent, but the diagonal matrix D destroys this pattern. This makes the construction of a structured low rank approximation of the Bernstein basis resultant matrix $S(f,g)$ more involved than the construction of its power basis equivalent.

It was stated in Section 1 that it is necessary to consider $S(f,\alpha g)$, rather than $S(f,g)$, when it is desired to compute approximate GCDs of $f(y)$ and $g(y)$. This requirement follows from the structure of $S(f,g)$, because the coefficients of $f(y)$ and $g(y)$ are decoupled in this matrix, with the consequence that the numerical properties of $S(f,\alpha g)$ (for example, its singular values and distance to singularity) are strongly dependent upon α.

2.1. Subresultants of the Bernstein basis Sylvester resultant matrix. The subresultant matrices $S_k(f,\alpha g)$, $k = 1,\ldots,\min(m,n)$, are formed from the resultant matrix $S(f,\alpha g)$ by deletion of some of its rows and columns. In particular, the kth subresultant matrix $S_k(f,\alpha g)$ is formed by deleting the last $(k-1)$ columns of the coefficients of $f(y)$, the last $(k-1)$ columns of the coefficients of $\alpha g(y)$, and the last $(k-1)$ rows, from which it follows that $S_k(f,\alpha g)$ is of size $(m+n-k+1)\times(m+n-2k+2)$.

EXAMPLE 2.1. Consider the polynomials $f(y)$ and $g(y)$ for $m=4$ and $n=3$. Then, the first subresultant matrix $S_1(f,\alpha g) = S(f,\alpha g)$ is

$$\begin{bmatrix}
\frac{a_0\binom{4}{0}}{\binom{6}{0}} & & & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{0}} & & & \\
\frac{a_1\binom{4}{1}}{\binom{6}{1}} & \frac{a_0\binom{4}{0}}{\binom{6}{1}} & & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{1}} & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{1}} & & \\
\frac{a_2\binom{4}{2}}{\binom{6}{2}} & \frac{a_1\binom{4}{1}}{\binom{6}{2}} & \frac{a_0\binom{4}{0}}{\binom{6}{2}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{2}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{2}} & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{2}} & \\
\frac{a_3\binom{4}{3}}{\binom{6}{3}} & \frac{a_2\binom{4}{2}}{\binom{6}{3}} & \frac{a_1\binom{4}{1}}{\binom{6}{3}} & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{3}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{3}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{3}} & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{3}} \\
\frac{a_4\binom{4}{4}}{\binom{6}{4}} & \frac{a_3\binom{4}{3}}{\binom{6}{4}} & \frac{a_2\binom{4}{2}}{\binom{6}{4}} & & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{4}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{4}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{4}} \\
 & \frac{a_4\binom{4}{4}}{\binom{6}{5}} & \frac{a_3\binom{4}{3}}{\binom{6}{5}} & & & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{5}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{5}} \\
 & & \frac{a_4\binom{4}{4}}{\binom{6}{6}} & & & & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{6}}
\end{bmatrix},$$

while the second and third subresultant matrices, $S_2(f,\alpha g)$ and $S_3(f,\alpha g)$, are

$$\begin{bmatrix}
\frac{a_0\binom{4}{0}}{\binom{6}{0}} & & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{0}} & & \\
\frac{a_1\binom{4}{1}}{\binom{6}{1}} & \frac{a_0\binom{4}{0}}{\binom{6}{1}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{1}} & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{1}} & \\
\frac{a_2\binom{4}{2}}{\binom{6}{2}} & \frac{a_1\binom{4}{1}}{\binom{6}{2}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{2}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{2}} & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{2}} \\
\frac{a_3\binom{4}{3}}{\binom{6}{3}} & \frac{a_2\binom{4}{2}}{\binom{6}{3}} & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{3}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{3}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{3}} \\
\frac{a_4\binom{4}{4}}{\binom{6}{4}} & \frac{a_3\binom{4}{3}}{\binom{6}{4}} & & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{4}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{4}} \\
 & \frac{a_4\binom{4}{4}}{\binom{6}{5}} & & & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{5}}
\end{bmatrix}
\quad\text{and}\quad
\begin{bmatrix}
\frac{a_0\binom{4}{0}}{\binom{6}{0}} & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{0}} & \\
\frac{a_1\binom{4}{1}}{\binom{6}{1}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{1}} & \frac{\alpha b_0\binom{3}{0}}{\binom{6}{1}} \\
\frac{a_2\binom{4}{2}}{\binom{6}{2}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{2}} & \frac{\alpha b_1\binom{3}{1}}{\binom{6}{2}} \\
\frac{a_3\binom{4}{3}}{\binom{6}{3}} & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{3}} & \frac{\alpha b_2\binom{3}{2}}{\binom{6}{3}} \\
\frac{a_4\binom{4}{4}}{\binom{6}{4}} & & \frac{\alpha b_3\binom{3}{3}}{\binom{6}{4}}
\end{bmatrix}.$$

The following theorem is easily established.

THEOREM 2.2. *A necessary and sufficient condition for the polynomials $f(y)$ and $\alpha g(y)$ to have a common divisor of degree $k \geq 1$ is that the rank of $S_k(f,\alpha g)$ is less than $(m+n-2k+2)$.*

ETNA
Kent State University
http://etna.math.kent.edu

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

$$S_k(f, \alpha g) = \left[\begin{array}{c|c} c_k & A_k \end{array}\right] = \left[\begin{array}{c|c|c} c_k & \text{coeffs. of } f(y) & \text{coeffs. of } \alpha g(y) \end{array}\right]. \quad (2.3)$$

The coefficients of $f(y)$ occupy $(n-k)$ columns, while those of $\alpha g(y)$ occupy $(m-k+1)$ columns of $S_k(f, \alpha g)$. Theorems 2.2 and 2.3 show that the computation of the GCD of $f(y)$ and $\alpha g(y)$ requires that the equation

$$A_k x = c_k, \qquad x \in \mathbb{R}^{m+n-2k+1}, \quad (2.4)$$

be considered [12, 14, 29].

THEOREM 2.3. *If $k \leq \min(m, n)$ is a positive integer, then the dimension of the null space of $S_k(f, \alpha g)$ is greater than or equal to one if and only if (2.4) possesses a solution.*

The polynomials $f(y)$ and $g(y)$ are inexact and coprime in many examples, in which case (2.4) does not possess a solution for all $k = 1, \ldots, \min(m, n)$. The construction of a structured low rank approximation of $S(f, \alpha g)$ therefore requires the computation of perturbations $\delta f(y)$ and $\alpha \delta g(y)$, such that the polynomials $f(y) + \delta f(y)$ and $\alpha(g(y) + \delta g(y))$ have a non-constant GCD, which necessarily implies that a perturbed form of (2.4) must be considered. Furthermore, the coefficient matrix in (2.4) and its perturbed form must have the same structure, and the right hand side vector in (2.4) and its perturbed form must have the same structure, in order to guarantee that the perturbed equation is derived from the subresultant matrices of $S(f + \delta f, \alpha(g + \delta g))$.

It was noted above that the matrix D in (2.1) destroys the diagonal pattern of $S(f, \alpha g)$, but that $T(f, \alpha g)$, the Sylvester resultant matrix of $f(y)$ and $\alpha g(y)$ when they are expressed in the scaled Bernstein basis (2.2), preserves this pattern. It is therefore desirable to perform computations on the scaled Bernstein basis forms of $f(y)$ and $\alpha g(y)$, since this makes the imposition of the structure on the perturbed form of (2.4) easier. It may be argued that this choice leads to inferior numerical results because the Bernstein basis is numerically superior to the scaled Bernstein basis, but it is shown in Section 5 that excellent numerical results are obtained with the scaled Bernstein basis.

2.2. Subresultants of the scaled Bernstein basis Sylvester resultant matrix. The subresultant matrix $S_k(f, \alpha g)$ can be decomposed as

$$S_k(f, \alpha g) = D_k^{-1} T_k(f, \alpha g),$$

where $T_k(f, \alpha g) \in \mathbb{R}^{(m+n-k+1)\times(m+n-2k+2)}$ is the kth subresultant matrix of $T(f, \alpha g)$, that is, $T_k(f, \alpha g)$ is formed from $T(f, \alpha g)$ by deleting the last $(k-1)$ columns of $f(y)$, the last $(k-1)$ columns of $\alpha g(y)$, and the last $(k-1)$ rows. Similarly, the diagonal matrix D_k, of order $m+n-k+1$, is obtained by deleting the last $(k-1)$ rows and the last $(k-1)$ columns of D.

The matrix $T_k(f, \alpha g)$ is written in the form given in (2.3) for $S_k(f, \alpha g)$,

$$T_k(f, \alpha g) = \left[\begin{array}{c|c} d_k & F_k \end{array}\right] = \left[\begin{array}{c|c|c} d_k & \text{coeffs. of } f(y) & \text{coeffs. of } \alpha g(y) \end{array}\right],$$

where $d_k \in \mathbb{R}^{m+n-k+1}$ and $F_k \in \mathbb{R}^{(m+n-k+1)\times(m+n-2k+1)}$. It therefore follows that

$$S_k(f, \alpha g) = D_k^{-1} \left[\begin{array}{c|c} d_k & F_k \end{array}\right] = \left[\begin{array}{c|c} D_k^{-1} d_k & D_k^{-1} F_k \end{array}\right],$$

and thus (2.3) yields

$$c_k = D_k^{-1} d_k, \quad (2.5)$$

ETNA
Kent State University
http://etna.math.kent.edu

and

$$A_k = D_k^{-1} F_k. \tag{2.6}$$

It therefore follows from (2.5) and (2.6) that (2.4) is replaced by

$$F_k x = d_k, \tag{2.7}$$

which assumes that $f(y)$ and $g(y)$ are expressed in the scaled Bernstein basis.

It was noted above that $f(y)$ and $g(y)$ are inexact and coprime in many examples, and thus a perturbed form of (2.7) must be considered,

$$(F_k + E_k)\, x = d_k + h_k, \tag{2.8}$$

where $h_k \in \mathbb{R}^{m+n-k+1}$ and $E_k \in \mathbb{R}^{(m+n-k+1)\times(m+n-2k+1)}$, such that this equation has an exact solution. The matrices F_k and E_k have the same structure, as well as the vectors d_k and h_k, and it is therefore necessary to determine the smallest structured perturbations E_k and h_k such that (2.8), which is an over-determined linear system, possesses a solution. This constrained equation is solved by the method of structured total least norm [18], which is considered in the next section.

3. The method of structured total least norm. It is shown in this section that the method of structured total least norm (STLN) can be used to compute the smallest perturbations E_k and h_k such that (2.8) is satisfied, where F_k and E_k have the same structure, and d_k and h_k have the same structure.

The Sylvester resultant matrix $T_k(f, \alpha g)$ for the scaled Bernstein basis forms of $f(y)$ and $g(y)$ is given by

$$T_k(f,\alpha g) = \left[\begin{array}{c|c} d_k & F_k \end{array}\right]$$
$$= \begin{bmatrix}
a_0\binom{m}{0} & & & \alpha b_0\binom{n}{0} & & \\
a_1\binom{m}{1} & \ddots & & \alpha b_1\binom{n}{1} & \ddots & \\
\vdots & \ddots & a_0\binom{m}{0} & \vdots & \ddots & \alpha b_0\binom{n}{0} \\
a_{m-1}\binom{m}{m-1} & \ddots & a_1\binom{m}{1} & \alpha b_{n-1}\binom{n}{n-1} & \ddots & \alpha b_1\binom{n}{1} \\
a_m\binom{m}{m} & \ddots & \vdots & \alpha b_n\binom{n}{n} & \ddots & \vdots \\
 & \ddots & a_{m-1}\binom{m}{m-1} & & \ddots & \alpha b_{n-1}\binom{n}{n-1} \\
 & & a_m\binom{m}{m} & & & \alpha b_n\binom{n}{n}
\end{bmatrix},$$

and thus the structured error matrix $B_k(z)$ is given by

$$B_k(z) := \left[\begin{array}{c|c} h_k & E_k \end{array}\right]$$
$$= \begin{bmatrix}
z_0 & & & & z_{m+1} & & & \\
z_1 & z_0 & & & z_{m+2} & z_{m+1} & & \\
\vdots & z_1 & \ddots & & \vdots & z_{m+2} & \ddots & \\
z_{m-1} & \vdots & \ddots & z_0 & z_{m+n} & \vdots & \ddots & z_{m+1} \\
z_m & z_{m-1} & \ddots & z_1 & z_{m+n+1} & z_{m+n} & \ddots & z_{m+2} \\
 & z_m & \ddots & \vdots & & z_{m+n+1} & \ddots & \vdots \\
 & & \ddots & z_{m-1} & & & \ddots & z_{m+n} \\
 & & & z_m & & & & z_{m+n+1}
\end{bmatrix}, \tag{3.1}$$

ETNA
Kent State University
http://etna.math.kent.edu

where h_k is equal to the first column of $B_k(z)$, and E_k is equal to the last $(m+n-2k+1)$ columns of $B_k(z)$. It is immediate that $B_k(z)$ is a Sylvester resultant matrix, and thus also $T_k(f,\alpha g)+B_k(z)$ is a resultant matrix. The element z_i of $z\in\mathbb{R}^{m+n+2}$ is the perturbation of the coefficient a_i, $i=0,\ldots,m$, of $f(y)$, and the element z_{m+1+j} is the perturbation of the coefficient αb_j, $j=0,\ldots,n$, of $\alpha g(y)$.

It follows from the definitions of the vectors h_k and z that there exists a matrix P_k of size $(m+n-k+1)\times(m+n+2)$ such that

$$h_k = P_k z = \left[\begin{array}{cc} I_{m+1} & 0_{m+1,n+1} \\ 0_{n-k,m+1} & 0_{n-k,n+1}\end{array}\right] z,$$

where I_{m+1} is the identity matrix of order $m+1$ and the subscripts on the zero matrices indicate their order.

The residual $r(z,x)$ that is associated with an approximate solution of (2.8), due to the structured perturbations h_k and E_k, is

$$r(z,x) = d_k + h_k - (F_k+E_k)x, \qquad h_k = P_k z, \qquad E_k = E_k(z). \tag{3.2}$$

Since it is required to compute the smallest perturbations z_i such that (2.8) has a solution, it is necessary to minimise $\|Hz\|$ subject to the constraint $r(z,x)=0$. The matrix H, of order $m+n+2$, accounts for the repetition of the elements of z in $B_k(z)$. In particular, it follows from (3.1) that each of the perturbations $z_i, i=0,\ldots,m$, occurs $(n-k+1)$ times, and each of the perturbations $z_i, i=m+1,\ldots,m+n+1$, occurs $(m-k+1)$ times in $B_k(z)$, and thus

$$H = \left[\begin{array}{cc} H_1 & 0 \\ 0 & H_2\end{array}\right] = \left[\begin{array}{cc} (n-k+1)I_{m+1} & 0 \\ 0 & (m-k+1)I_{n+1}\end{array}\right].$$

The objective function in [12, 14, 29] is $\|z\|$, but this measure fails to consider the number of occurrences of $z_i, i=0,\ldots,m+n+1$, in $B_k(z)$. The 2-norm is used, $\|\cdot\|=\|\cdot\|_2$, and thus this constrained minimisation is an equality constrained least squares problem (the LSE problem).

It is necessary to replace the vector $E_k x$ by the vector $Y_k z$,

$$Y_k z = E_k x, \qquad Y_k = Y_k(x), \qquad E_k = E_k(z),$$

and thus

$$Y_k \delta z = \delta E_k x, \tag{3.3}$$

where $Y_k\in\mathbb{R}^{(m+n-k+1)\times(m+n+2)}$. Closed form expressions for the elements of Y_k are derived using the expressions for the elements of E_k, which are easily obtained from (3.1).

The function $r(z,x)$ is non-linear, and thus iterative algorithms based on a linear approximation are used for the solution of $r(z,x)=0$. In particular, if it is assumed that second order terms are sufficiently small such that they can be neglected, then, since $h_k=h_k(z)$ and

$E_k = E_k(z)$, using (3.3) we obtain

$$\begin{aligned}
r(z+\delta z, x+\delta x) &= d_k + h_k(z+\delta z) - (F_k + E_k(z+\delta z))\,(x+\delta x) \\
&\approx d_k + P_k z + P_k \delta z - F_k x - F_k \delta x \\
&\quad - E_k x - E_k \delta x - \left(\sum_{i=0}^{m+n+1} \frac{\partial E_k}{\partial z_i} \delta z_i \right) x \\
&= d_k + P_k z + P_k \delta z - F_k x - F_k \delta x \\
&\quad - E_k x - E_k \delta x - \delta E_k x \\
&= r(z,x) + P_k \delta z - F_k \delta x - E_k \delta x - Y_k \delta z \\
&= r(z,x) - (Y_k - P_k)\delta z - (F_k + E_k)\delta x.
\end{aligned} \tag{3.4}$$

The constrained minimisation can therefore be approximated by

$$\min_{\delta z} \|H(z+\delta z)\|$$

subject to

$$\tilde{r} = r(z,x) - (Y_k - P_k)\delta z - (F_k + E_k)\delta x = 0.$$

This problem is of the form

$$\min \|E\omega - p\| \quad \text{subject to} \quad C\omega = q, \tag{3.5}$$

where the function to be minimised represents the norm of the perturbations z_i,

$$\|E\omega - p\| = \left\| \begin{bmatrix} H & 0 \end{bmatrix} \begin{bmatrix} \delta z \\ \delta x \end{bmatrix} - (-Hz) \right\|,$$

and the constraint $\tilde{r} = 0$ follows from (3.4),

$$C\omega - q = \begin{bmatrix} (Y_k - P_k) & (F_k + E_k) \end{bmatrix} \begin{bmatrix} \delta z \\ \delta x \end{bmatrix} - r(z,x) = 0.$$

The constraint equation $C\omega = q$ is under-determined, which guarantees that the minimisation is performed over an infinite number of vectors ω. Methods for the solution of (3.5) are discussed in Section 3.1.

It is recalled that z is the vector of perturbations of $f(y)$ and $\alpha g(y)$ when they are expressed in the scaled Bernstein basis, and thus the corrected scaled Bernstein basis polynomials are

$$\sum_{i=0}^{m} \left(a_i \binom{m}{i} + z_i \right) (1-y)^{m-i} y^i$$

and

$$\sum_{i=0}^{n} \left(\alpha b_i \binom{n}{i} + z_{m+1+i} \right) (1-y)^{n-i} y^i,$$

respectively. The Bernstein basis form of these polynomials is

$$\tilde{f}(y) = \sum_{i=0}^{m} \left(a_i + \frac{z_i}{\binom{m}{i}} \right) \binom{m}{i} (1-y)^{m-i} y^i$$

ETNA
Kent State University
http://etna.math.kent.edu

and

$$\tilde{g}(y) = \sum_{i=0}^{n} \left(b_i + \frac{z_{m+1+i}}{\alpha \binom{n}{i}} \right) \binom{n}{i} (1-y)^{n-i} y^i,$$

respectively, where the multiplier α has been omitted from $\tilde{g}(y)$. It follows that the perturbations of the coefficients of the given inexact Bernstein basis polynomials $f(y)$ and $g(y)$ are

$$\frac{z_i}{\binom{m}{i}}, \quad i = 0, \ldots, m, \qquad \text{and} \qquad \frac{z_{m+1+i}}{\alpha \binom{n}{i}}, \quad i = 0, \ldots, n, \tag{3.6}$$

respectively.

3.1. Methods for solution of the LSE problem. There exist several methods for the solution of the LSE problem, including the method of weights (also called the penalty method) and the null space method [9, 10]. The method of weights requires a weight whose value is defined heuristically, which is a disadvantage because an incorrect value may lead to the algorithm converging very slowly, or not converging at all, or converging to an inaccurate solution [1]. The QR decomposition does not suffer from this disadvantage, and it was therefore used in the example in Section 5.

It is recalled that (3.4) is a linear approximation of a non-linear expression, and the extensions of the QR decomposition for the solution of the LSE problem that include this linearisation, and the calculation of the initial forms of x and z in the iterative scheme, are considered in Algorithms 3.1 and 4.1 in [24].

4. Bounds on the magnitude of the perturbations. The coefficients of $f(y)$ and $g(y)$ are known inexactly in many practical examples, and repetition of the experimental procedure or the use of another computational algorithm that generated them will yield slightly different values of these coefficients. This shows that the given inexact polynomials are one realisation of an infinite number of noisy realisations of the theoretically exact data, and thus the coefficients of the given inexact polynomials $f(y)$ and $g(y)$ can be perturbed by an amount that is governed by the signal-to-noise ratio μ of the data. In particular, the smaller the value of μ, the larger the allowable perturbations of the coefficients of the inexact polynomials.

It follows from (3.6) that if the vectors $z_f \in \mathbb{R}^{m+1}$ and $z_g \in \mathbb{R}^{n+1}$ are defined as

$$z_f = \left[\begin{array}{cccccc} \frac{z_0}{\binom{m}{0}} & \frac{z_1}{\binom{m}{1}} & \cdots & \frac{z_{m-1}}{\binom{m}{m-1}} & \frac{z_m}{\binom{m}{m}} \end{array} \right]$$

and

$$z_g = \left[\begin{array}{cccccc} \frac{z_{m+1}}{\binom{n}{0}} & \frac{z_{m+2}}{\binom{n}{1}} & \cdots & \frac{z_{m+n}}{\binom{n}{n-1}} & \frac{z_{m+n+1}}{\binom{n}{n}} \end{array} \right],$$

respectively, then z_f and z_g/α are the structured perturbations computed by the method of STLN that are applied to the coefficients of the Bernstein basis polynomials $f(y)$ and $g(y)$. Since $\|f\|/\mu$ and $\|g\|/\mu$ are the maximum allowable perturbations of $f(y)$ and $g(y)$, respectively, it follows that

$$\|z_f\| \leq \frac{\|f\|}{\mu} \qquad \text{and} \qquad \frac{\|z_g\|}{\alpha} \leq \frac{\|g\|}{\mu}, \tag{4.1}$$

and all vectors z_f and z_g that satisfy these inequalities and the LSE problem are legitimate perturbations of the given inexact data.

5. An example. This section contains an example that illustrates how the theory of the previous sections can be used to construct structured low rank approximations of $S(f, \alpha g)$ in order to compute approximate GCDs of $f(y)$ and $g(y)$. The noisy polynomials $f(y)$ and $g(y)$ are normalised by the geometric mean of their Bernstein basis coefficients because computational experiments showed that this form of scaling yields very good results for power basis polynomials, and, in particular, significantly better results than monic scaling.

Consider the Bernstein basis form of the exact polynomials

$$\hat{f}(y) = (y-0.6)^8 (y-0.8)^9 (y-0.9)^{10} (y-0.95)^5$$

and

$$\hat{g}(y) = (y-0.6)^{12} (y-0.7)^4 (y-0.9)^5 ,$$

whose GCD is of degree 13, and thus the rank of $S(\hat{f}, \hat{g})$ is equal 40. The 13th subresultant matrix, corresponding to the value $k = 13$, was selected, and μ was set equal to 10^8. The noisy polynomials $f(y)$ and $g(y)$ are given by

$$f(y) = \hat{f}(y) + \delta\hat{f}(y) \qquad \text{and} \qquad g(y) = \hat{f}(y) + \delta\hat{g}(y),$$

where

$$\mu = \frac{\left\|\hat{f}(y)\right\|}{\left\|\delta\hat{f}(y)\right\|} = \frac{\left\|\hat{g}(y)\right\|}{\left\|\delta\hat{g}(y)\right\|}$$

and, as noted above, $f(y)$ and $g(y)$ are normalised by the geometric mean of their coefficients.[1]

The results of the computational experiments are shown in Figure 5.1. In particular, Figure 5.1-(i) shows the ratio $\|f\|/\mu$, which is the maximum allowable magnitude of the perturbations of the coefficients of $f(y)$, and the variation with α of the computed value of $\|z_f\|$, where both quantities are defined in (4.1). Figure 5.1-(ii) is the same as Figure 5.1-(i), but for $g(y)$ instead of $f(y)$.

It is seen that the first inequality in (4.1) is satisfied for all values of α in the specified range, but the second inequality in (4.1) is only satisfied for $\log_{10}\alpha > 1.72$. This is the minimum value of α for which the bounds on the structured perturbations of the coefficients of $f(y)$ and $g(y)$ are satisfied. Figure 5.1-(iii) shows the variation with α of the normalised residual $\|r_{\text{norm}}\|$, where

$$r_{\text{norm}} = \frac{r(z,x)}{\|d_k + h_k\|},$$

and $r(z,x)$ is defined in (3.2). It is seen that this variation is significant and, in particular, the graph shows that there exist values of α for which the normalised residual is large. It therefore follows that there does not exist a structured matrix E_k and a structured vector h_k such that (3.2) is satisfied for these values of α. By contrast, it is also seen that there exist values of α for which the normalised residual is equal to $O\left(10^{-16}\right)$, which implies that (3.2) is satisfied (to within machine precision). The normalised residual is minimum when $\alpha = 10^{2.8}$, and this is therefore the optimal value of α.

[1] The coefficients of the perturbed polynomials $f(y)$ and $g(y)$, and not the coefficients of the exact polynomials $\hat{f}(y)$ and $\hat{g}(y)$, are normalised since the latter polynomials are not known because of the inexact nature of the data in most practical examples.

ETNA
Kent State University
http://etna.math.kent.edu

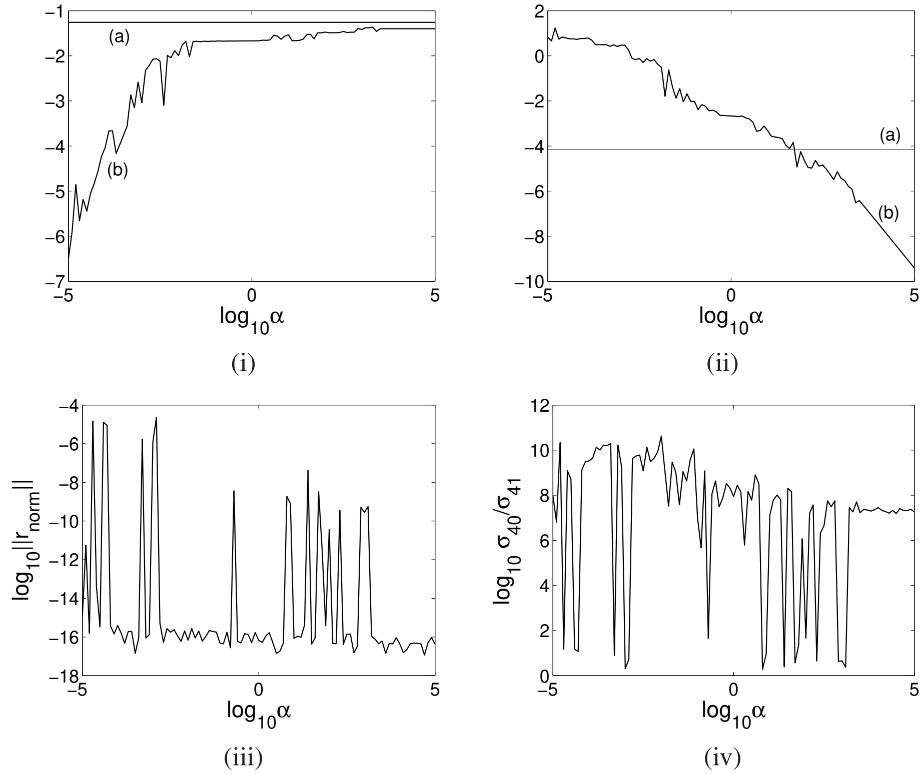

FIGURE 5.1. *The variation with α of: (i) the maximum allowable value of $\|z_f\|$ (a), which is equal to $\|f\|/\mu$, and the computed value of $\|z_f\|$ (b); (ii) the maximum allowable value of $\|z_g\|/\alpha$ (a), which is equal to $\|g\|/\mu$, and the computed value of $\|z_g\|/\alpha$ (b); (iii) the normalised residual $\|r_{norm}\|$; (iv) the singular value ratio σ_{40}/σ_{41}. The horizontal and vertical axes are logarithmic in the four plots.*

The theoretically exact rank of $S(\hat{f},\hat{g})$ is equal to 40, and thus a measure of the effectiveness of the method of STLN is the ratio $\gamma = \sigma_{40}/\sigma_{41}$ of the Sylvester resultant matrix $S(\tilde{f},\tilde{g})$, where $\tilde{f} = \tilde{f}(y)$ and $\tilde{g} = \tilde{g}(y)$ are the polynomials that are computed by the method of STLN, σ_i is the ith singular value of $S(\tilde{f},\tilde{g})$, and the singular values are arranged in non-increasing order. Figure 5.1-(iv) shows the variation of γ with α, and it is seen that it is approximately reciprocal, to within a scale factor, with respect to the variation of $\|r_{\text{norm}}\|$ with α, as shown in Figure 5.1-(iii). In particular, it is seen that large values of γ are associated with small values of $\|r_{\text{norm}}\|$, and that $\gamma = O\left(10^8\right)$ at the optimal value of α. Figures 5.1-(iii) and (iv) clearly show the importance of including the parameter α in the analysis, and specifically, a poor choice of α can lead to unsatisfactory results (the ratio γ is small and the normalised residual $\|r_{\text{norm}}\|$ is large). It is also noted that the default value ($\alpha = 1$) may lead to poor results, as shown in Figure 4.4 in [24].

Figure 5.2 shows the normalised singular values of the Sylvester resultant matrix of the theoretically exact polynomials $\hat{f}(y)$ and $\hat{g}(y)$, the given inexact polynomials $f(y)$ and $g(y)$, and the corrected polynomials $\tilde{f}(y)$ and $\tilde{g}(y)$ for $\alpha = 10^{2.8}$, which is the optimal value of α. All the polynomials are scaled by the geometric mean of their coefficients. It is seen from Figure 5.2-(i) that $S(\hat{f},\hat{g})$ is of full rank, which is incorrect because $\hat{f}$ and $\hat{g}$ are not coprime, and Figure 5.2-(ii) shows that the addition of noise affects its small singular values more than its large singular values. The results for $S(\tilde{f},\tilde{g})$, which are shown in Figure 5.2-(iii), are

ETNA
Kent State University
http://etna.math.kent.edu

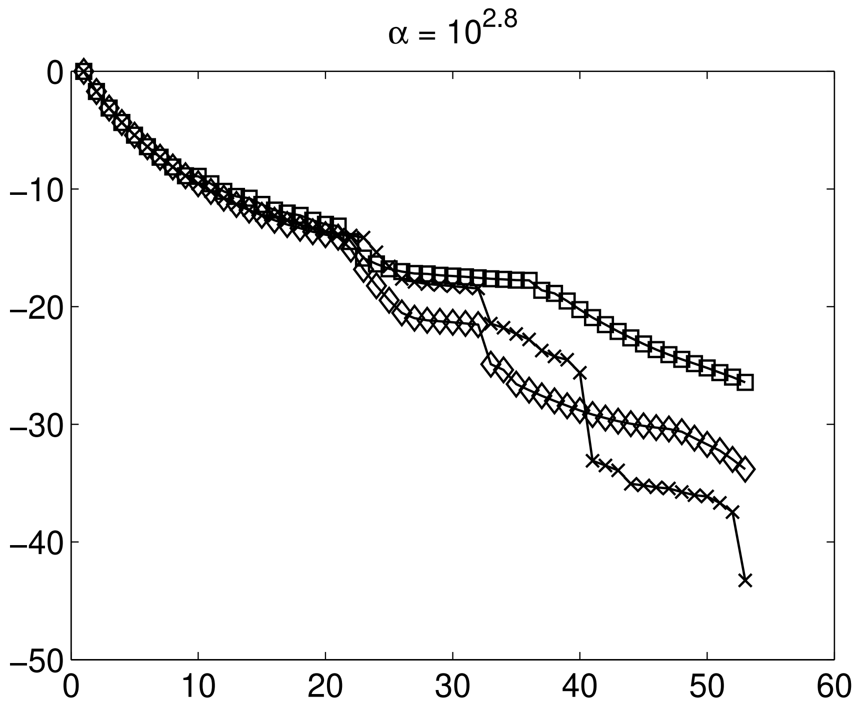

FIGURE 5.2. *The normalised singular values, on a logarithmic scale, of the Sylvester resultant matrix for (i) the theoretically exact polynomials $\hat{f}(y)$ and $\hat{g}(y)$, $\diamond$; (ii) the given inexact polynomials $f(y)$ and $g(y)$, $\square$; (iii) the corrected polynomials $\tilde{f}(y)$ and $\tilde{g}(y)$ for $\alpha = 10^{2.8}$, $\times$. All the polynomials are normalised by the geometric mean of their coefficients.*

significantly better because its (numerical) rank is 40, which is the correct value. Since the perturbations that are used for the formation of this matrix are, by construction, structured, and its rank is equal to 40, this matrix is a structured low rank approximation of $S(f, \alpha g)$, for $\alpha = 10^{2.8}$, that can be used to compute an approximate GCD of $f(y)$ and $g(y)$.

It is seen from Figures 5.1-(iii) and (iv) that there are many values of $\alpha > 10^{1.72}$ such that the ratio γ is large and the normalised residual $\|r_{\text{norm}}\|$ is sufficiently small. There exist therefore many structured low rank approximations of $S(f, \alpha g)$ that satisfy tight bounds on $\|r_{\text{norm}}\|$, which is an error bound for the satisfaction of (3.2), and also satisfy tight bounds on the ratio γ, which is a measure of the numerical rank of $S(\tilde{f}, \tilde{g})$. Each of these approximations yields a different approximate GCD, and additional constraints can be imposed on the optimisation algorithm in order to select a particular structured low rank approximation of $S(f, \alpha g)$, and thus a particular approximate GCD.

This example only considers one value of the signal-to-noise ratio μ and one subresultant matrix. The effects of varying one or both of these parameters is considered in Section 4 in [24]. For example, if the upper bounds of $\|z_f\|$ and $\|z_g\| / \alpha$ are too small because μ is large, then it may not be possible to compute a structured low rank approximation of a Sylvester resultant matrix whose perturbations z_i are within specified bounds.

The computational results in this example, as shown in Figures 5.1 and 5.2, are typical of those obtained from the large number of computational experiments that were performed on different pairs of Bernstein basis polynomials, all of which were of high degree and had a large number of multiple roots. Furthermore, the results in Figures 5.1 and 5.2 for Bernstein basis polynomials are similar to the results in [24] for power basis polynomials.

The integer k is the degree of the common divisor, and this is set by the user. For given error tolerances and signal-to-noise ratio μ, the degree of the approximate GCD is defined by the largest value of k for which the LSE problem has a solution, but it cannot be guaranteed that there exists an approximate GCD of specified degree for given error tolerances and signal-to-noise ratio μ. The absence of an approximate GCD that satisfies these criteria manifests itself by the normalised residual $\|r_{\text{norm}}\|$ being significantly greater than machine precision (10^{-16}), that is, the equality constraint $C\omega = q$ is not satisfied, and the numerical rank of the Sylvester resultant matrix $S(\tilde{f}, \tilde{g})$ is not defined. In this case, it may be necessary to reduce the value of the integer k in order to compute an approximate GCD. Finally, Figures 5.1-(iii) and 5.1-(iv) provide the necessary certification of the computed result because if $\|r_{\text{norm}}\| \approx 10^{-16}$, then the equality constraint is satisfied, and Figure 5.1-(iv) shows that $S(\tilde{f}, \tilde{g})$ is singular at the solution, in which case $\tilde{f}(y)$ and $\tilde{g}(y)$ have a non-constant GCD.

An approximate GCD of the inexact polynomials $f(y)$ and $g(y)$ can be computed from the corrected polynomials $\tilde{f}(y)$ and $\tilde{g}(y)$. The simplest method involves an LU or QR factorisation of the Sylvester resultant matrix $S(\tilde{f}, \tilde{g})$, in which case the coefficients of the GCD are contained in the last non-zero row of the upper triangular matrices U and R. This method does not involve the coprime factors of $\tilde{f}(y)$ and $\tilde{g}(y)$, which must be compared with the method in [28], which involves the solution of a non-linear least squares equation for the GCD and the coprime factors.

6. Future work. The example in Section 5 shows that the method of STLN allows a structured low rank approximation of a Sylvester resultant matrix, and an approximate GCD of two polynomials, to be computed. Although this example, and other examples on high degree polynomials with multiple roots that were obtained but are not included, show that the method of STLN is effective, there exist several issues that must be addressed in order to improve its viability:

- This paper has shown the importance of including the scaling parameter α, but a criterion and algorithm for the computation of its optimal value were not discussed. Figures 5.1-(iii) and 5.1-(iv) are highly oscillatory, and thus a criterion that leads to a non-linear equation for the calculation of the optimal value of α, and for which an initial estimate is required, may not converge to the optimal, or a satisfactory, solution. A possible solution to this problem is the reformulation of the LSE problem, such that instead of minimising $\|Hz\|^2$, the revised objective function

$$(n-k+1)^2 \sum_{i=0}^{m} \left(\frac{z_i}{a_i \binom{m}{i}} \right)^2 + (m-k+1)^2 \sum_{i=0}^{n} \left(\frac{z_{m+1+i}}{b_i \binom{n}{i}} \right)^2$$

  is minimised. This objective function, which is equivalent to the minimisation of the backward error, may be particularly important for polynomials whose coefficients differ by several orders of magnitude.
- The determination of the degree of the GCD, that is, the order of the subresultant matrix, is an important part of the application of the method of STLN to the computation of an approximate GCD of two polynomials. One method involves the application of a threshold to the small singular values, where the value of the threshold is a function of the noise [28]. The noise level may not be known, or it may only be known approximately, which raises disadvantages about these threshold dependent measures. A potential solution to this problem that does not require knowledge of the noise level is the use of minimum description length, which is an information

theoretic measure for the estimation of the rank of a matrix [26].

7. Summary. This paper has considered the use of the method of STLN for the construction of structured low rank approximations of the Sylvester resultant matrix in order to compute approximate GCDs of two inexact polynomials. It has been shown that the method is effective in computing these structured approximations, but that it is necessary to include the parameter α in the optimisation, because failure to do so could lead to unsatisfactory results. This parameter arises because the coefficients of $f(y)$ and $g(y)$ in $S(f, g)$ are decoupled, with the consequence that one polynomial can be scaled relative to the other polynomial.

The important issue of the development of fast algorithms that exploit the structures of the matrices has not been addressed because this paper has only considered the feasibility of the method of STLN, with the new contributions stated in Section 1, for the computation of an approximate GCD of two polynomials. The very good computational results in this paper and [24] show that the development of these fast algorithms must be considered in order to enhance the practical viability of the method.

REFERENCES

[1] J. BARLOW, *Error analysis and implementation aspects of deferred correction for equality constrained least squares problems*, SIAM J. Numer. Anal., 25 (1988), pp. 1340–1358.
[2] J. BARLOW AND U. VEMULAPATI, *A note on deferred correction for equality constrained least squares problems*, SIAM J. Numer. Anal., 29 (1992), pp. 249–256.
[3] S. BARNETT, *Polynomials and Linear Control Systems*, Marcel Dekker, New York, USA, 1983.
[4] D. A. BINI AND L. GEMIGNANI, *Bernstein-Bezoutian matrices*, Theoret. Comput. Sci., 315 (2004), pp. 319–333.
[5] J. F. CANNY, *The Complexity of Robot Motion Planning*, The MIT Press, Cambridge, USA, 1988.
[6] 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, ACM Press, New York, 1995, pp. 195–207.
[7] R. M. CORLESS, S. M. WATT, AND L. ZHI, *QR factoring to compute the GCD of univariate approximate polynomials*, IEEE Trans. Signal Process., 52 (2004), pp. 3394–3402.
[8] I. EMIRIS, A. GALLIGO, AND H. LOMBARDI, *Certified approximate univariate GCDs*, J. Pure Appl. Algebra, 117,118 (1997), pp. 229–251.
[9] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, John Hopkins University Press, Baltimore, USA, 1996.
[10] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, USA, 2002.
[11] J. T. KAJIYA, *Ray tracing parametric patches*, Comput. Graphics, 16 (1982), pp. 245–254.
[12] E. KALTOFEN, Z. YANG, AND L. ZHI, *Structured low rank approximation of a Sylvester matrix*, in Trends in Mathematics, D. Wang and L. Zhi, eds., Birkhäuser Verlag, Basel, Switzerland, 2006, pp. 69–83.
[13] N. KARMARKAR AND Y. N. LAKSHMAN, *Approximate polynomial greatest common divsior and nearest singular polynomials*, in Proc. Int. Symp. Symbolic and Algebraic Computation, ACM Press, New York, 1996, pp. 35–39.
[14] 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 (2005), pp. 165–174.
[15] C. V. LOAN, *On the method of weighting for equality-constrained least squares problems*, SIAM J. Numer. Anal., 22 (1985), pp. 851–864.
[16] V. Y. PAN, *Computation of approximate polynomial GCDs and an extension*, Information and Computation, 167 (2001), pp. 71–85.
[17] S. PETITJEAN, *Algebraic geometry and computer vision: Polynomial systems, real and complex roots*, J. Math. Imaging Vision, 10 (1999), pp. 191–220.
[18] J. B. ROSEN, H. PARK, AND J. GLICK, *Total least norm formulation and solution for structured problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 110–128.
[19] D. RUPPRECHT, *An algorithm for computing certified approximate GCD of n univariate polynomials*, J. Pure Appl. Algebra, 139 (1999), pp. 255–284.
[20] J. V. USPENSKY, *Theory of Equations*, McGraw-Hill, New York, USA, 1948.
[21] J. R. WINKLER, *A companion matrix resultant for Bernstein polynomials*, Linear Algebra Appl., 362 (2003), pp. 153–175.

[22] ———, *The transformation of the companion matrix resultant between the power and Bernstein polynomial bases*, Appl. Numer. Math., 48 (2004), pp. 113–126.

[23] ———, *Numerical and algebraic properties of Bernstein basis resultant matrices*, in Computational Methods for Algebraic Spline Surfaces, T. Dokken and B. Jüttler, eds., Germany, 2005, Springer-Verlag, pp. 107–118.

[24] J. R. WINKLER AND J. D. ALLAN, *Structured total least norm and approximate GCDs of inexact polynomials*, J. Comput. Appl. Math., 215 (2008), pp. 1–13.

[25] J. R. WINKLER AND R. N. GOLDMAN, *The Sylvester resultant matrix for Bernstein polynomials*, in Curve and Surface Design: Saint-Malo 2002, T. Lyche, M. Mazure, and L. L. Schumaker, eds., Tennessee, USA, 2003, Nashboro Press, pp. 407–416.

[26] C. J. ZAROWSKI, *The MDL criterion for rank determination via effective singular values*, IEEE Trans. Signal Process., 46 (1998), pp. 1741–1744.

[27] 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 Process., 48 (2000), pp. 3042–3051.

[28] Z. ZENG, *Computing multiple roots of inexact polynomials*, Math. Comp., 74 (2005), pp. 869–903.

[29] L. ZHI AND Z. YANG, *Computing approximate GCD of univariate polynomials by structured total least norm*, tech. report, Institute of Systems Science, AMSS, Academia Sinica, Beijing, China, 2004.