Fast Low Rank Approximation of a Sylvester Matrix

Bingyu Li, Zhuojun Liu and Lihong Zhi
Key Lab of Mathematics Mechanization
AMSS, Academia Sinica
Beijing 100080, China
email: { liby, zliu, lzhi}@mmrc.iss.ac.cn

Extended Abstract for SNC 2005

In [2, 8], authors described an algorithm based on STLN [5] for constructing a Sylvester matrix of given lower rank and obtaining the nearest perturbed polynomials with exact GCD of given degree. For their algorithm, the overall computation time depends on solving a sequence least squares(LS) problems. In the paper [4], based on the displacement structure of the coefficient matrices, the authors described a fast algorithm using the generalized Schur algorithm [1] for solving the LS problems and deriving a fast version of the approximate GCD algorithm. In this report, we present a new generator pair of the augment matrix (5) and show the performance of the fast algorithm.

Given two polynomials a, $b \in \mathbb{C}[x]$ with $a = a_m x^m + \cdots + a_1 x + a_0$ and $b = b_n x^n + \cdots + b_1 x + b_0$, $a_m \neq 0, b_n \neq 0$. S is the Sylvester matrix for a and b. The perturbations of a and b are denoted by $\Delta a = \Delta a_m x^m + \cdots + \Delta a_1 x + \Delta a_0$, $\Delta b = \Delta b_n x^n + \cdots + \Delta b_1 x + \Delta b_0$ respectively. We consider the minimal perturbation problem: *minimize* $\|\Delta a\|_2^2 + \|\Delta b\|_2^2$ *preserving that* $a + \Delta a$ *and* $b + \Delta b$ *have an exact GCD of a given degree.*

Let us denote $S_k = [\mathbf{a}\ A_k]$ as the k-th Sylvester matrix,

$$S_k = \begin{pmatrix} a_m & 0 & \cdots & 0 & 0 & b_n & 0 & \cdots & 0 & 0 \\ a_{m-1} & a_m & \cdots & 0 & 0 & b_{n-1} & b_n & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_0 & a_1 & 0 & 0 & \cdots & b_0 & b_1 \\ 0 & 0 & \cdots & 0 & a_0 & 0 & 0 & \cdots & 0 & b_0 \end{pmatrix},$$
$$\underbrace{\qquad\qquad\qquad}_{n-k+1} \quad \underbrace{\qquad\qquad\qquad}_{m-k+1}$$

*The work is Partially supported by a National Key Basic Research Project of China 2004CB318000 and Chinese National Science Foundation under Grant 10371127 and 10401035.

where $\mathbf{a}$ is the first column of S_k, A_k consists of the last $m+n-2k+1$ columns of S_k.

The perturbations Δa and Δb are expressed by a $m+n+2$-dimensional vector $\mathbf{d}$,

$$\mathbf{d} = (d_1, d_2, \ldots, d_{m+n+1}, d_{m+n+2})^T.$$

The k-th Sylvester structured perturbation of S_k is represented as $[\Delta\mathbf{a}\, D_k]$.

The following two theorems were given in [2, 8].

Theorem 1 *Given* $a(x), b(x) \in \mathbb{C}[x]$ *with* $\deg(a) = m$ *and* $\deg(b) = n$*. Let* $S(a,b)$ *be the Sylvester matrix of* $a(x)$ *and* $b(x)$*,* S_k *be the* k*-th Sylvester matrix,* $1 \le k \le \min(m,n)$*. Then the following statements are equivalent:*

(1) $\deg(\gcd(a,b)) \ge k$.

(2) Rank deficiency of S_k *is not less than one.*

Theorem 2 *In the same assumption as in Theorem 1, let* $S_k = [\mathbf{a}\ A_k]$*, where* $\mathbf{a}$ *is the first column of* S_k*, then* $\deg(\gcd(a,b)) \ge k$ *if and only if* $A_k\mathbf{x} = \mathbf{a}$ *has a solution.*

Based on the above two theorems we know that, for a given degree k, it is always possible to find a k-th Sylvester structured perturbation matrix $[\Delta\mathbf{a}\, D_k]$ such that $\mathbf{a} + \Delta\mathbf{a} \in \mathrm{Range}(A_k + D_k)$. The minimal perturbation problem can be formulated as the following equality-constrained least squares problem:

$$\min_{\mathbf{x}} \|\mathbf{d}\|_2, \quad \text{subject to } \mathbf{r} = 0, \tag{1}$$

where the structured residual $\mathbf{r}$ is given by

$$\mathbf{r} = \mathbf{a} + \Delta\mathbf{a} - (A_k + D_k)\mathbf{x}.$$

Applying the penalty method [5], we transform (1) into

$$\min_{\mathbf{d},\mathbf{x}} \left\| \begin{pmatrix} w\mathbf{r} \\ \mathbf{d} \end{pmatrix} \right\|_2, \tag{2}$$

where w is a large penal value.

The vectors $\Delta\mathbf{a}$ and $D_k\mathbf{x}$ can be expressed as

$$\Delta\mathbf{a} = P_k\,\mathbf{d}, \;\; D_k\mathbf{x} = X_k\mathbf{d}.$$

Here

$$P_k = \begin{pmatrix} I_{m+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where I_{m+1} is an identity matrix of order $m+1$, and

$$X_k = \begin{pmatrix} \begin{matrix} 0 & & \\ x_1 & \ddots & \\ \vdots & \ddots & 0 \\ x_{n-k} & & x_1 \\ & \ddots & \vdots \\ & & x_{n-k} \end{matrix} & \begin{matrix} x_{n+1-k} & & \\ x_{n+2-k} & \ddots & \\ \vdots & \ddots & x_{n+1-k} \\ x_{m+n+1-2k} & & x_{n+2-k} \\ & \ddots & \vdots \\ & & x_{m+n+1-2k} \end{matrix} \end{pmatrix}.$$

$$\underbrace{\qquad\qquad\qquad}_{m+1} \quad \underbrace{\qquad\qquad\qquad\qquad\qquad}_{n+1}$$

Then (2) becomes the following least squares problem:

$$\min_{\Delta\mathbf{x}\,\Delta\mathbf{d}} \left\| \begin{pmatrix} w(X_k - P_k) & w(A_k + D_k) \\ I_{m+n+2} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{d} \\ \Delta\mathbf{x} \end{pmatrix} + \begin{pmatrix} -w\mathbf{r} \\ \mathbf{d} \end{pmatrix} \right\|_2, \tag{3}$$

where I_{m+n+2} is an identity matrix of order $m+n+2$.

Let us denote the coefficient matrix of the system in (3) by M,

$$M = \begin{pmatrix} w(X_k - P_k) & w(A_k + D_k) \\ I_{m+n+2} & \mathbf{0} \end{pmatrix},$$

the following result was given in [4].

Theorem 3 *M is a structured matrix of displacement rank at most* 4.

Let $\mathbf{y} = \begin{pmatrix} \Delta\mathbf{d} \\ \Delta\mathbf{x} \end{pmatrix}$, $\mathbf{z} = \begin{pmatrix} w\mathbf{r} \\ -\mathbf{d} \end{pmatrix}$, the least squares problem (3) can be rewritten as

$$\min_{\mathbf{y}} \|M\mathbf{y} - \mathbf{z}\|_2. \tag{4}$$

As in [1, 3], in order to solve (4), we consider the augmented matrix

$$T = \begin{pmatrix} M^T M & M^T \\ M & 0 \end{pmatrix}. \tag{5}$$

Theorem 4 *The matrix T is a structured matrix of displacement rank at most* 8.

We construct a generator pair (G, F) for T so that $T - FTF^T = GJG^T$, where

$$F = \mathrm{diag}(Z_{m+1}, Z_{n+1}, Z_{n-k}, Z_{m-k+1}, Z_{m+n-k+1}, Z_{m+n+2}),$$

$J = \mathrm{diag}(I_4, -I_4)$, and G consists of eight columns. In [4], we express G as columns of T which are of order w^2 where w is the large penal value. It is undesirable for numerical stability. In the following, we introduce a new generator matrix of T which is of order w. Let $t_1, \ldots, t_4$ be:

$$\begin{aligned} t_1 &= \|M(:,1)\|_2,\ t_2 = \|M(:,m+2)\|_2, \\ t_3 &= \|M(:,m+n+3)\|_2,\ t_4 = \|M(:,2n+m-k+3)\|_2. \end{aligned}$$

The eight columns of G are given as:

$$\begin{aligned}
\mathbf{g}_1 &= T(1,:)/t_1, \\
\mathbf{g}_2 &= T(m+2,:)/t_2, \quad \text{except that} \quad \mathbf{g}_2[1] = 0, \\
\mathbf{g}_3 &= T(m+n+3,:)/t_3, \quad \text{except that } \mathbf{g}_3[1] = 0,\ \mathbf{g}_3[m+2] = 0, \\
\mathbf{g}_4 &= T(2n+m-k+3,:)/t_4, \quad \text{except that } \mathbf{g}_4[1] = 0,\ \mathbf{g}_4[m+2] = 0, \\
&\quad\ \mathbf{g}_4[m+n+3] = 0, \\
\mathbf{g}_5 &= [0, \mathbf{g}_1^T(2:4m+4n-3k+6)]^T, \\
\mathbf{g}_6 &= [\mathbf{g}_2^T(1:m+1), 0, \mathbf{g}_2^T(m+3:4m+4n-3k+6)]^T, \\
\mathbf{g}_7 &= [\mathbf{g}_3^T(1:m+n+2), 0, \mathbf{g}_3^T(m+n+4:4m+4n-3k+6)]^T, \\
\mathbf{g}_8 &= [\mathbf{g}_4^T(1:2n+m-k+2), 0, \mathbf{g}_4^T(2n+m-k+4:4m+4n-3k+6)]^T.
\end{aligned}$$

As discussed in [1], we need to make the following normalization assumption:

$$\|M\|_2 \le \frac{1}{5},$$

which can always be guaranteed by scaling M as

$$M = M/5\|M\|_F.$$

Here $\|M\|_F$ is the Frobenius norm of M. Due to the large penal value w, after normalization, the identity I_{m+n+2} (down-left corner of M) has very small diagonal elements. This causes the rank deficiency of M^TM and leads to early breakdown of the generalized Schur algorithm directly applied to (G, F). Moreover, M has more rows than columns for $k \ge 1$, so the Schur complement of T with respect to M^TM is only negative semi-definite. In order to complete the positive and negative steps of the generalized Schur algorithm successfully, we consider the perturbed matrix:

$$\bar{T} = \begin{pmatrix} M^T M + \alpha I & M^T \\ M & -\beta I \end{pmatrix}.$$

The displacement rank of $\bar{T}$ increases by two, the new generators are given as follows:

$$\bar{T} - F\bar{T}F^T = \bar{G}\bar{J}\bar{G}^T,$$

where $\bar{J} = \mathrm{diag}(I_4, -I_6)$. The first 8 columns of $\bar{G}$ have similar forms as G, except that we need update $t_1, \ldots, t_4$ as:

$$\begin{aligned} t_1 &= \sqrt{T[1,1]+\alpha},\ t_2 = \sqrt{T[m+2,m+2]+\alpha}, \\ t_3 &= \sqrt{T[m+n+3,m+n+3]+\alpha}, \\ t_4 &= \sqrt{T[2n+m-k+3,2n+m-k+3]+\alpha}. \end{aligned}$$

Another two extra columns of $\bar{G}$ are:

$$\begin{aligned} \bar{\mathbf{g}}_9 &= [\ \underbrace{0,\ \ldots,\ 0}_{2m+2n-2k+3},\sqrt{\beta},\ 0,\ \ldots,\ 0,\ 0\ ]^T, \\ \bar{\mathbf{g}}_{10} &= [\ \underbrace{0,\ \ldots,\ 0,\ 0}_{3m+3n-3k+4},\ \sqrt{\beta},\ 0,\ \ldots,\ 0\ ]^T. \end{aligned}$$

After applying $2m+2n-2k+3$ positive steps and $2m+2n-k+3$ negative steps of the generalized Schur algorithm to the generator pair $(\bar{G}, \bar{J})$, we can obtain a backward stable triangular factorization of T:

$$\begin{pmatrix} \hat{R}^T & 0 \\ \hat{Q} & \hat{D} \end{pmatrix} \begin{pmatrix} \hat{R} & \hat{Q}^T \\ 0 & -\hat{D}^T \end{pmatrix}, \tag{6}$$

where $\hat{R}$ is upper triangular and $\hat{D}$ is lower triangular. The solution $\mathbf{y}$ is obtained as

$$\mathbf{y} = \hat{R}^{-1}\hat{Q}^T\hat{D}^{-T}\hat{D}^{-1}\mathbf{z}. \tag{7}$$

In the following table, we show the performance of the classic algorithms and fast algorithms for computing ϵ-GCD of univariate polynomials randomly generated in Maple 9 with $\mathit{Digits} = 10$ and $\alpha = \beta = 10^{-9}$.

Ex	m,n	k	σ_k	*error (Zeng)*	*error (classic)*	*error (fast)*	*error*
1	$2,2$	1	$1.08e-2$	$2.29e-4$	$1.047e-4$	$1.08e-4$	$1.048e-4$
2	$3,2$	2	$1.95e-2$	$5.96e-4$	$3.77e-4$	$3.85e-4$	$3.77e-4$
3	$10,10$	5	$4.3e-3$	$5.91e-4$	$1.082e-4$	$1.16e-4$	$1.09e-4$
4	$3,3$	2	$1.05e-1$	$1.91e-2$	$9.49e-3$	$9.54e-3$	$9.49e-3$
5	$6,4$	3	$6.38e-2$	$1.30e-2$	$3.99e-3$	$4.21e-3$	$3.99e-3$
6	$6,4$	3	$4.14e-1$	$5.74e-1$	$2.39e-1$	$2.47e-1$	$2.39e-1$
7	$28,28$	8	$8.15e-3$	$2.16e-3$	$1.328e-4$	$1.334e-4$	$1.328e-4$
8	$14,13$	7	$1.31e-2$	$5.67e-4$	$2.832e-4$	$3.014e-4$	$2.832e-4$

Let m, n denote degrees of the polynomials a and b. k is a given positive integer. σ_k is the last k-th singular value of $S(a, b)$. *error (Zeng)*, *error (classic)*, *error (fast)* and *error* denote the backward errors of computing GCD and cofactors by the algorithms given in [7], [2], [4] and this report respectively. The overall computation time of our fast algorithm is $O(s^2+st+t^2)$, where s, t denote the row dimension $2m+2n-k+3$ and column dimension $2m+2n-2k+3$ of the involving coefficient matrix respectively. Since we always need nonzero small perturbations α and β to avoid the breakdown of the generalized Schur algorithm, this complicates the error analysis of the fast algorithm. The detail error analysis of our new algorithm is still under investigation.

References

[1] S.Chandrasekaran and A.H.Sayed, A fast stable solver for nonsymmetric Toeplitz and quasi-Toeplitz systems of linear equations, *SIMAX,* Vol.19, No.1, 1998, pp. 107-139.

[2] E. Kaltofen, Z. Yang and L. Zhi, Structured low rank approximation of a Sylvester matrix. Submitted to SNC2005.

[3] Thomas Kailath and Ali H. Sayed: Displacement structure: theory and applications. *SIAM Review.* **37** (1995) 297–386.

[4] Bingyu Li, Zhengfeng Yang and Lihong Zhi: Fast low rank approximation of a Sylvester matrix by Structured Total Least Norm. *J.JSSAC.* **11** (2005) 165–174.

[5] Park, H., Zhang, L. and Rosen, J.B., Low rank approximation of a Hankel matrix by structured total least norm, *BIT Numerical Mathematics*, vol.39: pp. 757–779,1999.

[6] Rosen, J.B., Park, H. and Glick, J., Total least norm formulation and solution for structured problems, *SIAM Journal on Matrix Anal. Appl.* vol.17-1: pp. 110–128, 1996.

[7] Zeng, Z. The approximate GCD of inexact polynomials. Part I: a univariate algorithm. Manuscript, 2004.

[8] Lihong Zhi and Zhengfeng Yang, Computing approximate GCD of univariate polynomials by structured total least norm. In*MM Research Preprints, 11-19, No.23, December 2004.*