The extended block Arnoldi method for solving generalized differential Sylvester equations

Laklifa Sadek∗, Hamad Talibi Alaoui
Faculté des Sciences, University Chouaib Doukkali, Morocco
Email(s): sadek.l@ucd.ac.ma, talibi_1@hotmail.fr

Abstract. In the present paper, we propose a new method for solving large-scale generalized differential Sylvester equations, by projecting the initial problem onto the extended block Krylov subspace with an orthogonality Galerkin condition. This projection gives rise to a low-dimensional generalized differential Sylvester matrix equation. The low-dimensional equations is then solved by Rosenbrock or BDF method. We give some theoretical results and report some numerical experiments to show the effectiveness of the proposed method.

Keywords: Extended block Krylov subspace, Generalized differential Sylvester matrix equation, low-rank approximate solution, Rosenbrock method, BDF method.

AMS Subject Classification 2010: 65F50, 15A24.

1 Introduction

In the present paper, we consider the generalized differential Sylvester matrix equation (GDSME) of the form

\[
\begin{aligned}
\dot{X}(t) &= AX(t) + X(t)B^T + \sum_{i=1}^{k} N_i X(t) M_i^T - EF^T, \quad t \in [t_0, T_f] \\
X(t_0) &= X_0,
\end{aligned}
\]  

(1)
where \( A, N_i \in \mathbb{R}^{n \times n}, B, M_i \in \mathbb{R}^{p \times p}, \) and \( E \in \mathbb{R}^{n \times s}, F \in \mathbb{R}^{p \times s}, \) with \( s \ll n, p. \) The matrices \( A \) and \( B \) are assumed to be large, sparse, and nonsingular.

Generalized differential Sylvester matrix equations play a fundamental role in many problems in control, filter design theory, model reduction problems, differential equations and robust control problems; see, [1–3,5,6,9,11,12,16,20,23] and the references therein. For small or medium-sized differential Sylvester matrix equations, there are several methods to solve this equation, for example Backward Differentiation Formula (BDF) method and Rosenbrock method [6, 10, 16, 22]. For large generalized differential Sylvester matrix equations, we propose a new method based on projection onto extended block Krylov subspaces [3,7,8,14,17,18,24] with an orthogonality Galerkin condition.

The rest of the paper is organized as follows. In Section 2, we recall the extended block Arnoldi process with some of its properties. In Section 3, we give a low-rank method for solving large-scale generalized differential Sylvester equations, by using projections onto extended block Krylov subspaces \( \mathcal{K}_m^e(A,E) \) and \( \mathcal{K}_m^e(B,F), \) and Galerkin orthogonality condition. Then, in Section 4, we give some iterative methods for solving the obtained low dimensional problem. Finally, Section 5 is devoted to numerical experiments.

Throughout the paper, we use the following notations. The Frobenius inner product of the matrices \( X \) and \( Y \) is defined by \( \langle X,Y \rangle_F = \text{tr}(X^TY), \) where \( \text{tr}(Z) \) denotes the trace of a square matrix \( Z. \) The associated norm is the Frobenius norm denoted by \( \| \cdot \|_F. \)

## 2 The extended block Arnoldi process

We will consider the extended block Krylov subspaces associated to the pair \( (A,E) \) which is defined as follows

\[
\mathcal{K}_m^e(A,E) = \text{range}\{ E, A^{-1}E, AE, A^{-2}E, A^2E, \ldots, A^{m-1}E, A^{-m}E \}.
\]

We recall the extended block Arnoldi (EBA) [3,7,8,17] algorithm, when applied to the pair \( (A,E) \). The extended block Arnoldi is described in Algorithm 1 as follows:

After \( m \) steps, Algorithm 1 builds an orthonormal basis \( V_m = [V_1, \ldots, V_m] \) of the extended block Krylov subspace \( \mathcal{K}_m^e(A,E). \) Let \( T_{m,A} = V_m^T A V_m \) be a \( 2s \times 2s \) block upper Hessenberg matrix. Then we have the following
Algorithm 1 The extended block Arnoldi algorithm (EBA)

Inputs: $A$ an $n \times n$ matrix, $E$ an $n \times s$ matrix and $m$ an integer.

1. Compute the $QR$ decomposition of $[E, A^{-1}E]$, i.e, $[E, A^{-1}E] = V_1 \Lambda$;
2. Set $V_0 = []$;
3. For $j = 1, 2, 3, \ldots, m$
4. Set $V^{(1)}_j = V_j(:, 1 : s)$ and $V^{(2)}_j = V_j(:, s + 1 : 2s)$
5. $V_j = [V_{j-1}, V_j]$; $\hat{V}_{j+1} = [AV^{(1)}_j, A^{-1}V^{(2)}_j]$;
6. $V_j = [V_j - 1, V_j]$; $\hat{V}_{j+1} = \hat{V}_{j+1} - V_i H_{i,j}$;
7. $H_{i,j} = V^T_i \hat{V}_{j+1}$;
8. $\hat{V}_{j+1} = \hat{V}_{j+1} - V_i H_{i,j}$;
9. End For $i$
10. Compute the $QR$ decomposition of $U$ i.e., $\hat{V}_{j+1} = V_{j+1} H_{j+1,j}$;
11. End For $j$.

Output: $V_m = [V_1, \ldots, V_m]$.

relations

$$AV_m = V_m T_{m,A} + V_{m+1} T^A_{m+1,m} E_m^T = V_{m+1} \begin{bmatrix} T_{m,A} & T^A_{m+1,m} E_m^T \end{bmatrix},$$

where $E_m^T = [0_{2s \times 2s(m-1)}, I_{2s}]$ is the matrix formed by the last $2s$ columns of the $2ms \times 2ms$ identity matrix $I_{2ms}$.

3 Low rank approximate solutions

In this section, we show how to obtain low rank approximate solutions to the generalized differential Sylvester equation (1) by first projecting directly the initial problem onto extended block Krylov subspaces and then solve the obtained low dimensional differential problem. We first apply the extended block Arnoldi algorithm to the pairs $(A, E)$ and $(B, F)$ to get the orthonormal matrices $V_m$ and $W_m$, whose columns form orthonormal bases of the extended block Krylov subspaces $K^e_m(A, E)$ and $K^e_m(B, F)$, respectively. We also get the upper block Hessenberg matrices $T_{m,A} = V^T_m A V_m$ and $T_{m,B} = W^T_m B W_m$. After $m$ iterations, we consider the low rank approximate solutions $X_m(t)$ of exact solution $X(t)$ to equation (1) of the form

$$X_m(t) = V_m Y_m(t) W_m^T. \quad (2)$$

The matrix $Y_m(t)$ can be obtained by the Petrov-Galerkin orthogonality condition

$$V^T_m R_m(t) W_m = 0, \quad t \in [t_0, T_f], \quad (3)$$
where

\[ R_m(t) = \dot{X}_m(t) - AX_m(t) - X_m(t)B^T - \sum_{i=1}^{k} N_i X_m(t) M_i^T + EF^T. \]  

(4)

Using this condition and the relation (2), we obtain the reduced generalized differential Sylvester matrix equation

\[ \dot{Y}_m(t) = T_{m,A} Y_m(t) + Y_m(t)T_{m,B}^T + \sum_{i=1}^{k} N_{i,m} Y_m(t) M_i^T - \tilde{E}_m \tilde{F}_m^T, \]  

(5)

where

\[
\begin{align*}
N_{i,m} &= V_{m+1}^T N_i V_{m}, \\
M_{i,m} &= W_{m+1}^T M_i W_{m}, \\
\tilde{E}_m &= V_{m+1}^T E, \\
\tilde{F}_m &= W_{m+1}^T F.
\end{align*}
\]

Next, we give a result that allows us the computation of the norm of the residual without forming the approximation \(X_m(t)\) at each step of the extended block Arnoldi process. The approximation \(X_m(t)\) is computed in a factored form only when convergence is achieved.

**Theorem 1.** Let \(X_m(t) = V_m Y_m(t) W_m^T\) be the approximation obtained at step \(m\) by extended block Arnoldi process. Then, the Frobenius norm of the residual \(R_m(t)\) associated to the approximation \(X_m(t)\) satisfies the relation

\[
\|R_m(t)\|_F = \sqrt{\|T_{m+1,m}^{A} Y_m^{(1)}(t)\|_F^2 + \|T_{m+1,m}^{B} Y_m^{(2)}(t)\|_F^2},
\]

(6)

where \(Y_m^{(1)}(t)\) and \(Y_m^{(2)}(t)\) are the \(2s \times 2ms\) matrix corresponding to the last \(2s\) rows of \(Y_m(t)\) and \(Y_m^T(t)\) respectively.

**Proof.** We have

\[
R_m(t) = \dot{V}_m \dot{Y}_m(t) W_m^T - AV_m Y_m(t) W_m^T - V_m Y_m(t) W_m^T B^T + \sum_{i=1}^{k} N_i V_m Y_m(t) W_m^T M_i^T - EF^T.
\]

Using the relations

\[
\begin{align*}
AV_m &= V_{m+1} \begin{bmatrix} T_{m,A} \\ T_{m+1,m}^{A} E_m \end{bmatrix}, \\
W_m B^T &= \begin{bmatrix} T_{m,B}^T \\ T_{m+1,m}^{B} E_m(T_{m+1,m}^{B}) \end{bmatrix} W_m^T, \\
V_m &= V_{m+1} \begin{bmatrix} I_{2sm} \\ 0_{2s \times 2ms} \end{bmatrix},
\end{align*}
\]


and $Y_m(t)$ is the solution of reduced generalized differential Sylvester equation (5), we get

$$R_m(t) = \mathcal{V}_{m+1} \begin{bmatrix} 0_{2ms \times 2ms} & -Y_m(t)E_m \left(T_{m+1,m}^B\right)^T \\ -T_{m+1,m}^A \mathcal{E}_m^T Y_m(t) & 0_{2s \times 2s} \end{bmatrix} \mathcal{W}_{m+1}^T.$$  

Let $Y_m^{(1)}(t) := E_m^T Y_m(t)$ and $Y_m^{(2)}(t) := E_m^T (Y_m(t))^T$, so

$$R_m(t) = \mathcal{V}_{m+1} \begin{bmatrix} 0_{2ms \times 2ms} & -Y_m^{(2)}(t) \left(T_{m+1,m}^B\right)^T \\ -T_{m+1,m}^A Y_m^{(1)}(t) & 0_{2s \times 2s} \end{bmatrix} \mathcal{W}_{m+1}^T.$$  

Since $\mathcal{V}_{m+1}$ and $\mathcal{W}_{m+1}$ are the orthonormal matrices, we have

$$\|R_m(t)\|_F = \sqrt{\|T_{m+1,m}^A Y_m^{(1)}(t)\|^2_F + \|T_{m+1,m}^B Y_m^{(2)}(t)\|^2_F},$$  

which completes the proof.

To save memory, the solution $X_m(t) = \mathcal{V}_m Y_m(t) \mathcal{W}_m^T$ can be given as a product of two matrices of low-rank. For that, we consider the singular value decomposition of the $2ms \times 2ms$ matrix $Y_m = UDV^T$, where $D$ is the diagonal matrix of the singular values of $Y_m(t)$ sorted in decreasing order. Let $U_i$ and $V_i$ be the $2ms \times l$ matrix of the first $l$ columns of $U$ and $V$ respectively, corresponding to the $l$ singular values of magnitude greater than some tolerance $dtol$. We obtain the truncated singular value decomposition $Y_m \approx U_i D_i V_i^T$ where $D_i = \text{diag}[\lambda_1, \ldots, \lambda_l]$. Setting $\tilde{Z}_{m,1} = \mathcal{V}_m U_i D_i^{\frac{1}{2}}$ and $\tilde{Z}_{m,2} = \mathcal{W}_m V_i D_i^{\frac{1}{2}}$ it follows that

$$X_m = \tilde{Z}_{m,1} \tilde{Z}_{m,2}^T.$$  

The following result shows that the approximation $X_m(t)$ is an exact solution of a perturbed generalized Sylvester differential equation.

**Theorem 2.** Let $X_m(t) = \mathcal{V}_m Y_m(t) \mathcal{W}_m^T$ be the approximate solution obtained after running $m$ steps of the extended block Arnoldi process. Then we have

$$\dot{X}_m(t) = (A - F_{m,A})X_m(t) + X_m(t)(B - F_{m,B})^T + \sum_{i=1}^k N_i X_m(t)M_i^T - EF^T,$$

where $F_{m,A} = \mathcal{V}_{m+1}T_{m+1,m}^A \mathcal{V}_m^T$ and $F_{m,B} = \mathcal{W}_{m+1}T_{m+1,m}^B \mathcal{W}_m^T$. 

Proof. By multiplying from the left Eq. (5) by $V_m$ and from the right by $W_m^T$, we obtain

$$
\dot{X}_m(t) = (AV_m - V_{m+1}T_{m+1,m}^A E_m^T)Y_m(t)W_m^T + V_m Y_m(t)(BW_m - W_{m+1}E_m^T) + \sum_{i=1}^{k} N_i X_m(t) M_i^T - EF^T.
$$

On the other hand, since $V_m^T V_m = I_{2ms}$, and $E_m^T V_m = V_m^T$, we have $E_m^T Y_m(t) = V_m^T Y_m(t)$. Then, we have

$$
\dot{X}_m(t) = (A - F_{m,A})X_m(t) + X_m(t)(B - F_{m,B})^T + \sum_{i=1}^{k} N_i X_m(t) M_i^T - EF^T,
$$

where $F_{m,A} = V_{m+1}T_{m+1,m}^A V_m^T$ and $F_{m,B} = W_{m+1}T_{m+1,m}^B W_m^T$.

The following result indicates that the error matrix $E_m(t) = X(t) - X_m(t)$ satisfies a generalized differential Sylvester equation.

**Theorem 3.** Let $X_m(t) = V_m Y_m(t) W_m^T$ and $E_m(t) = X(t) - X_m(t)$. Then we have

$$
\hat{E}_m(t) = A E_m(t) + E_m(t) B^T + \sum_{i=1}^{k} N_i E_m(t) M_i^T - R_m(t). \tag{10}
$$

Proof. According to (1) and (4), we obtain

$$
\hat{E}_m(t) = \dot{X}(t) - \dot{X}_m(t)
$$

$$
= AX(t) + X(t)B^T + \sum_{i=1}^{k} N_i X(t) M_i^T - EF^T - AX_m(t) - X_m(t)B^T
$$

$$
- \sum_{i=1}^{k} N_i X_m(t) M_i^T + EF^T - R_m(t)
$$

$$
= A(X(t) - X_m(t)) + (X(t) - X_m(t))B^T + \sum_{i=1}^{k} N_i (X(t) - X_m(t)) M_i^T
$$

$$
- R_m(t)
$$

$$
= A E_m(t) + E_m(t) B^T + \sum_{i=1}^{k} N_i E_m(t) M_i^T - R_m(t).
$$

So the proof is complete.
The error $E_m(t)$ satisfies in the following differential equation

$$
\dot{E}_m(t) = A E_m(t) + E_m(t) B^T + \sum_{i=1}^{k} N_i E_m(t) M_i^T - R_m(t).
$$

Equation (10) is equivalent to

$$
\begin{align*}
\dot{E}_m(t) &= A E_m(t) - b_m(t), \\
E_0 &= vec(E_m(t_0)).
\end{align*}
$$

where

$$
\begin{align*}
A &= I_p \otimes A + B \otimes I_n + \sum_{i=1}^{k} M_i \otimes N_i, \\
E_m(t) &= vec(E_m(t)), \\
b_m(t) &= vec(R_m(t)).
\end{align*}
$$

The solution of (11) is given by (see for example [1,23])

$$
E_m(t) = e^{(t-t_0)A} E_0 - \int_{t_0}^{t} e^{(t-\tau)A} b_m(\tau) d\tau, \quad t \in [t_0, T_f].
$$

The 2-logarithmic norm of the matrix $A$ is defined by $\mu_2(A) = \lambda_{\max}(A + A^T)/2$. The 2-logarithmic norm satisfies the following property for the matrix exponential $\|e^{tA}\|_2 \leq e^{\mu_2(A)t}$, $t \geq 0$. In the following result, we give an upper bound for the norm of the error $E_m(t) = X(t) - X_m(t)$:

$$
\begin{align*}
\|E_m(t)\|_2 &\leq \|e^{(t-t_0)A} E_0\|_2 - \int_{t_0}^{t} e^{(t-\tau)A} b_m(\tau) d\tau_2 \\
&\leq \|e^{(t-t_0)A} E_0\|_2 + \int_{t_0}^{t} e^{(t-\tau)A} b_m(\tau) d\tau_2 \\
&\leq e^{(t-t_0)\mu_2(A)} \|E_0\|_2 + \int_{t_0}^{t} e^{(t-\tau)\mu_2(A)} \|b_m(\tau)\|_2 d\tau \\
&\leq e^{(t-t_0)\mu_2(A)} \|E_0\|_2 + \max_{\tau \in [t_0,t]} \|b_m(\tau)\|_2 \int_{t_0}^{t} e^{(t-\tau)\mu_2(A)} d\tau \\
&\leq e^{(t-t_0)\mu_2(A)} \|E_0\|_2 + \max_{\tau \in [t_0,t]} \|b_m(\tau)\|_2 e^{(t-t_0)\mu_2(A)} - 1 \mu_2(A) \\
&\leq e^{(t-t_0)\mu_2(A)} \|E_0\|_2 + \max_{\tau \in [t_0,t]} \|vec(R_m(\tau))\|_2 e^{(t-t_0)\mu_2(A)} - 1 \mu_2(A).
\end{align*}
$$

As $\|vec(E_m(t))\|_2 = \|E_m(t)\|_F$, so

$$
\begin{align*}
\|E_m(t)\|_F &\leq e^{(t-t_0)\mu_2(A)} \|E_m(t_0)\|_F + \max_{\tau \in [t_0,t]} \|R_m(\tau)\|_F \frac{e^{(t-t_0)\mu_2(A)} - 1 \mu_2(A)}{\mu_2(A)}.
\end{align*}
$$
Since
\[
\max_{\tau \in [t_0, t]} \|R_m(\tau)\|_F \leq \sqrt{\|T^A_{m+1,m}\|^2_F + \|T^B_{m+1,m}\|^2_F}
\]
\[
\times \max \left\{ \max_{\tau \in [t_0, t]} \|Y_m^{(1)}(\tau)\|_F, \max_{\tau \in [t_0, t]} \|Y_m^{(2)}(\tau)\|_F \right\}.
\]
Then, we have the following upper bound for the norm of the error,
\[
\|E_m(t)\|_F \leq e^{(t-t_0)\mu_2(A)} \|E_m(t_0)\|_F + \alpha_m \rho_m e^{(t-t_0)\mu_2(A)} - \frac{1}{\mu_2(A)},
\]
where
\[
\alpha_m = \max \left\{ \max_{\tau \in [t_0, t]} \|Y_m^{(1)}(\tau)\|_F, \max_{\tau \in [t_0, t]} \|Y_m^{(2)}(\tau)\|_F \right\},
\]
\[
\rho_m = \sqrt{\|T^A_{m+1,m}\|^2_F + \|T^B_{m+1,m}\|^2_F}.
\]

In the next section, we give some iterative methods for solving the reduced order differential Sylvester matrix equation (5).

### 4 Methods for solving the reduced generalized differential Sylvester equation

#### 4.1 Rosenbrock method

In this section, we apply the Rosenbrock method [10, 22] to the low dimensional generalized differential Sylvester matrix equation (5). The new approximation \(Y_{m,j+1}\) of \(Y_m(t_{j+1})\) obtained at step \(j + 1\) is defined, by the relation
\[
Y_{m,j+1} = Y_{m,j} + \frac{3}{2} K_1 + \frac{1}{2} K_2,
\]
where \(K_1\) and \(K_2\) solve the following generalized Sylvester matrix equations
\[
T_{m,A} K_1 + K_1 T_{m,B}^T + \sum_{i=1}^{k} V^T_m N_i V_m K_1 W^T_m M_i^T W_m = g(Y_{m,j}), \tag{13}
\]
and
\[
T_{m,A} K_2 + K_2 T_{m,B}^T + \sum_{i=1}^{k} V^T_m N_i V_m K_2 W^T_m M_i^T W_m = g(Y_{m,j} + K_1) + \frac{2}{h} K_1, \tag{14}
\]
where
\[
\begin{align*}
\mathcal{T}_{m,A} &= \frac{1}{2\gamma} I_{2m} - \gamma \mathcal{T}_{m,A}, \\
\mathcal{T}_{m,B} &= \frac{1}{2\gamma} I_{2m} - \gamma \mathcal{T}_{m,B}, \\
g(Y) &= \mathcal{T}_{m,A}Y + Y^T \mathcal{T}_{m,B} + \sum_{i=1}^{k} V_i^T N_i Y \mathcal{W}_m^T M_i^T \mathcal{W}_m - V_i^T \mathcal{E} \mathcal{F}^T \mathcal{W}_m.
\end{align*}
\]

The equations (13) and (14) are written as the form
\[
A_m X + X B_m^T + \sum_{i=1}^{k} N_{mi} X M_{mi}^T = C_m. \quad \text{(GSME) (15)}
\]

Let \(A_m = Q_A U_A Q_A^T\) and \(B_m = Q_B U_B Q_B^T\) be the real Schur decompositions of the matrices \(A_m\) and \(B_m\), respectively. Then, to solve the small or medium size of generalized Sylvester matrix equation (15) we will apply the following algorithm

**Algorithm 2** The GSME-small method for solving (15)

**Input:** Matrices \(A_m, B_m, N_{m1}, \ldots, N_{mk}, M_{m1}, \ldots, M_{mk}\) and \(C_m\).

1. Choose a tolerance \(tol > 0\).
2. Compute: \(A_m = Q_A U_A Q_A^T\).
3. Compute: \(B_m = Q_B U_B Q_B^T\).
4. Compute: \(N_i = Q_A^T N_{mi} Q_A, M_i = Q_B^T M_{mi} Q_B\) for \(i = 1, \ldots, k\).
5. Compute: \(C = Q_A^T C_m Q_B\).
6. Solve \(U_A Y_0 + Y_0 U_B^T = C\).
7. Set \(Z = Y_0\)
8. For \(j = 0, 1, \ldots\) until convergence do
   - (a) Solve \(U_A Y_{j+1} + Y_{j+1} U_B^T = -\sum_{i=1}^{k} N_i Y_j M_i^T\).
   - (b) Set \(Z = Z + Y_{j+1}\).
   - (c) If \(\|R^{(j+1)}\|_F \leq tol\) then
   - (d) Set \(l = j + 1\)
   - (e) Break
   - (f) End
9. End For \(j\)
10. Return \(X^{(l)} = Q_A Z Q_B^T\).

**Output:** \(X^{(l)}\).

For more details on this approach to solve generalized Sylvester matrix equation (15) see [4, 5, 15, 19]. Now, we summarize the steps of the Rosenbrock in the following algorithm
Algorithm 3 The (Ros-2) method for solving reduced GDSE (5)

**Input:** $T_{m,A}, T_{m,B}, V_m, W_m, E, F, t_0, T_f, N_i$ and $M_i$, for $i = 1, \ldots, k$.

1. Choose $h$.
2. Compute: $r = \frac{T_f - t_0}{h}$
3. Compute: $T_{m,A} = \frac{1}{2h} I_{2ms} - T_{m,A}$
4. Compute: $T_{m,B} = \frac{1}{2h} I_{2ms} - T_{m,B}$
5. Compute: $N_{i,m} = Y_m^T N_i V_m$
6. Compute: $M_{i,m} = W_m^T M_i W_m$
7. For $j = 1 : r$
   (a) Apply GSME–small (Algorithm 2) to (13)
   (b) Apply GSME–small (Algorithm 2) to (14)
   (c) Calculate $Y_{m,j+1}$ by (12)
8. End For $j$.

**Output:** $Y_{m,j+1}$.

We summarize the steps of this approach extended block Arnoldi and Rosenbrock method to solving the generalized differential Sylvester matrix equation (1) in the following algorithm.

Algorithm 4 The extended block Arnoldi–Rosenbrock (EBA-Ros) method for Solving GDSE

**Input:** $X_0, A, B, E$ and $F$ an matrix.

1. Choose a tolerance $tol > 0$ and an integer $m_{max}$.
2. For $m = 1 : m_{max}$
   (a) Apply EBA (Algorithm 1) to $(A, E)$ and $(B, F)$ to get $V_m, W_m$, $T_{m,A}$ and $T_{m,B}$.
   (b) Apply the Ros-2 method (Algorithm 3) to solve the low dimensional generalized differential Sylvester equation (5).
   (c) If $\|R_m\|_F < tol$, stop.
3. End For $m$
4. Compute the approximate solution $X_m$ in the factored form given by the relation (8).

**Output:** $X_m$. 
4.2 BDF method

We use the Backward Differentiation Formula (BDF) method for solving the reduced generalized differential Sylvester matrix equation (5). At each time $t_j$, let $Y_{m,j}$ of the approximation of $Y_m(t_j)$, where $Y_m$ is a solution of (5). Then, the new approximation $Y_{m,j+1}$ of $Y_m(t_{j+1})$ obtained at step $j+1$ by BDF2 is defined by the implicit relation

$$
Y_{m,j+1} = \frac{4}{3}Y_{m,j} - \frac{1}{3}Y_{m,j-1} + \frac{2h}{3}f(Y_{m,j+1}),
$$

(16)

where $h = t_{j+1} - t_j$ is the step size, and $f(Y)$ is given by

$$
f(Y) = T_{m,A}Y + YT_{m,B}^T + \sum_{i=1}^{k} N_{i,m}YM_{i,m}^T - V_mE^TF^TW_m.
$$

The approximate $Y_{m,j+1}$ solves the following matrix equation

$$
-Y_{m,j+1} + \frac{2h}{3}f(Y_{m,j+1}) + \frac{4}{3}Y_{m,j} - \frac{1}{3}Y_{m,j-1} = 0.
$$

(17)

Let

$$
\begin{align*}
\mathcal{T}_{m,A} &= \frac{2h}{3}T_{m,A} - \frac{1}{3}I_{2ms}, \\
\mathcal{T}_{m,B} &= \frac{2h}{3}T_{m,B} - \frac{1}{3}I_{2ms}, \\
Q_{m,j+1} &= -\frac{2h}{3}V_mE^TF^TW_m + \frac{1}{3}Y_{m,j} - \frac{1}{3}Y_{m,j-1}, \\
N_{i,m} &= \sqrt{\frac{2h}{3}}V_mE^TF^TW_m, \\
M_{i,m} &= \sqrt{\frac{2h}{3}}W_mE^TF^TW_m.
\end{align*}
$$

Therefore, we can write equation (17) as the following generalized Sylvester matrix equation:

$$
\mathcal{T}_{m,A}Y_{m,j+1} + Y_{m,j+1}\mathcal{T}_{m,B}^T + \sum_{i=1}^{k} N_{i,m}YM_{i,m}^T + Q_{m,j+1} = 0.
$$

(18)

To solve this equation we will apply the GSME–small algorithm (Algorithm 2). We summarize the steps of the BDF2 method in the following algorithm

**Algorithm 5** The BDF2 method for reduced GDSE (5)

**Input:** $T_{m,A}, T_{m,B}, V_m, W_m, E, F, N_{i,m}, M_{i,m}, t_0, T_f$.

1. Choose $h$.
2. Compute: $r = \frac{T_f - t_0}{h}$.
3. Compute: $\mathcal{T}_{m,A}, \mathcal{T}_{m,B}, N_{i,m}, M_{i,m}$.
4. For $j = 1 : r$
   (a) Compute: $Q_{j+1}$.
   (b) Apply GSME–small (Algorithm 2) for Solving the GSME (18).
5. End For $j$.

**Output:** $Y_{m,T_f}$. 
We summarize the steps of this approach extended block Arnoldi and BDF2 method for solving the generalized differential Sylvester matrix equation (1) in the following algorithm

**Algorithm 6** The extended block Arnoldi–BDF (EBA–BDF2) method for GDSE

**Input**: \( X_0, A, B, E \) and \( F \) an matrix.

1. Choose a tolerance \( tol > 0 \) and an integer \( m_{\text{max}} \).
2. For \( m = 1 : m_{\text{max}} \):
3. Apply EBA (Algorithm 1) to \((A, E)\) and \((B, F)\) to get \( V_m, W_m, T_{m,A} \) and \( T_{m,B} \).
4. Apply BDF2 (Algorithm 5) to find the approximate solution of equation (5).
5. If \( \|R_m\|_F < tol \).
6. End For \( m \)
7. Compute the approximate solution \( X_m \) by using (8).

**Output**: \( X_m \).

## 5 Numerical experiments

In this section, we present some numerical experiments of large and sparse generalized differential Sylvester matrix equations. We give approach to low-rank approximate solutions by extended block Arnoldi algorithm via Rosenbrock method (EBA–Ros) and BDF method (EBA–BDF). The algorithms are coded in MATLAB R2018b. All the experiments were performed on a Laptop with an Intel Core i3 processor and 4GB of RAM. In all of the examples, the matrices \( E \) and \( F \) were generated randomly and their coefficients were uniformly distributed in \([0, 1]\). The time interval considered was \([1, 2]\) and the initial condition \( X_0 = 0 \).

**Example 1.** For the first experiment, we considered the generalized differential Sylvester equation of the form

\[
\dot{X}(t) = AX(t) + X(t)B^T + \gamma^2 NX(t)N^T - EF^T.
\]

In Figure 1, we compared the component \( X_{11} \) of the solution obtained by the EBA–Ros and EBA–BDF methods, to the solution provided by the ode23s method from MATLAB. And on the left, the graph shows the variation of the residual norm with the number of iterations, where \( A = \text{tridiag}(2, -5, 2) \), \( B = \text{tridiag}(1, -4, 1) \), \( N = \text{tridiag}(3, -7, 3) \), \( h = 0.005 \), \( s = 2 \), \( \gamma = \frac{1}{6} \), the
Figure 1: Residual norms vs the number of extended block Arnoldi iterations $m$ (left plot) and values of $X(1,1)(t)$ for $t \in [1, 2]$ computed by ode23s, EBA-Ros and EBA-BDF methods (right plot).

Example 2. In this second example, we considered the particular case of a general differential Lyapunov equation

$$\begin{align*}
\dot{X}(t) &= AX(t) + X(t)A^T + NX(t)N^T - EE^T, \\
X(0) &= 0_{n,n},
\end{align*}$$

(19)

In Figure 2, we plotted the Frobenius residual norm at final time $T_f$ in function of the number $m$ of iterations for the EBA–Ros and EBA–BDF methods, with $A = \text{tridiag}(2, -5, 2)$, $N = \text{tridiag}(\frac{1}{12}, 1, \frac{1}{12})$, the tolerance was set to $10^{-9}$ for the stop test on the residual, we used a constant time step $h = 0.01$. Their rank were set to $s = 2$.

In Table 2, we give the obtained runtimes in seconds, the number of iterations and the Frobenius residual norm at final time, for both methods EBA–BDF and EBA–Ros applied to Eq. (19), with $A = \text{tridiag}(2, -5, 2)$, $N = \text{tridiag}(\frac{1}{12}, 1, \frac{1}{12})$, the tolerance was set to $10^{-9}$ for the stop test on
Table 1: Runtimes and the Frobenius residual norms for Example 1.

<table>
<thead>
<tr>
<th>Test</th>
<th>Methods</th>
<th>CPU time</th>
<th>Iterations</th>
<th>$|R_m(T_f)|_F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A = \text{tridiag}(2, -5, 2), n = 6400$</td>
<td>EBA–BDF</td>
<td>1.98s</td>
<td>12</td>
<td>$4.04 \times 10^{-10}$</td>
</tr>
<tr>
<td>$B = \text{tridiag}(1, -4, 1), p = 6400$</td>
<td>EBA–Ros</td>
<td>0.58s</td>
<td>12</td>
<td>$5.19 \times 10^{-10}$</td>
</tr>
<tr>
<td>$N = \text{tridiag}(3, -7, 3), \gamma = \frac{1}{2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A = \text{tridiag}(2, -5, 2), n = 3456$</td>
<td>EBA–BDF</td>
<td>10.48s</td>
<td>9</td>
<td>$2.85 \times 10^{-10}$</td>
</tr>
<tr>
<td>$B = \text{tridiag}(1, -4, 1), p = 3456$</td>
<td>EBA–Ros</td>
<td>0.26s</td>
<td>9</td>
<td>$2.14 \times 10^{-10}$</td>
</tr>
<tr>
<td>$N = I_n, \gamma = \frac{1}{2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A = \text{tridiag}(2, -5, 2), n = 4960$</td>
<td>EBA–BDF</td>
<td>30.75s</td>
<td>13</td>
<td>$4.75 \times 10^{-10}$</td>
</tr>
<tr>
<td>$B = \text{add32 mtx}, p = 4960$</td>
<td>EBA–Ros</td>
<td>0.50s</td>
<td>13</td>
<td>$8.53 \times 10^{-10}$</td>
</tr>
<tr>
<td>$N = I_n, n = 4960, \gamma = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A = \text{fdm}(\cos(xy), e^{xy}, 10)$</td>
<td>EBA–BDF</td>
<td>23.12s</td>
<td>12</td>
<td>$5.56 \times 10^{-10}$</td>
</tr>
<tr>
<td>$B = \text{fdm}(xy, x^2 + y^2, 1), \gamma = \frac{1}{2}$</td>
<td>EBA–Ros</td>
<td>0.45s</td>
<td>11</td>
<td>$9.63 \times 10^{-10}$</td>
</tr>
<tr>
<td>$N = \text{tridiag}(1, 0, 1), n = p = 3000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Residual norm vs number $m$ of extended block Arnoldi iterations for Example 2.
Extended Krylov method for generalized Sylvester equation

Table 2: Results for Example 2.

<table>
<thead>
<tr>
<th>Test</th>
<th>Methods</th>
<th>CPU time</th>
<th>Iterations</th>
<th>$|R_m(T_f)|_F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 3600$</td>
<td>EBA–BDF</td>
<td>3.66s</td>
<td>13</td>
<td>$1.32 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>2.42s</td>
<td>13</td>
<td>$2.50 \times 10^{-10}$</td>
</tr>
<tr>
<td>$n = 6400$</td>
<td>EBA–BDF</td>
<td>8.12s</td>
<td>13</td>
<td>$4.26 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>2.37s</td>
<td>13</td>
<td>$4.39 \times 10^{-10}$</td>
</tr>
<tr>
<td>$n = 8100$</td>
<td>EBA–BDF</td>
<td>8.65s</td>
<td>13</td>
<td>$6.04 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>4.98s</td>
<td>13</td>
<td>$5.71 \times 10^{-10}$</td>
</tr>
<tr>
<td>$n = 36100$</td>
<td>EBA–BDF</td>
<td>15.50s</td>
<td>14</td>
<td>$1.97 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>4.48s</td>
<td>14</td>
<td>$2.00 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

Table 3: Runtimes and the Frobenius residual norms for Example 3.

<table>
<thead>
<tr>
<th>Test</th>
<th>Methods</th>
<th>CPU time</th>
<th>Iterations</th>
<th>$|R_m(T_f)|_F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = p = 1600$</td>
<td>EBA–BDF</td>
<td>8.88s</td>
<td>13</td>
<td>$9.36 \times 10^{-11}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>5.72s</td>
<td>13</td>
<td>$8.49 \times 10^{-11}$</td>
</tr>
<tr>
<td>$n = 6400, p = 3600$</td>
<td>EBA–BDF</td>
<td>9.16s</td>
<td>12</td>
<td>$5.73 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>5.85s</td>
<td>12</td>
<td>$5.00 \times 10^{-10}$</td>
</tr>
<tr>
<td>$n = 10000, p = 8100$</td>
<td>EBA–BDF</td>
<td>9.90s</td>
<td>13</td>
<td>$2.00 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>3.06s</td>
<td>13</td>
<td>$2.14 \times 10^{-10}$</td>
</tr>
<tr>
<td>$n = 14400, p = 12100$</td>
<td>EBA–BDF</td>
<td>10.61s</td>
<td>13</td>
<td>$3.00 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td>EBA–Ros</td>
<td>3.27s</td>
<td>13</td>
<td>$3.05 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

the residual, we used a constant time step $h = 0.01$. Their rank were set to $s = 2$, for the EBA–Ros and EBA–BDF methods.

Example 3. In this example, we considered the generalized differential Sylvester equation of the form

$$\dot{X}(t) = AX(t) + X(t)B^T + N_1X(t)M_1^T + N_2X(t)M_2^T - EF^T.$$  

In Table 3, we give the obtained runtimes in seconds, the number of iterations and the Frobenius residual norm at final time, for solving equation (1), where $A = \text{tridiag}(2, -5, 2)$, $B = \text{tridiag}(1, -4, 1)$, $N_1 = \frac{1}{5}\text{tridiag}(3, -7, 3)$, $N_2 = \frac{1}{5}\text{tridiag}(1, -2, 1)$, $M_1 = \frac{1}{5}\text{tridiag}(2, 5, 2)$, $M_2 = \frac{1}{5}\text{tridiag}(3, 4, 3)$, $h = 0.01$, $s = 2$ and the tolerance was set to $10^{-9}$ for the stop test on the residual.
6 Conclusion

We presented a new approach for computing approximate solutions to large scale general differential Sylvester matrix equations. The approach is based on projecting the initial problem onto a extended block Krylov subspace to obtain a low dimensional general differential Sylvester equation which is solved by using the well known BDF or Rosenbrock methods. We gave some theoretical results such as the exact expression of the residual norm. Numerical experiments show that the proposed method is effective for large-scale problems.

Acknowledgements

The authors are grateful to the Prof. Davod Khojasteh Salkuyeh, editor in chief and anonymous referee for their valuable comments which helped to improve the quality of this paper.

References


