Journal of Applied Analysis and Computation
Volume 16, Number 2, April 2026, 579–635
Website:http://www.jaac-online.com
DOI:10.11948/20240566

DIAGONAL AND NEW MODIFIED GRADIENT-BASED ITERATIVE ALGORITHMS FOR SYLVESTER TENSOR EQUATIONS*

Fengqing Li[1], Zhengge Huang[1,†], Jingjing Cui[1] and Xiuwen Zheng[1]

Abstract The main objective of this work is to construct some new gradient-based iterative (GI)-like algorithms for solving the (coupled) Sylvester tensor equations. In this paper, we first derive the optimal parameter and the corresponding optimal convergence factor of the GI algorithm (Math. Probl. Eng. 819479 (2013) 1-7) in terms of matricization of a tensor and straightening operator. In order to reduce the computation cost of each iteration of the GI algorithm, enlightened by the idea of the Jacobi method, we replace the system matrices in the GI algorithm by their diagonal parts, and design the diagonal GI (DGI) algorithm for the Sylvester tensor equations. And we derive the sufficient convergence condition, quasi-optimal parameter and quasi-optimal convergence factor of the DGI algorithm. Furthermore, we apply a new update strategy to the GI algorithm and develop the new modified GI (NMGI) algorithm for the Sylvester tensor equations. The proposed NMGI algorithm is different from the MGI one (Math. Probl. Eng. 819479 (2013) 1-7), and can make more full use of the latest computed results and has faster convergence rate than the MGI one for many cases. Also, by utilizing the properties of the tensor norm and techniques of inequalities, we prove that the proposed NMGI algorithm is convergent under proper restrictions. Lastly, some numerical examples are given to validate the efficiencies and advantages of the proposed algorithms for the (coupled) Sylvester tensor equations.

Keywords Sylvester tensor equation, optimal parameter, DGI algorithm, quasi-optimal parameter, NMGI algorithm, convergence properties.

MSC(2010) 65F10, 65H10.

1. Introduction

In this paper, we aim to compute the numerical solution of the following Sylvester tensor equation

$$\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W}, \tag{1.1}$$

where $V_1 \in \mathbb{R}^{L_1 \times L_1}$, $V_2 \in \mathbb{R}^{L_2 \times L_2}$, $V_3 \in \mathbb{R}^{L_3 \times L_3}$, $\mathcal{W} \in \mathbb{R}^{L_1 \times L_2 \times L_3}$ are given, and the unknown tensor $\mathcal{Y} \in \mathbb{R}^{L_1 \times L_2 \times L_3}$ needs to be computed. The operators $\times_j$ $(j = 1, 2, 3)$ in (1.1) are defined as in the next section. When $\mathcal{Y} \in \mathbb{R}^{L_1 \times L_2}$ and $\mathcal{W} \in \mathbb{R}^{L_1 \times L_2}$ are matrices, Sylvester tensor

†The corresponding author.
[1]School of Mathematical Sciences, Center for Applied Mathematics of Guangxi, Guangxi Minzu University, 530006 Nanning, China
*The authors were supported by National Natural Science Foundation of China (No. 12361078) and the Guangxi Natural Science Foundation (No. 2024GXNSFAA010498).
Email: FengqingLi@163.com(F. Li), ZhenggeHuang@163.com(Z. Huang), JingjingCui1990@163.com(J. Cui), XiuwenZheng@163.com(X. Zheng)

equation (1.1) reduces to the Sylvester matrix equation as follows

$$V_1 Y + Y V_2^T = W. \tag{1.2}$$

Sylvester matrix equations are widely used in control theory [8, 9, 11, 28, 44], model reduction [2], image process [5], quantum information [35], disturbance decoupling problem [7] and system identification [10, 27, 36]. Thus it is meaningful to design efficient methods to establish the numerical solutions of the Sylvester matrix equations. These methods mainly can be divided into to two categories: iterative methods and direct methods. The simplest direct method for the matrix equation (1.2) is the Kronecker product method [15], but it requires huge computation and memory space. Another direct method is to transform the coefficient matrix into a special form and then solve another matrix equation that is easy to compute, for instance, Schur canonical form, Hessenberg-Schur form and so forth [16, 17, 26, 31, 34], but these approaches are costly.

Therefore, using iterative methods to solve the Sylvester matrix equation (1.2) may be more efficient than the direct methods. Hence, the research of iterative methods for the matrix equations has attracted considerable attention from many researchers, and a large number of efficient iterative methods have been developed for solving different kinds of Sylvester matrix equations. For example, Ding and Chen [9, 11, 13] proposed some simple and efficient iterative methods for solving the matrix equations, which are easy to implement. Bai [1] designed a Hermitian and skew-Hermitian splitting (HSS) method to solve (1.2). Then the modified HSS method has been proposed in [45] to improve the convergence speed of the HSS one. In addition, Li and Wu [25] extended the single-step HSS (SHSS) method for saddle point problems. Yan and Ma [41] designed an iterative algorithm to solve a class of generalized coupled Sylvester-transpose matrix equations over bisymmetric or skew-anti-symmetric matrices. And Wu and Zeng [39] proposed the ADMM-based methods to solve the nearness symmetric solution of the system of matrix equations $A_1XB_1 = C_1$ and $A_2XB_2 = C_2$. Recently, Wang and Song [38] proposed a new BCR algorithm to compute the constraint solution of the coupled operator equations. Apart from the methods mentioned above, Krylov subspace methods for solving (1.2) have also been developed in [19, 20, 29]. Based on the hierarchical identification principle [10, 12], Ding et al. [9, 14] established the GI algorithms to solve the coupled Sylvester matrix equations and generalized Sylvester matrix equations. In order to improve the computational efficiency of the GI algorithm, its many improved versions have been proposed in recent years. For example, Niu et al. [32] derived the relaxed GI (RGI) method for solving the Sylvester equation. Besides, Xie and Ma [40] combined the update strategy with and the relaxation technique, and constructed the accelerated GI (AGI) algorithm for solving the generalized Sylvester-transpose matrix equations.

As the generalization of the Sylvester matrix equation (1.2), Sylvester tensor equation (1.1) have widely applications in many fields, such as finite element, finite difference, image processing, discretization of high-dimensional linear partial differential equations and convection-diffusion equations, spectral methods and so forth, see [3, 21, 24, 30, 43] and the references therein. There are some illustrations:

- By applying a standard finite difference discretization on equidistant nodes and a second order convergent scheme to the following convection-diffusion equation [43]

$$\begin{aligned} &-x\Phi y + z^T \Psi y = g, \quad \text{in } \Gamma = [0,1]\times[0,1]\times[0,1], \\ &x = 0, \quad \text{on } \partial\Gamma, \end{aligned}$$

we can obtain the Sylvester tensor equation (1.1) with

$$V_i = \begin{pmatrix} 2vh^{-2}+\frac{3}{4}ch^{-1} & -vh^{-2}-\frac{5}{4}ch^{-1} & \frac{1}{4}ch^{-1} & & \\ -vh^{-2}+\frac{1}{4}ch^{-1} & 2vh^{-2}+\frac{3}{4}ch^{-1} & -ch^{-2}-\frac{5}{4}ch^{-1} & \frac{1}{4}ch^{-1} & \\ \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & -ch^{-2}-\frac{5}{4}ch^{-1} & \\ 0 & \cdots & -vh^{-2}+\frac{1}{4}ch^{-1} & 2vh^{-2}+\frac{3}{4}ch^{-1} & \end{pmatrix}_{n\times n}, \quad i=1,2,3.$$

- The three-dimensional microscopic heat transport problem can be discretized and solved via a mixed collocation-finite difference technique. In accordance with this description, the Sylvester tensor equation (1.1) is constructed such that the entries of the coefficient matrices $V_1 = (v_{ij}^{(1)}) \in \mathbb{R}^{L_1\times L_1}$, $V_2 = (v_{ij}^{(2)}) \in \mathbb{R}^{L_2\times L_2}$ and $V_3 = (v_{ij}^{(3)}) \in \mathbb{R}^{L_3\times L_3}$, scaled through a mesh grid $L_1 \times L_2 \times L_3$, are expressed as [30]

$$v_{ij}^{(1)} = \frac{1}{L_1L_2L_3}\begin{cases} -\tau_q\left(\frac{\pi}{\delta_x}\right)^2\left(\frac{M_1^2+2}{3}\right) - \frac{(\Delta t+\tau_q)}{\alpha\Delta t}, & \text{if } i=j,\\ -2\tau_q\left(\frac{\pi}{\delta_x}\right)^2\frac{(-1)^{i+j}}{\sin^2\left[\frac{1}{2}\left(\frac{2\pi\xi_j}{\delta_x}-x_i\right)\right]}, & \text{if } i\neq j,\end{cases}$$

$$v_{ij}^{(2)} = \frac{1}{L_1L_2L_3}\begin{cases} -\tau_q\left(\frac{\pi}{\delta_y}\right)^2\left(\frac{M_2^2+2}{3}\right), & \text{if } i=j,\\ -2\tau_q\left(\frac{\pi}{\delta_y}\right)^2\frac{(-1)^{i+j}}{\sin^2\left[\frac{1}{2}\left(\frac{2\pi\eta_j}{\delta_y}-y_i\right)\right]}, & \text{if } i\neq j,\end{cases}$$

$$v_{ij}^{(3)} = \frac{1}{L_1L_2L_3}\begin{cases} -\tau_U\left(\frac{\pi}{\delta_z}\right)^2\left(\frac{M_3^2+2}{3}\right), & \text{if } i=j,\\ -2\tau_U\left(\frac{\pi}{\delta_z}\right)^2\frac{(-1)^{i+j}}{\sin^2\left[\frac{1}{2}\left(\frac{2\pi\zeta_j}{\delta_z}-z_i\right)\right]}, & \text{if } i\neq j.\end{cases}$$

As mentioned before, the Sylvester tensor equations arise widely in scientific and engineering fields. Thus it is meaningful to design efficient algorithms for solving the Sylvester tensor equations, and a great many of methods for Sylvester tensor equation (1.1) have been presented in recent years. In these methods, GI-like algorithms is a type of effective methods for the Sylvester tensor equations. So far, some GI-like algorithms for the Sylvester matrix equations have been extended to solve the Sylvester tensor equation (1.1). For example, Chen and Lu [6] extended the GI and the modified GI (MGI) algorithms to compute the numerical solutions of (1.1). Then Zhang and Wang [43] applied the relaxation technique to the GI and MGI algorithms, and proposed the relaxed GI (RGI) and modified RGI (MRGI) algorithms for the Sylvester tensor equation (1.1). Numerical results in [43] show that the RGI and MRGI algorithms have better numerical performances than the GI and MGI ones, respectively, by choosing proper relaxation parameters. Nevertheless, it can be found that the optimal parameter of the GI algorithm has not been derived, which plays a key role in the implementation of the GI algorithm. Moreover, when the system matrices V_1, V_2 and V_3 in (1.1) are large and dense, the computations of the mode products in the GI and MGI algorithms may require a large amount of calculation and consume much time. Thus the purpose of this paper is to perfect the theories of the existing GI algorithm in [6], and construct some new GI-like algorithms to reduce the computation of the GI and MGI ones in [6] and ameliorate their numerical performances. We first deduce the optimal parameter and the corresponding optimal convergence factor of the GI algorithm in [6] based on mode-1 matricization and straightening operator of a tensor. Besides, inspired by the ideas of [18, 37], we replace the system matrices in the GI algorithm by their diagonal parts, then develop the diagonal GI (DGI) algorithm for the Sylvester tensor equation (1.1). Compared with the GI algorithm in [6], the proposed DGI algorithm requires less computational cost. To further improve the efficiency of the MGI algorithm proposed in [6] and reduce its computation, we introduce a new update strategy for the GI algorithm and replace the system matrices by their diagonal parts, then establish a new MGI (NMGI) algorithm for (1.1). In addition, we deduce the convergence properties of the proposed DGI and NMGI algorithms, including the quasi-optimal parameter and quasi-optimal convergence factor of the DGI algorithm, and sufficient convergence conditions of the DGI and NMGI algorithms. Also, we extend the DGI and NMGI algorithms to solve the coupled Sylvester tensor equations (with two unknowns). Lastly, several numerical examples are provided to verify that the proposed algorithms are efficient and outperform the GI, RGI, MGI and MRGI ones in terms of numerical performance.

The remainder of this paper is organized as below. Several useful notations, definitions and lemmas are listed in Section 2. In Section 3, we deduce the optimal parameter and the corresponding optimal convergence factor of the GI algorithm in [6]. Then we propose the diagonal GI (DGI) algorithm by replacing the system matrices in the GI algorithm by their diagonal parts, and derive its convergence conditions, quasi-optimal parameter and quasi-optimal convergence factor. In addition, we apply a new update strategy and the diagonal substitution technique to the GI algorithm, then construct a new MGI (NMGI) algorithm and investigate its convergence condition. We extend the DGI and NMGI algorithms to solve the coupled Sylvester tensor equations (with two unknowns) and discuss their convergence properties in Sections 4–5. In Sections 6–8, several numerical examples are given to illustrate the effectivenesses and advantages of the proposed DGI and NMGI algorithms. Lastly, some conclusions and outlooks are given to end this paper in Section 9.

2. Preliminaries

In this section, we list some notations, definitions and lemmas, which will be used in the subsequent sections.

To improve readability, matrices and tensor are written as capital and Euler script letters, respectively. I_k stands for the $k \times k$ identity matrix. Let $H = [h_1, \cdots, h_l] \in \mathbb{R}^{k\times l}$ with h_t $(1 \le t \le l)$ being the t-th column of H, then the vector stretching operation of H is defined as

$$\mathrm{vec}(H) = [h_1^T, h_2^T, \cdots, h_l^T]^T.$$

The set of all order m dimension $L_1 \times L_2 \times \cdots \times L_m$ tensors over $\mathbb{R}$ is denoted by $\mathbb{R}^{L_1\times L_2\times\cdots\times L_m}$. If $\mathcal{U} \in \mathbb{R}^{L_1\times L_2\times\cdots\times L_m}$, then it can be expressed as

$$\mathcal{U} = (u_{i_1 i_2 \cdots i_m}),\ u_{i_1 i_2\cdots i_m} \in \mathbb{R}, 1 \le i_j \le L_j, 1 \le j \le m.$$

In addition, we present several useful definitions below.

Definition 2.1 ([22]). If $\mathcal{Y} \in \mathbb{R}^{L_1\times L_2\times\cdots\times L_m}$ and $G \in \mathbb{R}^{J\times L_n}$, then the t-mode product of $\mathcal{Y}$ and G is defined as

$$(\mathcal{Y}\times_t G)_{p_1\cdots p_{t-1}qp_{t+1}\cdots p_m} = \sum_{p_t=1}^{L_t} y_{p_1\cdots p_t\cdots p_m} g_{qp_t},$$

which is an $L_1\times\cdots\times L_{t-1}\times J\times L_{t+1}\times\cdots\times L_m$ tensor.

Definition 2.2 ([42]). If $\mathcal{Y} = (y_{i_1i_2\cdots i_m}) \in \mathbb{R}^{L_1\times L_2\times\cdots\times L_m}$, then its mode-$n$ matricization $\mathcal{Y}_{(n)}$ is an $L_n \times L_1\cdots L_{n-1}L_{n+1}\cdots L_m$ matrix with

$$(\mathcal{Y}_{(n)})_{i_nl} = y_{i_1i_2\cdots i_m},\ l = 1 + \sum_{q=1,q\neq n}^{m} (i_q-1)J_q,\ J_q = \prod_{p=1,p\neq n}^{q-1} L_p.$$

Definition 2.3 ([43]). Let $\mathcal{Y} \in \mathbb{R}^{L_1\times\cdots\times L_m}$, the operator $\mathrm{vec}(\mathcal{Y})$ is the column stacking form of the corresponding matrix $\mathcal{Y}_{(1)}$. The inner product of $\mathcal{G} = (g_{p_1\cdots p_m}) \in \mathbb{R}^{L_1\times\cdots\times L_m}$ and $\mathcal{H} = (h_{p_1\cdots p_m}) \in \mathbb{R}^{L_1\times\cdots\times L_m}$ is defined as

$$\langle\mathcal{G},\mathcal{H}\rangle = \mathrm{vec}(\mathcal{G})^T\mathrm{vec}(\mathcal{H}) = \sum_{p_1=1}^{L_1}\sum_{p_2=1}^{L_2}\cdots\sum_{p_m=1}^{L_m} g_{p_1p_2\cdots p_m}h_{p_1p_2\cdots p_m}.$$

Then the induced Frobenious norm of $\mathcal{Y}$ is

$$\|\mathcal{Y}\|^2 = \sum_{p_1=1}^{L_1}\sum_{p_2=1}^{L_2}\cdots\sum_{p_m=1}^{L_m} y^2_{p_1p_2\cdots p_m}.$$

Next, some significant lemmas are reviewed in the following.

Lemma 2.1 ([23, 33]). *Let $\mathcal{G},\mathcal{H} \in \mathbb{R}^{L_1\times\cdots L_n\times\cdots\times L_m}$ be two tensors, then the following conclusions hold:*
1) If $\mathcal{H} = \mathcal{G}\times_1 C_1\times_2 C_2\times_3\cdots\times_m C_m$ holds and $C_p \in \mathbb{R}^{L_p\times L_p}$ $(1\le p\le m)$, then

$$\mathcal{H}_{(t)} = C_t\mathcal{G}_{(t)}(C_m\otimes\cdots\otimes C_{t+1}\otimes C_{t-1}\otimes\cdots\otimes C_1)^T.$$

2) For $C_p \in \mathbb{R}^{L_p\times L_p}$ and $C_q \in \mathbb{R}^{L_q\times L_q}$ $(1\le p,q\le m)$, then it has

$$\mathcal{G}\times_p C_p\times_q C_q = \begin{cases}\mathcal{G}\times_p(C_qC_p),\ \text{if}\, p=q,\\ \mathcal{G}\times_q C_q\times_p\times C_p,\ \text{if}\, p\neq q.\end{cases}$$

3) For $C_p \in \mathbb{R}^{L_p\times L_p}$ $(1\le p\le m)$, it holds that $\langle\mathcal{G},\mathcal{H}\times_p C_p\rangle = \langle\mathcal{G}\times_p C_p^T,\mathcal{H}\rangle$.
4) $2\langle\mathcal{G},\mathcal{H}\rangle \le \|\mathcal{G}\|^2 + \|\mathcal{H}\|^2$, $\langle\mathcal{G},\mathcal{H}\rangle \le \|\mathcal{G}\|\|\mathcal{H}\|$.
5) Let $C_p \in \mathbb{R}^{L_p\times L_p}$, then $\|\mathcal{G}\times_p C_p\| \le \|\mathcal{G}\|\|C_p\|_2$ $(1\le p\le m)$, where $\|C_p\|_2$ denotes the spectral norm of the matrix C_p.

By making use of Lemma 2.1, Sylvester tensor equation (1.1) can be transformed into the following linear equation

$$(I_{L_3}\otimes I_{L_2}\otimes V_1 + I_{L_3}\otimes V_2\otimes I_{L_1} + V_3\otimes I_{L_2}\otimes I_{L_1})\mathrm{vec}(\mathcal{Y}) = \mathrm{vec}(\mathcal{W}). \tag{2.1}$$

Lemma 2.2. *Let* $\mathcal{G} = (g_{xyz}), \mathcal{H} = (h_{xyz}) \in \mathbb{C}^{L_1 \times L_2 \times L_3}$, *then* $\|\mathcal{G} + \mathcal{H}\| \le \|\mathcal{G}\| + \|\mathcal{H}\|$.

Proof. By straightforward calculations, it has

$$\begin{aligned}
\|\mathcal{G} + \mathcal{H}\|^2 &= \sum_{xyk} (g_{ijk} + h_{ijk})^2 \\
&= \sum_{xyk} (g_{ijk}^2 + 2g_{ijk}h_{ijk} + h_{ijk}^2) \\
&= \sum_{xyk} g_{ijk}^2 + 2\sum_{xyk} g_{ijk}h_{ijk} + \sum_{xyk} h_{ijk}^2 \\
&= \|\mathcal{G}\|^2 + 2\langle \mathcal{G}, \mathcal{H}\rangle + \|\mathcal{H}\|^2.
\end{aligned} \tag{2.2}$$

By Lemma 2.1, we get $\langle \mathcal{G}, \mathcal{H}\rangle \le \|\mathcal{G}\|\|\mathcal{H}\|$. Thus $2\langle \mathcal{G}, \mathcal{H}\rangle \le 2\|\mathcal{G}\|\|\mathcal{H}\|$, which together with (2.2) yields that $\|\mathcal{G} + \mathcal{H}\|^2 \le \|\mathcal{G}\|^2 + 2\|\mathcal{G}\|\|\mathcal{H}\| + \|\mathcal{H}\|^2 = (\|\mathcal{G}\| + \|\mathcal{H}\|)^2$, and therefore $\|\mathcal{G} + \mathcal{H}\| \le \|\mathcal{G}\| + \|\mathcal{H}\|$. □

3. The DGI and NMGI algorithms and their convergence properties

In this section, we first deduce the optimal parameter of the GI algorithm in [6], then construct two new GI-like algorithms referred to as the DGI and NMGI algorithms for the Sylvester tensor equation (1.1). At last, we deduce the convergence properties of the proposed algorithms. Before that, we first review the GI algorithm in [6]. Based on the hierarchical identification principle, Chen and Lu [6] designed the GI algorithm for solving the Sylvester tensor equation (1.1). The framework of the GI algorithm is as below.

Algorithm 3.1. The gradient-based iterative (GI) algorithm (see [6]):

Step 1. Given matrices $V_1 \in \mathbb{R}^{L_1 \times L_1}$, $V_2 \in \mathbb{R}^{L_2 \times L_2}$, $V_3 \in \mathbb{R}^{L_3 \times L_3}$, a tensor $\mathcal{W} \in \mathbb{R}^{L_1 \times L_2 \times L_3}$, and two constants $\gamma, \eta > 0$. Choose the initial tensor $\mathcal{Y}^0$, and set $l = 0$.

Step 2. If $\tau_l = \frac{\|\mathcal{W} - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3\|}{\|\mathcal{W} - \mathcal{Y}^0 \times_1 V_1 - \mathcal{Y}^0 \times_2 V_2 - \mathcal{Y}^0 \times_3 V_3\|} = \frac{\|\mathcal{R}^k\|}{\|\mathcal{R}^0\|} < \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ by the following procedure

$$\begin{aligned}
\mathcal{Y}_1^{l+1} &= \mathcal{Y}^l + \gamma \mathcal{R}^l \times_1 V_1^T, \\
\mathcal{Y}_2^{l+1} &= \mathcal{Y}^l + \gamma \mathcal{R}^l \times_2 V_2^T, \\
\mathcal{Y}_3^{l+1} &= \mathcal{Y}^l + \gamma \mathcal{R}^l \times_3 V_3^T, \\
\mathcal{Y}^{l+1} &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1}}{3}.
\end{aligned}$$

Step 4. Set $l := l + 1$ and return to Step 2.

Theorem 3.1. *Let* $M = I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}$, *and* $\lambda_{\max}$ *and* $\lambda_{\min}$ *be the maximum and minimum eigenvalues of the matrix* $M^T M$, *respectively. Suppose that the Sylvester tensor equation* (1.1) *has a unique solution* $\mathcal{Y}^*$. *Then the GI algorithm is convergent if and only if*

$$0 < \gamma < \frac{6}{\lambda_{\max}}.$$

And the optimal parameter γ_{opt} and the corresponding optimal convergence factor ρ_{opt} of the GI algorithm are

$$\gamma_{opt} = \frac{6}{\lambda_{\max} + \lambda_{\min}}, \quad \rho_{opt} = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

Proof. Define the error tensors

$$\widetilde{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*, \quad \widetilde{\mathcal{Y}}_j^l = \mathcal{Y}_j^l - \mathcal{Y}^* \ (j = 1, 2, 3).$$

In view of the expression of $\widetilde{\mathcal{Y}}_1^{l+1}$, we derive

$$\widetilde{\mathcal{Y}}_1^{l+1} = \widetilde{\mathcal{Y}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_1 V_1^T.$$

Let $\mathcal{Y} = (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_1 V_1^T$, then from Lemma 2.1 we can deduce that

$$\begin{aligned}
\mathcal{Y} &= (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_1 V_1^T \\
&= \widetilde{\mathcal{Y}}^l \times_1 (V_1^T V_1) + \widetilde{\mathcal{Y}}^l \times_1 V_1^T \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_1 V_1^T \times_3 V_3 \\
&= \widetilde{\mathcal{Y}}^l \times_1 (V_1^T V_1) \times_2 I_{L_2} \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 V_1^T \times_2 V_2 \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 V_1^T \times_2 I_{L_2} \times_3 V_3.
\end{aligned}$$

By taking the mode-1 matricization of $\mathcal{Y}$, we have

$$\mathcal{Y}_{(1)} = V_1^T V_1 \widetilde{\mathcal{Y}}_{(1)}^l (I_{L_3} \otimes I_{L_2})^T + V_1^T \widetilde{\mathcal{Y}}_{(1)}^l (I_{L_3} \otimes V_2)^T + V_1^T \widetilde{\mathcal{Y}}_{(1)}^l (V_3 \otimes I_{L_2})^T. \tag{3.1}$$

Applying the stretching operator on both sides of (3.1) results in

$$\text{vec}(\mathcal{Y}) = [I_{L_3} \otimes I_{L_2} \otimes (V_1^T V_1) + I_{L_3} \otimes V_2 \otimes V_1^T + V_3 \otimes I_{L_2} \otimes V_1^T]\text{vec}(\widetilde{\mathcal{Y}}^l).$$

Therefore we can obtain the following results

$$\begin{aligned}
\text{vec}(\widetilde{\mathcal{Y}}_1^{l+1}) &= [I_{L_1L_2L_3} - \gamma(I_{L_3} \otimes I_{L_2} \otimes (V_1^T V_1) + I_{L_3} \otimes V_2 \otimes V_1^T + V_3 \otimes I_{L_2} \otimes V_1^T)]\text{vec}(\widetilde{\mathcal{Y}}^l) \\
&= M_1 \text{vec}(\widetilde{\mathcal{Y}}^l) \\
&= M_1 \text{vec}(\mathcal{Y}^l - \mathcal{Y}^*) \\
&= M_1 \text{vec}[\frac{1}{3}(\mathcal{Y}_1^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_2^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_3^l - \mathcal{Y}^*)] \\
&= M_1 \text{vec}[\frac{1}{3}\widetilde{\mathcal{Y}}_1^l + \frac{1}{3}\widetilde{\mathcal{Y}}_2^l + \frac{1}{3}\widetilde{\mathcal{Y}}_3^l] \\
&= \frac{1}{3} M_1 [\text{vec}(\widetilde{\mathcal{Y}}_1^l) + \text{vec}(\widetilde{\mathcal{Y}}_2^l) + \text{vec}(\widetilde{\mathcal{Y}}_3^l)],
\end{aligned} \tag{3.2}$$

where $M_1 = I_{L_1L_2L_3} - \gamma(I_{L_3} \otimes I_{L_2} \otimes (V_1^T V_1) + I_{L_3} \otimes V_2 \otimes V_1^T + V_3 \otimes I_{L_2} \otimes V_1^T)$. It follows from Line 2 of Step 3 of the GI algorithm that

$$\widetilde{\mathcal{Y}}_2^{l+1} = \widetilde{\mathcal{Y}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_2 V_2^T.$$

Let $\widetilde{\mathcal{Y}} = (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_2 V_2^T$, then it holds that

$$\begin{aligned}
\widetilde{\mathcal{Y}} &= (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_2 V_2^T \\
&= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 V_2^T + \widetilde{\mathcal{Y}}^l \times_2 (V_2^T V_2) + \widetilde{\mathcal{Y}}^l \times_2 V_2^T \times_3 V_3
\end{aligned}$$

$$= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 V_2^T \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 (V_2^T V_2) \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 V_2^T \times_3 V_3.$$

Taking the mode-1 matricization of $\widetilde{\mathcal{Y}}$ leads to

$$\widetilde{\mathcal{Y}}_{(1)} = V_1 \widetilde{\mathcal{Y}}^l_{(1)} (I_{L_3} \otimes V_2^T)^T + I_{L_1} \widetilde{\mathcal{Y}}^l_{(1)} (I_{L_3} \otimes V_2^T V_2)^T + I_{L_1} \widetilde{\mathcal{Y}}^l_{(1)} (V_3 \otimes V_2^T)^T. \tag{3.3}$$

By utilizing the stretching operator to both sides of (3.3), we obtain

$$\mathrm{vec}(\widetilde{\mathcal{Y}}) = [I_{L_3} \otimes V_2^T \otimes V_1 + I_{L_3} \otimes (V_2^T V_2) \otimes I_{L_1} + V_3 \otimes V_2^T \otimes I_{L_1}] \mathrm{vec}(\widetilde{\mathcal{Y}}^l),$$

from which one can deduce that

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}_2^{l+1}) &= [I_{L_1 L_2 L_3} - \gamma (I_{L_3} \otimes V_2^T \otimes V_1 + I_{L_3} \otimes (V_2^T V_2) \otimes I_{L_1} + V_3 \otimes V_2^T \otimes I_{L_1})] \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= M_2 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= M_2 \mathrm{vec}(\mathcal{Y}^l - \mathcal{Y}^*) \\
&= M_2 \mathrm{vec}[\frac{1}{3}(\mathcal{Y}_1^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_2^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_3^l - \mathcal{Y}^*)] \\
&= M_2 \mathrm{vec}[\frac{1}{3}\widetilde{\mathcal{Y}}_1^l + \frac{1}{3}\widetilde{\mathcal{Y}}_2^l + \frac{1}{3}\widetilde{\mathcal{Y}}_3^l] \\
&= \frac{1}{3} M_2 [\mathrm{vec}(\widetilde{\mathcal{Y}}_1^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^l)],
\end{aligned} \tag{3.4}$$

where $M_2 = I_{L_1 L_2 L_3} - \gamma (I_{L_3} \otimes V_2^T \otimes V_1 + I_{L_3} \otimes (V_2^T V_2) \otimes I_{L_1} + V_3 \otimes V_2^T \otimes I_{L_1})$. From Line 3 of Step 3 of the GI algorithm, it is obtained that

$$\widetilde{\mathcal{Y}}_3^{l+1} = \widetilde{\mathcal{Y}}^l - \gamma (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_3 V_3^T.$$

Let $\widetilde{\widetilde{\mathcal{Y}}} = (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_3 V_3^T$, then straightforward computations give

$$\begin{aligned}
\widetilde{\widetilde{\mathcal{Y}}} &= (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_3 V_3^T \\
&= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_3 V_3^T + \widetilde{\mathcal{Y}}^l \times_2 V_2 \times_3 V_3^T + \widetilde{\mathcal{Y}}^l \times_3 (V_3^T V_3) \\
&= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 I_{L_2} \times_3 V_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 V_2 \times_3 V_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 I_{L_2} \times_3 (V_3^T V_3).
\end{aligned}$$

By taking the mode-1 matricization of $\widetilde{\widetilde{\mathcal{Y}}}$, it has

$$\widetilde{\widetilde{\mathcal{Y}}}_{(1)} = V_1 \widetilde{\mathcal{Y}}^l_{(1)} (V_3^T \otimes I_{L_2})^T + I_{L_1} \widetilde{\mathcal{Y}}^l_{(1)} (V_3^T \otimes V_2)^T + I_{L_1} \widetilde{\mathcal{Y}}^l_{(1)} (V_3^T V_3 \otimes I_{L_2})^T. \tag{3.5}$$

Then using vector stretching operator to both sides of relation (3.5) and applying Lemma 2.1 yield that

$$\mathrm{vec}(\widetilde{\widetilde{\mathcal{Y}}}) = [V_3^T \otimes I_{L_2} \otimes V_1 + V_3^T \otimes V_2 \otimes I_{L_1} + (V_3^T V_3) \otimes I_{L_2} \otimes I_{L_1}] \mathrm{vec}(\widetilde{\mathcal{Y}}^l).$$

As a consequence, the following result is valid

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}_3^{l+1}) &= [I_{L_1 L_2 L_3} - \gamma (V_3^T \otimes I_{L_2} \otimes V_1 + V_3^T \otimes V_2 \otimes I_{L_1} + (V_3^T V_3) \otimes I_{L_2} \otimes I_{L_1})] \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= M_3 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= M_3 \mathrm{vec}(\mathcal{Y}^l - \mathcal{Y}^*)
\end{aligned}$$

$$
\begin{aligned}
&= M_3\text{vec}[\frac{1}{3}(\mathcal{Y}_1^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_2^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_3^l - \mathcal{Y}^*)] \\
&= M_3\text{vec}[\frac{1}{3}\widetilde{\mathcal{Y}}_1^l + \frac{1}{3}\widetilde{\mathcal{Y}}_2^l + \frac{1}{3}\widetilde{\mathcal{Y}}_3^l] \\
&= \frac{1}{3}M_3[\text{vec}(\widetilde{\mathcal{Y}}_1^l) + \text{vec}(\widetilde{\mathcal{Y}}_2^l) + \text{vec}(\widetilde{\mathcal{Y}}_3^l)],
\end{aligned}
$$

where $M_3 = I_{L_1L_2L_3} - \gamma(V_3^T \otimes I_{L_2} \otimes V_1 + V_3^T \otimes V_2 \otimes I_{L_1} + (V_3^T V_3) \otimes I_{L_2} \otimes I_{L_1})$. Since $\text{vec}(\widetilde{\mathcal{Y}}^{l+1}) = \frac{1}{3}[\text{vec}(\widetilde{\mathcal{Y}}_1^{l+1}) + \text{vec}(\widetilde{\mathcal{Y}}_2^{l+1}) + \text{vec}(\widetilde{\mathcal{Y}}_3^{l+1})]$ and $\text{vec}(\widetilde{\mathcal{Y}}^l) = \frac{1}{3}[\text{vec}(\widetilde{\mathcal{Y}}_1^l) + \text{vec}(\widetilde{\mathcal{Y}}_2^l) + \text{vec}(\widetilde{\mathcal{Y}}_3^l)]$, it has

$$
\begin{aligned}
\text{vec}(\widetilde{\mathcal{Y}}^{l+1}) &= \frac{1}{3}[\text{vec}(\widetilde{\mathcal{Y}}_1^{l+1}) + \text{vec}(\widetilde{\mathcal{Y}}_2^{l+1}) + \text{vec}(\widetilde{\mathcal{Y}}_3^{l+1})] \\
&= \frac{1}{3}(\frac{1}{3}M_1 + \frac{1}{3}M_2 + \frac{1}{3}M_3)(\text{vec}(\widetilde{\mathcal{Y}}_1^l) + \text{vec}(\widetilde{\mathcal{Y}}_2^l) + \text{vec}(\widetilde{\mathcal{Y}}_3^l)) \\
&= (\frac{1}{3}M_1 + \frac{1}{3}M_2 + \frac{1}{3}M_3)\text{vec}(\widetilde{\mathcal{Y}}^l).
\end{aligned} \tag{3.6}
$$

Let $P = I_{L_3} \otimes I_{L_2} \otimes (V_1^T V_1) + I_{L_3} \otimes V_2 \otimes V_1^T + V_3 \otimes I_{L_2} \otimes V_1^T + I_{L_3} \otimes V_2^T \otimes V_1 + I_{L_3} \otimes (V_2^T V_2) \otimes I_{L_1} + V_3 \otimes V_2^T \otimes I_{L_1} + V_3^T \otimes I_{L_2} \otimes V_1 + V_3^T \otimes V_2 \otimes I_{L_1} + (V_3^T V_3) \otimes I_{L_2} \otimes I_{L_1} = (I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1})^T (I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}) = M^T M$ with $M = I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}$. Then it follows from (3.6) that $\text{vec}(\widetilde{\mathcal{Y}}^{l+1}) = (\frac{1}{3}M_1 + \frac{1}{3}M_2 + \frac{1}{3}M_3)\text{vec}(\widetilde{\mathcal{Y}}^l) = \frac{1}{3}(3I_{L_1L_2L_3} - \gamma P)\text{vec}(\widetilde{\mathcal{Y}}^l) = (I_{L_1L_2L_3} - \frac{\gamma}{3}P)\text{vec}(\widetilde{\mathcal{Y}}^l)$ and $I_{L_1L_2L_3} - \frac{\gamma}{3}P$ is the iteration matrix of the GI algorithm. By assumptions, (2.1) has a unique solution, thus M is of full column rank and hence $P = M^T M$ is a symmetric positive-definite matrix. Let $\lambda_i > 0$ $(i = 1, 2, \cdots, L_1L_2L_3)$ be the eigenvalues of P, then $\rho(I_{L_1L_2L_3} - \frac{\gamma}{3}P) = \max\limits_{1 \le i \le L_1L_2L_3}\{|1 - \frac{\gamma}{3}\lambda_i|\} = \max\{|1 - \frac{\gamma}{3}\lambda_{\min}|, |1 - \frac{\gamma}{3}\lambda_{\max}|\}$. By some calculations, it holds that

$$
\rho(I_{L_1L_2L_3} - \frac{\gamma}{3}P) = \begin{cases} 1 - \dfrac{\gamma}{3}\lambda_{\min}, \text{ if } \gamma \le \dfrac{6}{\lambda_{\max} + \lambda_{\min}}, \\ \dfrac{\gamma}{3}\lambda_{\max} - 1, \text{ if } \gamma > \dfrac{6}{\lambda_{\max} + \lambda_{\min}}. \end{cases}
$$

When $\gamma \le \frac{6}{\lambda_{\max} + \lambda_{\min}}$, $\rho(I_{L_1L_2L_3} - \frac{\gamma}{3}P) = 1 - \frac{\gamma}{3}\lambda_{\min}$ is monotonic decreasing about γ. And when $\gamma > \frac{6}{\lambda_{\max} + \lambda_{\min}}$, $\rho(I_{L_1L_2L_3} - \frac{\gamma}{3}P) = \frac{\gamma}{3}\lambda_{\max} - 1$ is monotonic increasing about the variable γ, then the optimal parameter γ_{opt} is

$$
\gamma_{\text{opt}} = \frac{6}{\lambda_{\max} + \lambda_{\min}}.
$$

Taking γ_{opt} into $\rho(I - \frac{\gamma}{3}P) = 1 - \frac{\gamma}{3}\lambda_{\min}$ leads to the optimal convergence factor $\rho_{\text{opt}} = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}$. □

According to the framework of the GI algorithm, it can be observed that the Step 3 of this algorithm may consume much time when the matrices V_i $(i = 1, 2, 3)$ are large and dense. In order to reduce the computation of the GI algorithm, motivated by the ideas of [18, 37], we replace the system matrices V_i^T $(i = 1, 2, 3)$ by their diagonal parts in the GI algorithm. This can reduce the computing time of each iteration of the GI algorithm, and the total calculation time decreases. In view of it, we will design a new algorithm called the diagonal GI (DGI) algorithm for the Sylvester tensor equation (1.1). The derivation process of the DGI algorithm is as below.

We first split the system matrices V_j into $V_j = D_j + C_j$ $(j = 1, 2, 3)$ with C_j and D_j being the non-diagonal and diagonal parts of V_j, respectively. Then from (1.1), we can deduce that

$$\begin{aligned}
\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W} &\Rightarrow \mathcal{Y} \times_1 (D_1 + C_1) + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W}, \\
\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W} &\Rightarrow \mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 (D_2 + C_2) + \mathcal{Y} \times_3 V_3 = \mathcal{W}, \\
\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W} &\Rightarrow \mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 (D_3 + C_3) = \mathcal{W},
\end{aligned}$$

which results in

$$\begin{aligned}
\mathcal{Y} \times_1 D_1 &= \mathcal{W} - \mathcal{Y} \times_1 C_1 - \mathcal{Y} \times_2 V_2 - \mathcal{Y} \times_3 V_3, \\
\mathcal{Y} \times_2 D_2 &= \mathcal{W} - \mathcal{Y} \times_1 V_1 - \mathcal{Y} \times_2 C_2 - \mathcal{Y} \times_3 V_3, \\
\mathcal{Y} \times_3 D_3 &= \mathcal{W} - \mathcal{Y} \times_1 V_1 - \mathcal{Y} \times_2 V_2 - \mathcal{Y} \times_3 C_3.
\end{aligned} \tag{3.7}$$

Based on the GI algorithm and (3.7), we establish the following diagonal GI (DGI) algorithm.

Algorithm 3.2. The diagonal gradient-based iterative (DGI) algorithm:

Step 1. Given matrices $V_1 \in \mathbb{R}^{L_1 \times L_1}$, $V_2 \in \mathbb{R}^{L_2 \times L_2}$, $V_3 \in \mathbb{R}^{L_3 \times L_3}$, a tensor $\mathcal{W} \in \mathbb{R}^{L_1 \times L_2 \times L_3}$, and two constants $\gamma, \eta > 0$. Choose the initial tensor $\mathcal{Y}^0$, and set $l = 0$.

Step 2. If $\tau_l = \frac{\|\mathcal{W} - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3\|}{\|\mathcal{W} - \mathcal{Y}^0 \times_1 V_1 - \mathcal{Y}^0 \times_2 V_2 - \mathcal{Y}^0 \times_3 V_3\|} = \frac{\|\mathcal{R}^l\|}{\|\mathcal{R}^0\|} < \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ by the following procedure

$$\begin{aligned}
\mathcal{Y}_1^{l+1} &= \mathcal{Y}^l + \gamma \mathcal{R}^l \times_1 D_1^T, \\
\mathcal{Y}_2^{l+1} &= \mathcal{Y}^l + \gamma \mathcal{R}^l \times_2 D_2^T, \\
\mathcal{Y}_3^{l+1} &= \mathcal{Y}^l + \gamma \mathcal{R}^l \times_3 D_3^T, \\
\mathcal{Y}^{l+1} &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1}}{3}.
\end{aligned}$$

Step 4. Set $l := l + 1$ and return to Step 2.

In what follows, by using the techniques applied in Theorems 4.1 and 4.2 in [18], we deduce the sufficient convergence condition, quasi-optimal parameter and the corresponding quasi-optimal convergence factor of the DGI algorithm in the following two theorems.

Theorem 3.2. *Suppose that the assumptions in Theorem 3.1 are valid, and $Re(\lambda_p) > 0$ with λ_p $(p = 1, \cdots, L_1L_2L_3)$ being the eigenvalues of $G_1^T G_2$, where*

$$\begin{aligned}
G_1 &= I_{L_3} \otimes I_{L_2} \otimes D_1 + I_{L_3} \otimes D_2 \otimes I_{L_1} + D_3 \otimes I_{L_2} \otimes I_{L_1}, \\
G_2 &= I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}.
\end{aligned}$$

Let $Im_1 = \max_{1 \le p \le L_1L_2L_3} |Im(\lambda_p)|$, $Re_{\max} = \max_{1 \le p \le L_1L_2L_3} \{Re(\lambda_p)\}$ and $Re_{\min} = \min_{1 \le p \le L_1L_2L_3} \{Re(\lambda_p)\}$. Then the DGI algorithm is convergent provided that

$$0 < \gamma < \begin{cases} \dfrac{6Re_{\min}}{Re_{\min}^2 + Im_1^2}, & \text{as } Im_1 \ge \sqrt{Re_{\max}Re_{\min}}\,, \\[2ex] \dfrac{6Re_{\max}}{Re_{\max}^2 + Im_1^2}, & \text{as } Im_1 \le \sqrt{Re_{\max}Re_{\min}}\,. \end{cases}$$

Proof. Define the error tensors

$$\widetilde{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*, \quad \widetilde{\mathcal{Y}}_i^l = \mathcal{Y}_j^l - \mathcal{Y}^* \ (j = 1, 2, 3).$$

It follows from the expression of $\mathcal{Y}_1^{l+1}$ in the DGI algorithm that

$$\widetilde{\mathcal{Y}}_1^{l+1} = \widetilde{\mathcal{Y}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 A_1 + \widetilde{\mathcal{Y}}^l \times_2 A_2 + \widetilde{\mathcal{Y}}^l \times_3 A_3) \times_1 D_1^T. \tag{3.8}$$

Let $\mathcal{Z} = (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_1 D_1^T$. By making use of Lemma 2.1, it has

$$\begin{aligned}
\mathcal{Z} &= (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_1 D_1^T \\
&= \widetilde{\mathcal{Y}}^l \times_1 (D_1^T V_1) + \widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_3 V_3 \\
&= \widetilde{\mathcal{Y}}^l \times_1 (D_1^T V_1) \times_2 I_{L_2} \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_2 V_2 \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_2 I_{L_2} \times_3 V_3.
\end{aligned}$$

By taking the mode-1 matricization of $\mathcal{Z}$, we have

$$\mathcal{Z}_{(1)} = D_1^T V_1 \widetilde{\mathcal{Y}}_{(1)}^l (I_{L_3} \otimes I_{L_2})^T + D_1^T \widetilde{\mathcal{Y}}_{(1)}^l (I_{L_3} \otimes V_2)^T + D_1^T \widetilde{\mathcal{Y}}_{(1)}^l (V_3 \otimes I_{L_2})^T. \tag{3.9}$$

Then taking the stretching operator into both sides of (3.9) yields that

$$\mathrm{vec}(\mathcal{Z}) = [I_{L_3} \otimes I_{L_2} \otimes (D_1^T V_1) + I_{L_3} \otimes V_2 \otimes D_1^T + V_3 \otimes I_{L_2} \otimes D_1^T]\mathrm{vec}(\widetilde{\mathcal{Y}}^l).$$

Let $\widetilde{M_1} = I_{L_1 L_2 L_3} - \gamma(I_{L_3} \otimes I_{L_2} \otimes (D_1^T V_1) + I_{L_3} \otimes V_2 \otimes D_1^T + V_3 \otimes I_{L_2} \otimes D_1^T)$. Then it follows from (3.8) that

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}_1^{l+1}) &= [I_{L_1 L_2 L_3} - \gamma(I_{L_3} \otimes I_{L_2} \otimes (D_1^T V_1) + I_{L_3} \otimes V_2 \otimes D_1^T + V_3 \otimes I_{L_2} \otimes D_1^T)]\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= \widetilde{M_1}\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= \widetilde{M_1}\mathrm{vec}(\mathcal{Y}^l - \mathcal{Y}^*) \\
&= \widetilde{M_1}\mathrm{vec}[\frac{1}{3}(\mathcal{Y}_1^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_2^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_3^l - \mathcal{Y}^*)] \\
&= \widetilde{M_1}\mathrm{vec}[\frac{1}{3}\widetilde{\mathcal{Y}}_1^l + \frac{1}{3}\widetilde{\mathcal{Y}}_2^l + \frac{1}{3}\widetilde{\mathcal{Y}}_3^l] \\
&= \frac{1}{3}\widetilde{M_1}[\mathrm{vec}(\widetilde{\mathcal{Y}}_1^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^l)].
\end{aligned} \tag{3.10}$$

In view of the expression of $\mathcal{Y}_2^{l+1}$, it has

$$\widetilde{\mathcal{Y}}_2^{l+1} = \widetilde{\mathcal{Y}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_2 D_2^T. \tag{3.11}$$

Let $\widetilde{\mathcal{Z}} = (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_2 D_2^T$, then we can deduce that

$$\begin{aligned}
\widetilde{\mathcal{Z}} &= (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_2 D_2^T \\
&= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 D_2^T + \widetilde{\mathcal{Y}}^l \times_2 (D_2^T V_2) + \widetilde{\mathcal{Y}}^l \times_2 D_2^T \times_3 V_3 \\
&= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 D_2^T \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 (D_2^T V_2) \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 D_2^T \times_3 V_3,
\end{aligned}$$

in terms of Lemma 2.1. By taking the mode-1 matricization of $\widetilde{\mathcal{Z}}$, it holds that

$$\widetilde{\mathcal{Z}}_{(1)} = V_1 \widetilde{\mathcal{Y}}_{(1)}^l (I_{L_3} \otimes D_2^T)^T + I_{L_1} \widetilde{\mathcal{Y}}_{(1)}^l (I_{L_3} \otimes D_2^T V_2)^T + I_{L_1} \widetilde{\mathcal{Y}}_{(1)}^l (V_3 \otimes D_2^T)^T. \tag{3.12}$$

Using stretching operator to both sides of (3.12) leads to

$$\mathrm{vec}(\widetilde{\mathcal{Z}}) = [I_{L_3} \otimes D_2^T \otimes V_1 + I_{L_3} \otimes (D_2^T V_2) \otimes I_{L_1} + V_3 \otimes D_2^T \otimes I_{L_1}]\mathrm{vec}(\widetilde{\mathcal{Y}}^l). \tag{3.13}$$

Let $\widetilde{M_2} = I_{L_1L_2L_3} - \gamma(I_{L_3} \otimes D_2^T \otimes V_1 + I_{L_3} \otimes (D_2^T V_2) \otimes I_{L_1} + V_3 \otimes D_2^T \otimes I_{L_1})$. Then the combination of (3.11) and (3.13) results in

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}_2^{l+1}) &= [I_{L_1L_2L_3} - \gamma(I_{L_3} \otimes D_2^T \otimes V_1 + I_{L_3} \otimes (D_2^T V_2) \otimes I_{L_1} + V_3 \otimes D_2^T \otimes I_{L_1})]\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= \widetilde{M_2}\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= \widetilde{M_2}\mathrm{vec}(\mathcal{Y}^l - \mathcal{Y}^*) \\
&= \widetilde{M_2}\mathrm{vec}[\frac{1}{3}(\mathcal{Y}_1^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_2^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_3^l - \mathcal{Y}^*)] \\
&= \widetilde{M_2}\mathrm{vec}[\frac{1}{3}\widetilde{\mathcal{Y}}_1^l + \frac{1}{3}\widetilde{\mathcal{Y}}_2^l + \frac{1}{3}\widetilde{\mathcal{Y}}_3^l] \\
&= \frac{1}{3}\widetilde{M_2}[\mathrm{vec}(\widetilde{\mathcal{Y}}_1^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^l)].
\end{aligned} \tag{3.14}$$

Besides, from Line 3 of Step 3 of the DGI algorithm, we have

$$\widetilde{\mathcal{Y}}_3^{l+1} = \widetilde{\mathcal{Y}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_3 D_3^T.$$

Let $\widetilde{\widetilde{\mathcal{Z}}} = (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_3 D_3^T$, then it can be derived that

$$\begin{aligned}
\widetilde{\widetilde{\mathcal{Z}}} &= (\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_3 D_3^T \\
&= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_3 D_3^T + \widetilde{\mathcal{Y}}^l \times_2 V_2 \times_3 D_3^T + \widetilde{\mathcal{Y}}^l \times_3 (D_3^T V_3) \\
&= \widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 I_{L_2} \times_3 D_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 V_2 \times_3 D_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 I_{L_2} \times_3 (D_3^T V_3).
\end{aligned}$$

By taking the mode-1 matricization of $\widetilde{\widetilde{\mathcal{Z}}}$, it follows that

$$\widetilde{\widetilde{\mathcal{Z}}}_{(1)} = V_1\widetilde{\mathcal{Y}}_{(1)}^l(D_3^T \otimes I_{L_2})^T + I_{L_1}\widetilde{\mathcal{Y}}_{(1)}^l(D_3^T \otimes V_2)^T + I_{L_1}\widetilde{\mathcal{Y}}_{(1)}^l(D_3^T V_3 \otimes I_{L_2})^T.$$

Taking straightening operator into both sides of the relations $\widetilde{\widetilde{\mathcal{Z}}}$ and $\widetilde{\mathcal{Y}}_3^{l+1}$ results in

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}_3^{l+1}) &= [I_{L_1L_2L_3} - \gamma(D_3^T \otimes I_{L_2} \otimes V_1 + D_3^T \otimes V_2 \otimes I_{L_1} + (D_3^T V_3) \otimes I_{L_2} \otimes I_{L_1})]\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= \widetilde{M_3}\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&= \widetilde{M_3}\mathrm{vec}(\mathcal{Y}^l - \mathcal{Y}^*) \\
&= \widetilde{M_3}\mathrm{vec}[\frac{1}{3}(\mathcal{Y}_1^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_2^l - \mathcal{Y}^*) + \frac{1}{3}(\mathcal{Y}_3^l - \mathcal{Y}^*)] \\
&= \widetilde{M_3}\mathrm{vec}[\frac{1}{3}\widetilde{\mathcal{Y}}_1^l + \frac{1}{3}\widetilde{\mathcal{Y}}_2^l + \frac{1}{3}\widetilde{\mathcal{Y}}_3^l] \\
&= \frac{1}{3}\widetilde{M_3}[\mathrm{vec}(\widetilde{\mathcal{Y}}_1^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^l)],
\end{aligned}$$

where $\widetilde{M_3} = I_{L_1L_2L_3} - \gamma(D_3^T \otimes I_{L_2} \otimes V_1 + D_3^T \otimes V_2 \otimes I_{L_1} + (D_3^T V_3) \otimes I_{L_2} \otimes I_{L_1})$. Owing to $\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) = \frac{1}{3}[\mathrm{vec}(\widetilde{\mathcal{Y}}_1^{l+1}) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^{l+1}) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^{l+1})]$ and $\mathrm{vec}(\widetilde{\mathcal{Y}}^l) = \frac{1}{3}[\mathrm{vec}(\widetilde{\mathcal{Y}}_1^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^l)]$, we have

$$\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) = \frac{1}{3}[\mathrm{vec}(\widetilde{\mathcal{Y}}_1^{l+1}) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^{l+1}) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^{l+1})]$$

$$
\begin{aligned}
&= \frac{1}{3}(\frac{1}{3}\widetilde{M_1} + \frac{1}{3}\widetilde{M_2} + \frac{1}{3}\widetilde{M_3})(\mathrm{vec}(\widetilde{\mathcal{Y}}_1^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_2^l) + \mathrm{vec}(\widetilde{\mathcal{Y}}_3^l)) \\
&= (\frac{1}{3}\widetilde{M_1} + \frac{1}{3}\widetilde{M_2} + \frac{1}{3}\widetilde{M_3})\mathrm{vec}(\widetilde{\mathcal{Y}}^l). \qquad (3.15)
\end{aligned}
$$

Let

$$
\begin{aligned}
P_1 =& I_{L_3} \otimes I_{L_2} \otimes (D_1^T V_1) + I_{L_3} \otimes V_2 \otimes D_1^T + V_3 \otimes I_{L_2} \otimes D_1^T \\
&+ I_{L_3} \otimes D_2^T \otimes V_1 + I_{L_3} \otimes (D_2^T V_2) \otimes I_{L_1} + V_3 \otimes D_2^T \otimes I_{L_1} \\
&+ D_3^T \otimes I_{L_2} \otimes V_1 + D_3^T \otimes V_2 \otimes I_{L_1} + (D_3^T V_3) \otimes I_{L_2} \otimes I_{L_1} \\
=& (I_{L_3} \otimes I_{L_2} \otimes D_1 + I_{L_3} \otimes D_2 \otimes I_{L_1} + D_3 \otimes I_{L_2} \otimes I_{L_1})^T \\
&\times (I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}) \\
=& G_1^T G_2, \qquad (3.16)
\end{aligned}
$$

which together with the forms of $\widetilde{M_1}$, $\widetilde{M_2}$, $\widetilde{M_3}$ and (3.15) gives

$$
\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) = (\frac{1}{3}\widetilde{M_1} + \frac{1}{3}\widetilde{M_2} + \frac{1}{3}\widetilde{M_3})\mathrm{vec}(\widetilde{\mathcal{Y}}^l) = (I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2)\mathrm{vec}(\widetilde{\mathcal{Y}}^l).
$$

Then the matrix $I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2$ is the iteration matrix of the DGI algorithm, and hence the DGI algorithm is convergent if and only if $\rho(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) < 1$. It is not difficult to verify that

$$
\lambda_i(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) = 1 - \frac{1}{3}\gamma\lambda_i(G_1^T G_2),\ i = 1, 2, \cdots, L_1L_2L_3.
$$

We will study the condition of γ such that $\rho(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) < 1$. Let $\lambda_i(G_1^T G_2) = Re(\lambda_i) + iIm(\lambda_i)$ $(i = 1, 2, \cdots L_1L_2L_3)$ be the eigenvalues of the matrix $G_1^T G_2$, then $\rho(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) < 1$ is equivalent to

$$
\begin{aligned}
|\lambda_i(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2)| &= |1 - \frac{1}{3}\gamma\lambda_i(G_1^T G_2)| \\
&= |1 - \frac{1}{3}\gamma Re(\lambda_i) - \frac{1}{3}\gamma Im(\lambda_i)i| \\
&= \sqrt{(1 - \frac{1}{3}\gamma Re(\lambda_i))^2 + (\frac{1}{3}\gamma Im(\lambda_i))^2} \\
&< 1, \qquad (3.17)
\end{aligned}
$$

which together with $Re(\lambda_i) > 0$ for $i = 1, 2 \cdots L_1L_2L_3$ results in

$$
0 < \gamma < \min_{1\le i\le L_1L_2L_3} \frac{6Re(\lambda_i)}{Re(\lambda_i)^2 + Im(\lambda_i)^2}. \qquad (3.18)
$$

Then, it follows from straightforward computations that a sufficient condition for guaranteeing (3.18) holding is

$$
0 < \gamma < \min_{1\le i\le L_1L_2L_3} \frac{6Re(\lambda_i)}{Re(\lambda_i)^2 + Im_1^2}, \qquad (3.19)
$$

where $Im_1 = \max\limits_{1\le i\le L_1L_2L_3} |Im(\lambda_i)|$. Define a function $h(t) = \frac{t}{t^2+Im_1^2}$, $t \in [Re_{\min}, Re_{\max}]$ with $Re_{\max} = \max\limits_{1\le p\le L_1L_2L_3}\{Re(\lambda_p)\}$ and $Re_{\min} = \min\limits_{1\le p\le L_1L_2L_3}\{Re(\lambda_p)\}$. With concrete computations,

we obtain $h'(t) = \frac{Im_1^2 - t^2}{(t^2 + Im_1^2)^2}$. Below we distinguish three cases to discuss:
(i) If $Im_1 \geq Re_{\max}$, then $h(t)$ is monotonically increasing with t. Thus it follows that

$$\min_{1\leq i\leq L_1L_2L_3} \frac{6Re(\lambda_i)}{Re(\lambda_i)^2 + Im_1^2} = \frac{6Re_{\min}}{Re_{\min}^2 + Im_1^2}. \tag{3.20}$$

(ii) If $Im_1 \leq Re_{\min}$, then $h(t)$ is monotonic decreasing about t, which leads to

$$\min_{1\leq i\leq L_1L_2L_3} \frac{6Re(\lambda_i)}{Re(\lambda_i)^2 + Im_1^2} = \frac{6Re_{\max}}{Re_{\max}^2 + Im_1^2}. \tag{3.21}$$

(iii) If $Re_{\min} < Im_1 < Re_{\max}$, then by some calculations, it holds that

$$\min_{1\leq i\leq L_1L_2L_3} \frac{6Re(\lambda_i)}{Re(\lambda_i)^2 + Im_1^2} = \begin{cases} \dfrac{6Re_{\min}}{Re_{\min}^2 + Im_1^2}, \text{ if } Im_1 \geq \sqrt{Re_{\max}Re_{\min}}\ , \\ \dfrac{6Re_{\max}}{Re_{\max}^2 + Im_1^2}, \text{ if } Im_1 \leq \sqrt{Re_{\max}Re_{\min}}\ . \end{cases}$$

We obtain the following results by summarizing the above discussions

$$0 < \gamma < \begin{cases} \dfrac{6Re_{\min}}{Re_{\min}^2 + Im_1^2}, \text{ as } Im_1 \geq \sqrt{Re_{\max}Re_{\min}}\ , \\ \dfrac{6Re_{\max}}{Re_{\max}^2 + Im_1^2}, \text{ as } Im_1 \leq \sqrt{Re_{\max}Re_{\min}}\ . \end{cases}$$

The proof of this theorem is completed. □

Theorem 3.3. *Let the conditions of Theorem 3.2 be satisfied, then the quasi-optimal parameter γ_{opt} and the corresponding quasi-optimal convergence factor of the DGI algorithm are*
(1) If $Im_1 \geq \sqrt{Re_{\max}Re_{\min}}$, then

$$\gamma_{opt} = \frac{3Re_{\min}}{Re_{\min}^2 + Im_1^2}, \quad \rho_{opt}(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) = \frac{Im_1}{\sqrt{Re_{\min}^2 + Im_1^2}};$$

(2) If $Im_1 \leq \sqrt{Re_{\max}Re_{\min}}$, there are two cases:
(i) When $0 < \gamma \leq \frac{6}{Re_{\max}+Re_{\min}}$, if $Im_1^2 < \frac{Re_{\min}(Re_{\max}-Re_{\min})}{2}$, then $\gamma_{opt} = \frac{6}{Re_{\max}+Re_{\min}}$ and $\rho_{opt}(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) = \frac{\sqrt{(Re_{\max}-Re_{\min})^2+4Im_1^2}}{Re_{\max}+Re_{\min}}$. If $Im_1^2 \geq \frac{Re_{\min}(Re_{\max}-Re_{\min})}{2}$, then $\gamma_{opt} = \frac{3Re_{\min}}{Re_{\min}^2+Im_1^2}$ and

$$\rho_{opt}(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) = \frac{Im_1}{\sqrt{Re_{\min}^2 + Im_1^2}};$$

(ii) When $\frac{6}{Re_{\max}+Re_{\min}} \leq \gamma < \frac{6Re_{\max}}{Re_{\max}+Im_1^2}$, then $\gamma_{opt} = \frac{6}{Re_{\max}+Re_{\min}}$ and

$$\rho_{opt}(I_{L_1L_2L_3} - \frac{1}{3}\gamma G_1^T G_2) = \frac{\sqrt{(Re_{\max} - Re_{\min})^2 + 4Im_1^2}}{Re_{\max} + Re_{\min}}.$$

Proof. Define the function $B(Re(\lambda_i), Im(\lambda_i)) = (1-\frac{1}{3}\gamma Re(\lambda_i))^2 + (\frac{1}{3}\gamma Im(\lambda_i))^2$ with respect to $Re(\lambda_i)$ and $Im(\lambda_i)$. Inasmuch as $B(Re(\lambda_i), Im(\lambda_i))$ is increasing about $Im(\lambda_i)^2$, we have

$$\max B(Re(\lambda_i), Im(\lambda_i)) \le \max B(Re(\lambda_i), Im_1) = \max[(1-\frac{1}{3}\gamma Re(\lambda_i))^2 + (\frac{1}{3}\gamma Im_1)^2]. \tag{3.22}$$

Moreover, we can deduce that $\max B(Re(\lambda_i), Im_1) = \max[B(Re_{\max}, Im_1), B(Re_{\min}, Im_1)]$, from which one can deduce the following result

$$\max B(Re(\lambda_i), Im_1) = \begin{cases} (1-\frac{1}{3}\gamma Re_{\min})^2 + (\frac{1}{3}\gamma Im_1)^2, \text{ if } \gamma \le \dfrac{6}{Re_{\max}+Re_{\min}}, \\ (1-\frac{1}{3}\gamma Re_{\max})^2 + (\frac{1}{3}\gamma Im_1)^2, \text{ if } \gamma \ge \dfrac{6}{Re_{\max}+Re_{\min}}. \end{cases} \tag{3.23}$$

If $Im_1 \ge \sqrt{Re_{\max}Re_{\min}}$, then it holds that

$$\frac{6Re_{\min}}{Re_{\min}^2 + Im_1^2} \le \frac{6}{Re_{\min}+Re_{\max}}.$$

Therefore, it follows from (3.23) that

$$\max B(Re(\lambda_i), Im_1) = (1-\frac{1}{3}\gamma Re_{\min})^2 + (\frac{1}{3}\gamma Im_1)^2.$$

By some calculations, the minimum point of $\max B(Re(\lambda_i), Im_1)$ is $\gamma_{\text{opt}} = \frac{3Re_{\min}}{Re_{\min}^2+Im_1^2}$.

In addition, if $Im_1 \le \sqrt{Re_{\max}Re_{\min}}$, it follows that

$$\frac{6Re_{\max}}{Re_{\max}^2 + Im_1^2} \ge \frac{6}{Re_{\min}+Re_{\max}}.$$

Then, there are two cases to consider:
(i) If $0 < \gamma \le \frac{6}{Re_{\max}+Re_{\min}}$, then it follows from (3.23) that $\max B(Re(\lambda_i), Im_1) = (1-\frac{1}{3}\gamma Re_{\min})^2 + (\frac{1}{3}\gamma Im_1)^2$, which yields that $\bar{\gamma}_{\text{opt}} = \frac{3Re_{\min}}{Re_{\min}^2+Im_1^2}$. If $Im_1^2 < \frac{Re_{\min}(Re_{\max}-Re_{\min})}{2}$, then $\frac{6}{Re_{\max}+Re_{\min}} < \frac{3Re_{\min}}{Re_{\min}^2+Im_1^2}$ is valid. Taking into account that $\max B(Re(\lambda_i), Im_1)$ is decreasing about γ in the interval $(0, \frac{6}{Re_{\max}+Re_{\min}}]$, it has $\gamma_{\text{opt}} = \frac{6}{Re_{\min}+Re_{\max}}$. Besides, if $Im_1^2 \ge \frac{Re_{\min}(Re_{\max}-Re_{\min})}{2}$, then $\frac{3Re_{\min}}{Re_{\min}^2+Im_1^2} \le \frac{6}{Re_{\max}+Re_{\min}}$ and hence $\gamma_{\text{opt}} = \frac{3Re_{\min}}{Re_{\min}^2+Im_1^2}$.
(ii) If $\frac{6}{Re_{\min}+Re_{\max}} \le \gamma < \frac{6Re_{\max}}{Re_{\max}^2+Im_1^2}$, then from (3.23) we can obtain $\max B(Re(\lambda_i), Im_1) = (1-\frac{1}{3}\gamma Re_{\max})^2 + (\frac{1}{3}\gamma Im_1)^2$. After some computations, we derive $\tilde{\gamma}_{\text{opt}} = \frac{3Re_{\max}}{Re_{\max}^2+Im_1^2}$. Since $\max B(Re(\lambda_i), Im_1)$ is increasing about γ in the interval $[\frac{3Re_{\max}}{Re_{\max}^2+Im_1^2}, +\infty)$ and $\frac{6}{Re_{\max}+Re_{\min}} \ge \frac{3Re_{\max}}{Re_{\max}^2+Im_1^2}$, we conclude that $\max B(Re(\lambda_i), Im_1)$ is increasing about γ in the interval $[\frac{6}{Re_{\min}+Re_{\max}}, \frac{6Re_{\max}}{Re_{\max}^2+Im_1^2}]$ and therefore $\gamma_{\text{opt}} = \frac{6}{Re_{\min}+Re_{\max}}$. □

Algorithm 3.3. The new MGI (NMGI) algorithm:

Step 1. Given matrices $V_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3 \in \mathbb{R}^{L_3\times L_3}$, a tensor $\mathcal{W} \in \mathbb{R}^{L_1\times L_2\times L_3}$, and two constants $\gamma, \eta > 0$. Choose an initial tensor $\mathcal{Y}^0$, and set $l = 0$.

Step 2. If $\tau_l = \frac{\|\mathcal{W}-\mathcal{Y}^l\times_1 V_1-\mathcal{Y}^l\times_2 V_2-\mathcal{Y}^l\times_3 V_3\|}{\|\mathcal{W}-\mathcal{Y}^0\times_1 V_1-\mathcal{Y}^0\times_2 V_2-\mathcal{Y}^0\times_3 V_3\|} = \frac{\|\mathcal{R}^l\|}{\|\mathcal{R}^0\|} < \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ by the following procedure

$$\mathcal{Y}_1^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}^l\times_1 D_1^T,$$

$$
\begin{aligned}
\widetilde{\mathcal{Y}}^l &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}^l + \mathcal{Y}^l}{3}, \\
\mathcal{Y}_2^{l+1} &= \widetilde{\mathcal{Y}}^l + \gamma \widetilde{\mathcal{R}}^l \times_2 D_2^T, \\
\overline{\mathcal{Y}}^l &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}^l}{3}, \\
\mathcal{Y}_3^{l+1} &= \overline{\mathcal{Y}}^l + \gamma \overline{\mathcal{R}}^l \times_3 D_3^T, \\
\mathcal{Y}^{l+1} &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1}}{3}.
\end{aligned}
$$

Step 4. Set $l := l + 1$ and return to Step 2.

Theorem 3.4. *The NMGI algorithm is convergent if the parameter γ is selected to satisfy $\frac{1}{3}q_1 + \frac{1}{3}q_2q_3 + \frac{1}{9}q_1q_4 + \frac{1}{9}q_2q_3q_4 + \frac{1}{9}q_4 < 1$, where*

$$
\begin{aligned}
q_1 &= \|I_{N_1} - \gamma D_1^T V_1\|_2 + \gamma \|D_1\|_2 \|V_2\|_2 + \gamma \|D_1\|_2 \|V_3\|_2, \\
q_2 &= \|I_{N_1} - \frac{1}{3}\gamma D_1^T V_1\|_2 + \frac{1}{3}\gamma \|D_1\|_2 \|V_2\|_2 + \frac{1}{3}\gamma \|D_1\|_2 \|V_3\|_2, \\
q_3 &= \|I_{N_2} - \gamma D_2^T V_2\|_2 + \gamma \|D_2\|_2 \|V_1\|_2 + \gamma \|D_2\|_2 \|V_3\|_2, \\
q_4 &= \|I_{N_3} - \gamma D_3^T V_3\|_2 + \gamma \|D_3\|_2 \|V_1\|_2 + \gamma \|D_3\|_2 \|V_2\|_2.
\end{aligned}
$$

Proof. We define the error tensors

$$
\overrightarrow{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*, \ \overrightarrow{\widetilde{\mathcal{Y}}}^l = \widetilde{\mathcal{Y}}^l - \mathcal{Y}^*, \ \overrightarrow{\overline{\mathcal{Y}}}^l = \overline{\mathcal{Y}}^l - \mathcal{Y}^*, \ \overrightarrow{\mathcal{Y}}_i^{l+1} = \mathcal{Y}_i^{l+1} - \mathcal{Y}^* \ (i = 1, 2, 3).
$$

According to Line 1 of Step 3 of the NMGI algorithm, it follows that

$$
\begin{aligned}
\overrightarrow{\mathcal{Y}}_1^{l+1} &= \overrightarrow{\mathcal{Y}}^l - \gamma(\overrightarrow{\mathcal{Y}}^l \times_1 V_1 + \overrightarrow{\mathcal{Y}}^l \times_2 V_2 + \overrightarrow{\mathcal{Y}}^l \times_3 V_3) \times_1 D_1^T \\
&= \overrightarrow{\mathcal{Y}}^l - \gamma \overrightarrow{\mathcal{Y}}^l \times_1 (D_1^T V_1) - \gamma \overrightarrow{\mathcal{Y}}^l \times_1 D_1^T \times_2 V_2 - \gamma \overrightarrow{\mathcal{Y}}^l \times_1 D_1^T \times_3 V_3 \\
&= \overrightarrow{\mathcal{Y}}^l \times_1 (I_{L_1} - \gamma D_1^T V_1) - \gamma \overrightarrow{\mathcal{Y}}^l \times_1 D_1^T \times_2 V_2 - \gamma \overrightarrow{\mathcal{Y}}^l \times_1 D_1^T \times_3 V_3. \qquad (3.24)
\end{aligned}
$$

By making use of the expression of $\widetilde{\mathcal{Y}}^l$ in the NMGI algorithm, we deduce

$$
\overrightarrow{\widetilde{\mathcal{Y}}}^l = \overrightarrow{\mathcal{Y}}^l \times_1 (I_{L_1} - \frac{1}{3}\gamma D_1^T V_1) - \frac{1}{3}\gamma \overrightarrow{\mathcal{Y}}^l \times_1 D_1^T \times_2 V_2 - \frac{1}{3}\gamma \overrightarrow{\mathcal{Y}}^l \times_1 D_1^T \times_3 V_3. \qquad (3.25)
$$

From the expression of $\mathcal{Y}_2^{l+1}$ in the NMGI algorithm, we derive

$$
\begin{aligned}
\overrightarrow{\mathcal{Y}}_2^{l+1} &= \overrightarrow{\widetilde{\mathcal{Y}}}^l - \gamma[\overrightarrow{\widetilde{\mathcal{Y}}}^l \times_1 V_1 + \overrightarrow{\widetilde{\mathcal{Y}}}^l \times_2 V_2 + \overrightarrow{\widetilde{\mathcal{Y}}}^l \times_3 V_3] \times_2 D_2^T \\
&= \overrightarrow{\widetilde{\mathcal{Y}}}^l - \gamma[\overrightarrow{\widetilde{\mathcal{Y}}}^l \times_1 V_1 \times_2 D_2^T + \overrightarrow{\widetilde{\mathcal{Y}}}^l \times_2 D_2^T V_2 + \overrightarrow{\widetilde{\mathcal{Y}}}^l \times_2 D_2^T \times_3 V_3] \\
&= \overrightarrow{\widetilde{\mathcal{Y}}}^l \times_2 (I_{L_2} - \gamma D_2^T V_2) - \gamma \overrightarrow{\widetilde{\mathcal{Y}}}^l \times_1 V_1 \times_2 D_2^T - \gamma \overrightarrow{\widetilde{\mathcal{Y}}}^l \times_2 D_2^T \times_3 V_3. \qquad (3.26)
\end{aligned}
$$

In view of the expression of $\mathcal{Y}_3^{l+1}$ in the NMGI algorithm, it has

$$
\begin{aligned}
\overrightarrow{\mathcal{Y}}_3^{l+1} &= \overrightarrow{\overline{\mathcal{Y}}}^l - \gamma[\overrightarrow{\overline{\mathcal{Y}}}^l \times_1 V_1 + \overrightarrow{\overline{\mathcal{Y}}}^l \times_2 V_2 + \overrightarrow{\overline{\mathcal{Y}}}^l \times_3 V_3] \times_3 D_3^T \\
&= \overrightarrow{\overline{\mathcal{Y}}}^l - \gamma[\overrightarrow{\overline{\mathcal{Y}}}^l \times_1 V_1 \times_3 D_3^T + \overrightarrow{\overline{\mathcal{Y}}}^l \times_2 V_2 \times_3 D_3^T + \overrightarrow{\overline{\mathcal{Y}}}^l \times_3 (D_3^T V_3)]
\end{aligned}
$$

$$= \overrightarrow{\overline{\mathcal{Y}}}^{l} \times_3 (I_{L_3} - \gamma D_3^T V_3) - \gamma \overrightarrow{\overline{\mathcal{Y}}}^{l} \times_1 V_1 \times_3 D_3^T - \gamma \overrightarrow{\overline{\mathcal{Y}}}^{l} \times_2 V_2 \times_3 D_3^T. \tag{3.27}$$

The combination of Lemma 2.2 and (3.24) leads to

$$\begin{aligned}
\|\overrightarrow{\mathcal{Y}}_1^{l+1}\| &= \|\overrightarrow{\mathcal{Y}}^{l} \times_1 (I_{L_1} - \gamma D_1^T V_1) - \gamma \overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_2 V_2 - \gamma \overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_3 V_3\| \\
&\leq \|\overrightarrow{\mathcal{Y}}^{l} \times_1 (I_{L_1} - \gamma D_1^T V_1)\| + \gamma \|\overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_2 V_2\| + \gamma \|\overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_3 V_3\| \\
&\leq \|\overrightarrow{\mathcal{Y}}^{l}\| \|I_{L_1} - \gamma D_1^T V_1\|_2 + \gamma \|\overrightarrow{\mathcal{Y}}^{l}\| \|D_1\|_2 \|V_2\|_2 + \gamma \|\overrightarrow{\mathcal{Y}}^{l}\| \|D_1\|_2 \|V_3\|_2 \\
&= \|\overrightarrow{\mathcal{Y}}^{l}\| [\|I_{L_1} - \gamma D_1^T V_1\|_2 + \gamma \|D_1\|_2 \|V_2\|_2 + \gamma \|D_1\|_2 \|V_3\|_2] \\
&= q_1 \|\overrightarrow{\mathcal{Y}}^{l}\|,
\end{aligned} \tag{3.28}$$

where $q_1 = \|I_{L_1} - \gamma D_1^T V_1\|_2 + \gamma \|D_1\|_2 \|V_2\|_2 + \gamma \|D_1\|_2 \|V_3\|_2$. Besides, it follows from (3.25) that

$$\begin{aligned}
\|\overrightarrow{\widetilde{\mathcal{Y}}}^{l}\| &= \|\overrightarrow{\mathcal{Y}}^{l} \times_1 (I_{L_1} - \frac{1}{3}\gamma D_1^T V_1) - \frac{1}{3}\gamma \overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_2 V_2 - \frac{1}{3}\gamma \overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_3 V_3\| \\
&\leq \|\overrightarrow{\mathcal{Y}}^{l} \times_1 (I_{L_1} - \frac{1}{3}\gamma D_1^T V_1)\| + \frac{1}{3}\gamma \|\overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_2 V_2\| + \frac{1}{3}\gamma \|\overrightarrow{\mathcal{Y}}^{l} \times_1 D_1^T \times_3 V_3\| \\
&\leq \|\overrightarrow{\mathcal{Y}}^{l}\| \|I_{L_1} - \frac{1}{3}\gamma D_1^T V_1\|_2 + \frac{1}{3}\gamma \|\overrightarrow{\mathcal{Y}}^{l}\| \|D_1\|_2 \|V_2\|_2 + \frac{1}{3}\gamma \|\overrightarrow{\mathcal{Y}}^{l}\| \|D_1\|_2 \|V_3\|_2 \\
&= \|\overrightarrow{\mathcal{Y}}^{l}\| [\|I_{L_1} - \frac{1}{3}\gamma D_1^T V_1\|_2 + \frac{1}{3}\gamma \|D_1\|_2 \|V_2\|_2 + \frac{1}{3}\gamma \|D_1\|_2 \|V_3\|_2] \\
&= q_2 \|\overrightarrow{\mathcal{Y}}^{l}\|,
\end{aligned} \tag{3.29}$$

with $q_2 = \|I_{L_1} - \frac{1}{3}\gamma D_1^T V_1\|_2 + \frac{1}{3}\gamma \|D_1\|_2 \|V_2\|_2 + \frac{1}{3}\gamma \|D_1\|_2 \|V_3\|_2$. It can be derived from (3.26) that

$$\begin{aligned}
\|\overrightarrow{\mathcal{Y}}_2^{l+1}\| &= \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l} \times_2 (I_{L_2} - \gamma D_2^T V_2) - \gamma \overrightarrow{\widetilde{\mathcal{Y}}}^{l} \times_1 V_1 \times_2 D_2^T - \gamma \overrightarrow{\widetilde{\mathcal{Y}}}^{l} \times_2 D_2^T \times_3 V_3\| \\
&\leq \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l} \times_2 (I_{L_2} - \gamma D_2^T V_2)\| + \gamma \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l} \times_1 V_1 \times_2 D_2^T\| + \gamma \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l} \times_2 D_2^T \times_3 V_3\| \\
&\leq \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l}\| \|I_{L_2} - \gamma D_2^T V_2\|_2 + \gamma \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l}\| \|V_1\|_2 \|D_2\|_2 + \gamma \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l}\| \|D_2\|_2 \|V_3\|_2 \\
&= \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l}\| [\|I_{L_2} - \gamma D_2^T V_2\|_2 + \gamma \|V_1\|_2 \|D_2\|_2 + \gamma \|D_2\|_2 \|V_3\|_2] \\
&= q_3 \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l}\|,
\end{aligned} \tag{3.30}$$

in terms of Lemma 2.2, where $q_3 = \|I_{L_2} - \gamma D_2^T V_2\|_2 + \gamma \|V_1\|_2 \|D_2\|_2 + \gamma \|D_2\|_2 \|V_3\|_2$. By Line 4 of Step 3 of the NMGI algorithm, it holds that

$$\begin{aligned}
\|\overrightarrow{\overline{\mathcal{Y}}}^{l}\| &= \|\frac{1}{3}\overrightarrow{\mathcal{Y}}_1^{l+1} + \frac{1}{3}\overrightarrow{\mathcal{Y}}_2^{l+1} + \frac{1}{3}\overrightarrow{\mathcal{Y}}_3^{l}\| \\
&\leq \frac{1}{3}\|\overrightarrow{\mathcal{Y}}_1^{l+1}\| + \frac{1}{3}\|\overrightarrow{\mathcal{Y}}_2^{l+1}\| + \frac{1}{3}\|\overrightarrow{\mathcal{Y}}_3^{l}\| \\
&\leq \frac{1}{3}q_1 \|\overrightarrow{\mathcal{Y}}^{l}\| + \frac{1}{3}q_3 \|\overrightarrow{\widetilde{\mathcal{Y}}}^{l}\| + \frac{1}{3}\|\overrightarrow{\mathcal{Y}}^{l}\| \\
&\leq \frac{1}{3}q_1 \|\overrightarrow{\mathcal{Y}}^{l}\| + \frac{1}{3}q_3 q_2 \|\overrightarrow{\mathcal{Y}}^{l}\| + \frac{1}{3}\|\overrightarrow{\mathcal{Y}}^{l}\| \\
&= (\frac{1}{3}q_1 + \frac{1}{3}q_3 q_2 + \frac{1}{3})\|\overrightarrow{\mathcal{Y}}^{l}\|.
\end{aligned} \tag{3.31}$$

By applying the Frobenious norm to (3.27) and using Lemmas 2.1–2.2, we get

$$\begin{aligned}
\|\overrightarrow{\mathcal{Y}}_3^{l+1}\| &= \|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l \times_3 (I_{L_3} - \gamma D_3^T V_3) - \gamma \overrightarrow{\overrightarrow{\mathcal{Y}}}^l \times_1 V_1 \times_3 D_3^T - \gamma \overrightarrow{\overrightarrow{\mathcal{Y}}}^l \times_2 V_2 \times_3 D_3^T\| \\
&\leq \|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l \times_3 (I_{L_3} - \gamma D_3^T V_3)\| + \gamma\|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l \times_1 V_1 \times_3 D_3^T\| + \gamma\|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l \times_2 V_2 \times_3 D_3^T\| \\
&\leq \|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l\|\|I_{L_3} - \gamma D_3^T V_3\|_2 + \gamma\|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l\|\|V_1\|_2\|D_3\|_2 + \gamma\|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l\|\|V_2\|_2\|D_3\|_2 \\
&= \|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l\|[\|I_{L_3} - \gamma D_3^T V_3\|_2 + \gamma\|V_1\|_2\|D_3\|_2 + \gamma\|V_2\|_2\|D_3\|_2] \\
&= q_4\|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l\|,
\end{aligned} \tag{3.32}$$

where $q_4 = \|I_{L_3} - \gamma D_3^T V_3\|_2 + \gamma\|V_1\|_2\|D_3\|_2 + \gamma\|V_2\|_2\|D_3\|_2$. The combination of (3.31) and (3.32) gives $\|\overrightarrow{\mathcal{Y}}_3^{l+1}\| \leq q_4\|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l\| \leq q_4[\frac{1}{3}q_1 + \frac{1}{3}q_3q_2 + \frac{1}{3}]\|\overrightarrow{\mathcal{Y}}^l\|$, which combines with $\|\overrightarrow{\mathcal{Y}}_2^{l+1}\| \leq q_3\|\overrightarrow{\overrightarrow{\mathcal{Y}}}^l\| \leq q_3q_2\|\overrightarrow{\mathcal{Y}}^l\|$ and $\|\overrightarrow{\mathcal{Y}}_1^{l+1}\| \leq q_1\|\overrightarrow{\mathcal{Y}}^l\|$ leads to

$$\begin{aligned}
\|\overrightarrow{\mathcal{Y}}^{l+1}\| &= \|\frac{1}{3}\overrightarrow{\mathcal{Y}}_1^{l+1} + \frac{1}{3}\overrightarrow{\mathcal{Y}}_2^{l+1} + \frac{1}{3}\overrightarrow{\mathcal{Y}}_3^{l+1}\| \\
&\leq \frac{1}{3}\|\overrightarrow{\mathcal{Y}}_1^{l+1}\| + \frac{1}{3}\|\overrightarrow{\mathcal{Y}}_2^{l+1}\| + \frac{1}{3}\|\overrightarrow{\mathcal{Y}}_3^{l+1}\| \\
&\leq \frac{1}{3}q_1\|\overrightarrow{\mathcal{Y}}^l\| + \frac{1}{3}q_2q_3\|\overrightarrow{\mathcal{Y}}^l\| + \frac{1}{3}(\frac{1}{3}q_1q_4 + \frac{1}{3}q_2q_3q_4 + \frac{1}{3}q_4)\|\overrightarrow{\mathcal{Y}}^l\| \\
&= [\frac{1}{3}q_1 + \frac{1}{3}q_2q_3 + \frac{1}{9}q_1q_4 + \frac{1}{9}q_2q_3q_4 + \frac{1}{9}q_4]\|\overrightarrow{\mathcal{Y}}^l\| \\
&\leq [\frac{1}{3}q_1 + \frac{1}{3}q_2q_3 + \frac{1}{9}q_1q_4 + \frac{1}{9}q_2q_3q_4 + \frac{1}{9}q_4]^{l+1}\|\overrightarrow{\mathcal{Y}}^0\|.
\end{aligned}$$

Thus if $\frac{1}{3}q_1 + \frac{1}{3}q_2q_3 + \frac{1}{9}q_1q_4 + \frac{1}{9}q_2q_3q_4 + \frac{1}{9}q_4 < 1$, then $\lim_{l\to+\infty}\|\overrightarrow{\mathcal{Y}}^{l+1}\| = 0$ and therefore $\mathcal{Y}^l \to \mathcal{Y}^*(l \to +\infty)$, i.e., the NMGI algorithm is convergent. □

4. The DGI and NMGI algorithms for the coupled Sylvester tensor equations

In this section, we extend the DGI and NMGI algorithms to solve the more general coupled Sylvester tensor equations, whose form is as follows

$$\begin{cases} \mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W}_1, \\ \mathcal{Y} \times_1 T_1 + \mathcal{Y} \times_2 T_2 + \mathcal{Y} \times_3 T_3 = \mathcal{W}_2, \end{cases} \tag{4.1}$$

where $V_1, T_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3 \in \mathbb{R}^{L_3\times L_3}$, $\mathcal{W}_1, W_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and the unknown tensor $\mathcal{Y} \in \mathbb{R}^{L_1\times L_2\times L_3}$ needs to be computed.

Based on the hierarchical identification principle, we can establish the following GI algorithm for solving the coupled Sylvester tensor equation (4.1).

Algorithm 4.1. The gradient-based iterative (GI) algorithm:

Step 1. Given matrices $V_1, T_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3 \in \mathbb{R}^{L_3\times L_3}$, two tensors $\mathcal{W}_1, \mathcal{W}_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and positive constants γ, η. Choose the initial tensor $\mathcal{Y}^0$, and set $l = 0$.

Step 2. If $\tau_l = \dfrac{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^l\times_1 V_1-\mathcal{Y}^l\times_2 V_2-\mathcal{Y}^l\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^l\times_1 T_1-\mathcal{Y}^l\times_2 T_2-\mathcal{Y}^l\times_3 T_3\|^2}}{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^0\times_1 V_1-\mathcal{Y}^0\times_2 V_2-\mathcal{Y}^0\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^0\times_1 T_1-\mathcal{Y}^0\times_2 T_2-\mathcal{Y}^0\times_3 T_3\|^2}}$

$< \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ by the following procedure

$$\begin{aligned}
&\mathcal{R}_1^l = \mathcal{W}_1 - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3,\\
&\mathcal{R}_2^l = \mathcal{W}_2 - \mathcal{Y}^l \times_1 T_1 - \mathcal{Y}^l \times_2 T_2 - \mathcal{Y}^l \times_3 T_3,\\
&\mathcal{Y}_1^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_1^l \times_1 V_1^T,\\
&\mathcal{Y}_2^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_1^l \times_2 V_2^T,\\
&\mathcal{Y}_3^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_1^l \times_3 V_3^T,\\
&\mathcal{Y}_4^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_2^l \times_1 T_1^T,\\
&\mathcal{Y}_5^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_2^l \times_2 T_2^T,\\
&\mathcal{Y}_6^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_2^l \times_3 T_3^T,\\
&\mathcal{Y}^{l+1} = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}_6^{l+1}}{6}.
\end{aligned}$$

Step 4. Set $l := l + 1$ and return to Step 2.

We decompose the system matrices V_j into $V_j = D_j + C_j$ $(j = 1, 2, 3)$ with C_j and D_j being the non-diagonal and diagonal parts of V_j, respectively. Meanwhile, we split the system matrices T_j into $T_j = D_{jj} + C_{jj}$ $(j = 1, 2, 3)$ with C_{jj} and D_{jj} being the non-diagonal and diagonal parts of T_j, respectively. Then it follows from (4.1) that

$$\begin{aligned}
&\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W}_1 \ \Rightarrow \mathcal{Y} \times_1 (D_1 + C_1) + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W}_1,\\
&\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W}_1 \ \Rightarrow \mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 (D_2 + C_2) + \mathcal{Y} \times_3 V_3 = \mathcal{W}_1,\\
&\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 = \mathcal{W}_1 \ \Rightarrow \mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 (D_3 + C_3) = \mathcal{W}_1,\\
&\mathcal{Y} \times_1 T_1 + \mathcal{Y} \times_2 T_2 + \mathcal{Y} \times_3 T_3 = \mathcal{W}_2 \ \Rightarrow \mathcal{Y} \times_1 (D_{11} + C_{11}) + \mathcal{Y} \times_2 T_2 + \mathcal{Y} \times_3 T_3 = \mathcal{W}_2,\\
&\mathcal{Y} \times_1 T_1 + \mathcal{Y} \times_2 T_2 + \mathcal{Y} \times_3 T_3 = \mathcal{W}_2 \ \Rightarrow \mathcal{Y} \times_1 T_1 + \mathcal{Y} \times_2 (D_{22} + C_{22}) + \mathcal{Y} \times_3 T_3 = \mathcal{W}_2,\\
&\mathcal{Y} \times_1 T_1 + \mathcal{Y} \times_2 T_2 + \mathcal{Y} \times_3 T_3 = \mathcal{W}_2 \ \Rightarrow \mathcal{Y} \times_1 T_1 + \mathcal{Y} \times_2 T_2 + \mathcal{Y} \times_3 (D_{33} + C_{33}) = \mathcal{W}_2.
\end{aligned} \tag{4.2}$$

Equations (4.2) directly lead to the following results

$$\begin{aligned}
\mathcal{Y} \times_1 D_1 &= \mathcal{W}_1 - \mathcal{Y} \times_1 C_1 - \mathcal{Y} \times_2 V_2 - \mathcal{Y} \times_3 V_3,\ \mathcal{Y} \times_2 D_2\\
&= \mathcal{W}_1 - \mathcal{Y} \times_1 V_1 - \mathcal{Y} \times_2 C_2 - \mathcal{Y} \times_3 V_3,\\
\mathcal{Y} \times_3 D_3 &= \mathcal{W}_1 - \mathcal{Y} \times_1 V_1 - \mathcal{Y} \times_2 V_2 - \mathcal{Y} \times_3 C_3,\ \mathcal{Y} \times_1 D_{11}\\
&= \mathcal{W}_2 - \mathcal{Y} \times_1 C_{11} - \mathcal{Y} \times_2 T_2 - \mathcal{Y} \times_3 T_3,\\
\mathcal{Y} \times_2 D_{22} &= \mathcal{W}_2 - \mathcal{Y} \times_1 T_1 - \mathcal{Y} \times_2 C_{22} - \mathcal{Y} \times_3 T_3,\ \mathcal{Y} \times_3 D_{33}\\
&= \mathcal{W}_2 - \mathcal{Y} \times_1 T_1 - \mathcal{Y} \times_2 T_2 - \mathcal{Y} \times_3 C_{33}.
\end{aligned} \tag{4.3}$$

By combining (4.3) with the GI algorithm, we can establish the following diagonal GI (DGI) algorithm.

Algorithm 4.2. The diagonal GI (DGI) algorithm:

Step 1. Given matrices $V_1, T_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3 \in \mathbb{R}^{L_3\times L_3}$, two tensors $\mathcal{W}_1, \mathcal{W}_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and positive constants γ, η. Choose the initial tensor $\mathcal{Y}^0$, and set $l = 0$.

Step 2. If $\tau_l = \dfrac{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^l\times_1 V_1-\mathcal{Y}^l\times_2 V_2-\mathcal{Y}^l\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^l\times_1 T_1-\mathcal{Y}^l\times_2 T_2-\mathcal{Y}^l\times_3 T_3\|^2}}{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^0\times_1 V_1-\mathcal{Y}^0\times_2 V_2-\mathcal{Y}^0\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^0\times_1 T_1-\mathcal{Y}^0\times_2 T_2-\mathcal{Y}^0\times_3 T_3\|^2}}$ $< \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ by the following procedure

$$\begin{aligned}
&\mathcal{R}_1^l = \mathcal{W}_1 - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3,\\
&\mathcal{R}_2^l = \mathcal{W}_2 - \mathcal{Y}^l \times_1 T_1 - \mathcal{Y}^l \times_2 T_2 - \mathcal{Y}^l \times_3 T_3,\\
&\mathcal{Y}_1^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_1^l \times_1 D_1^T,\\
&\mathcal{Y}_2^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_1^l \times_2 D_2^T,\\
&\mathcal{Y}_3^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_1^l \times_3 D_3^T,\\
&\mathcal{Y}_4^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_2^l \times_1 D_{11}^T,\\
&\mathcal{Y}_5^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_2^l \times_2 D_{22}^T,\\
&\mathcal{Y}_6^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_2^l \times_3 D_{33}^T,\\
&\mathcal{Y}^{l+1} = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}_6^{l+1}}{6}.
\end{aligned}$$

Step 4. Set $l := l + 1$ and return to Step 2.

We will discuss the convergence properties of the GI algorithm for the coupled Sylvester tensor equation (4.1), which include its convergence condition, optimal parameter and optimal convergence factor.

Theorem 4.1. *Let $M_1 = I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}$, $N_1 = I_{L_3} \otimes I_{L_2} \otimes T_1 + I_{L_3} \otimes T_2 \otimes I_{L_1} + T_3 \otimes I_{L_2} \otimes I_{L_1}$, and $\overline{\lambda}_{\max}$ and $\overline{\lambda}_{\min}$ be the maximum and minimum eigenvalues of the matrix $M_1^T M_1 + N_1^T N_1$, respectively. Suppose that the coupled Sylvester tensor equation* (4.1) *has a unique solution $\mathcal{Y}^*$. Then the GI algorithm is convergent if and only if*

$$0 < \gamma < \frac{12}{\overline{\lambda}_{\max}}.$$

And the optimal parameter γ_{opt} and the corresponding optimal convergence factor ρ_{opt} of the GI algorithm are

$$\gamma_{opt} = \frac{12}{\overline{\lambda}_{\max} + \overline{\lambda}_{\min}}, \quad \rho_{opt} = \frac{\overline{\lambda}_{\max} - \overline{\lambda}_{\min}}{\overline{\lambda}_{\max} + \overline{\lambda}_{\min}}.$$

Proof. Define the error tensor $\widetilde{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*$. It follows from the framework of the GI algorithm that

$$\mathcal{Y}^{l+1} = \mathcal{Y}^l + \frac{\gamma}{6}(\mathcal{R}_1^l \times_1 V_1^T + \mathcal{R}_1^l \times_2 V_2^T + \mathcal{R}_1^l \times_3 V_3^T + \mathcal{R}_2^l \times_1 T_1^T + \mathcal{R}_2^l \times_2 T_2^T + \mathcal{R}_2^l \times_3 T_3^T).$$

Inasmuch as

$$\mathcal{R}_1^l = \mathcal{W}_1 - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3 = -(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3),$$

$$\mathcal{R}_2^l = \mathcal{W}_2 - \mathcal{Y}^l \times_1 T_1 - \mathcal{Y}^l \times_2 T_2 - \mathcal{Y}^l \times_3 T_3 = -(\widetilde{\mathcal{Y}}^l \times_1 T_1 + \widetilde{\mathcal{Y}}^l \times_2 T_2 + \widetilde{\mathcal{Y}}^l \times_3 T_3),$$

it has

$$\begin{aligned}\widetilde{\mathcal{Y}}^{l+1} = \widetilde{\mathcal{Y}}^l - \frac{\gamma}{6}[&\widetilde{\mathcal{Y}}^l \times_1 (V_1^T V_1) \times_2 I_{L_2} \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 V_1^T \times_2 V_2 \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 V_1^T \times_2 I_{L_2} \times_3 V_3 \\ &+\widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 V_2^T \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 (V_2^T V_2) \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 V_2^T \times_3 V_3 \\ &+\widetilde{\mathcal{Y}}^l \times_1 V_1 \times_2 I_{L_2} \times_3 V_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 V_2 \times_3 V_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 I_{L_2} \times_3 (V_3^T V_3) \\ &+\widetilde{\mathcal{Y}}^l \times_1 (T_1^T T_1) \times_2 I_{L_2} \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 T_1^T \times_2 T_2 \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 T_1^T \times_2 I_{L_2} \times_3 T_3 \\ &+\widetilde{\mathcal{Y}}^l \times_1 T_1 \times_2 T_2^T \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 (T_2^T T_2) \times_3 I_{L_3} + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 T_2^T \times_3 T_3 \\ &+\widetilde{\mathcal{Y}}^l \times_1 T_1 \times_2 I_{L_2} \times_3 T_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 T_2 \times_3 T_3^T + \widetilde{\mathcal{Y}}^l \times_1 I_{L_1} \times_2 I_{L_2} \times_3 (T_3^T T_3)]. \quad (4.4)\end{aligned}$$

Applying straightening operator to (4.4) yields that

$$\begin{aligned}&\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) \\ =\ &\mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}[I_{L_3} \otimes I_{L_2} \otimes (V_1^T V_1) + I_{L_3} \otimes V_2 \otimes V_1^T + V_3 \otimes I_{L_2} \otimes V_1^T + I_{L_3} \otimes V_2^T \otimes V_1 \\ &+I_{L_3} \otimes (V_2^T V_2) \otimes I_{L_1} + V_3 \otimes V_2^T \otimes I_{L_1} + V_3^T \otimes I_{L_2} \otimes V_1 \\ &+V_3^T \otimes V_2 \otimes I_{L_1} + (V_3^T V_3) \otimes I_{L_2} \otimes I_{L_1} \\ &+I_{L_3} \otimes I_{L_2} \otimes (T_1^T T_1) + I_{L_3} \otimes T_2 \otimes T_1^T + T_3 \otimes I_{L_2} \otimes T_1^T + I_{L_3} \otimes T_2^T \otimes T_1 \\ &+I_{L_3} \otimes (T_2^T T_2) \otimes I_{L_1} + T_3 \otimes T_2^T \otimes I_{L_1} + T_3^T \otimes I_{L_2} \otimes T_1 + T_3^T \otimes T_2 \otimes I_{L_1} \\ &+(T_3^T T_3) \otimes I_{L_2} \otimes I_{L_1}]\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ =\ &\mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}[(I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1})^T \\ &\times(I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}) \\ &+(I_{L_3} \otimes I_{L_2} \otimes T_1 + I_{L_3} \otimes T_2 \otimes I_{L_1} + T_3 \otimes I_{L_2} \otimes I_{L_1})^T \\ &\times(I_{L_3} \otimes I_{L_2} \otimes T_1 + I_{L_3} \otimes T_2 \otimes I_{L_1} + T_3 \otimes I_{L_2} \otimes I_{L_1})]\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ =\ &\mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}(M_1^T M_1 + N_1^T N_1)\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ =\ &\mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}\begin{pmatrix} M_1^T & N_1^T \end{pmatrix}\begin{pmatrix} M_1 \\ N_1 \end{pmatrix}\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ :=\ &\mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}K^T K\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ =\ &(I_{L_1L_2L_3} - \frac{\gamma}{6}K^T K)\mathrm{vec}(\widetilde{\mathcal{Y}}^l),\end{aligned}$$

where $K = \begin{pmatrix} M_1 \\ N_1 \end{pmatrix}$. Hence $I_{L_1L_2L_3} - \frac{\gamma}{6}K^T K$ is the iteration matrix of the GI algorithm. By assumptions, (4.1) has a unique solution, then it holds that the matrix K is of full column rank and hence $K^T K = M_1^T M_1 + N_1^T N_1$ is a symmetric positive-definite matrix. Let $\overline{\lambda}_i > 0$ $(i = 1, 2, \cdots, L_1L_2L_3)$ be the eigenvalues of $M_1^T M_1 + N_1^T N_1$, then $\rho(I_{L_1L_2L_3} - \frac{\gamma}{6}K^T K) = \max\limits_{1\le i\le L_1L_2L_3}\{|1 - \frac{\gamma}{6}\overline{\lambda}_i|\} = \max\{|1 - \frac{\gamma}{6}\overline{\lambda}_{\min}|, |1 - \frac{\gamma}{6}\overline{\lambda}_{\max}|\}$. Then the GI algorithm is convergent if and only if $\rho(I_{L_1L_2L_3} - \frac{\gamma}{6}K^T K) = \max\limits_{1\le i\le L_1L_2L_3}\{|1 - \frac{\gamma}{6}\overline{\lambda}_i|\} < 1$, which is equivalent to $0 < \gamma < \frac{12}{\overline{\lambda}_{\max}}$.

In addition, by some computations, we have

$$\rho(I_{L_1L_2L_3}-\frac{\gamma}{6}K^TK)=\begin{cases}1-\frac{\gamma}{6}\overline{\lambda}_{\min},\ if\ \gamma\le\frac{12}{\overline{\lambda}_{\max}+\overline{\lambda}_{\min}},\\ \frac{\gamma}{6}\overline{\lambda}_{\max}-1,\ if\ \gamma>\frac{12}{\overline{\lambda}_{\max}+\overline{\lambda}_{\min}}.\end{cases}$$

When $\gamma\le\frac{12}{\overline{\lambda}_{\max}+\overline{\lambda}_{\min}}$, $\rho(I_{L_1L_2L_3}-\frac{\gamma}{6}K^TK)=1-\frac{\gamma}{6}\overline{\lambda}_{\min}$ is monotonic decreasing about γ. And when $\gamma>\frac{12}{\overline{\lambda}_{\max}+\overline{\lambda}_{\min}}$, $\rho(I_{L_1L_2L_3}-\frac{\gamma}{6}K^TK)=\frac{\gamma}{6}\overline{\lambda}_{\max}-1$ is monotonic increasing about the variable γ, then the optimal parameter γ_{opt} is

$$\gamma_{\text{opt}}=\frac{12}{\overline{\lambda}_{\max}+\overline{\lambda}_{\min}}.$$

Taking γ_{opt} into $\rho(I_{L_1L_2L_3}-\frac{\gamma}{6}K^TK)=1-\frac{\gamma}{6}\overline{\lambda}_{\min}$ yields the optimal convergence factor $\rho_{\text{opt}}=\frac{\overline{\lambda}_{\max}-\overline{\lambda}_{\min}}{\overline{\lambda}_{\max}+\overline{\lambda}_{\min}}$. □

In what follows, we establish the convergence properties of the DGI algorithm for the coupled Sylvester tensor equation (4.1).

Theorem 4.2. *Assume that the conditions of Theorem 4.1 are satisfied, and $Re(\overline{\delta}_q)>0$ with $\overline{\delta}_q$ $(q=1,\cdots,L_1L_2L_3)$ being the eigenvalues of K_1^TK, where $K_1=\begin{pmatrix}\overline{M}_1\\ \overline{N}_1\end{pmatrix}$, $K=\begin{pmatrix}M_1\\ N_1\end{pmatrix}$ and*

$$\begin{aligned}\overline{M}_1&=I_{L_3}\otimes I_{L_2}\otimes D_1+I_{L_3}\otimes D_2\otimes I_{L_1}+D_3\otimes I_{L_2}\otimes I_{L_1},\\ \overline{N}_1&=I_{L_3}\otimes I_{L_2}\otimes D_{11}+I_{L_3}\otimes D_{22}\otimes I_{L_1}+D_{33}\otimes I_{L_2}\otimes I_{L_1},\\ M_1&=I_{L_3}\otimes I_{L_2}\otimes V_1+I_{L_3}\otimes V_2\otimes I_{L_1}+V_3\otimes I_{L_2}\otimes I_{L_1},\\ N_1&=I_{L_3}\otimes I_{L_2}\otimes T_1+I_{L_3}\otimes T_2\otimes I_{L_1}+T_3\otimes I_{L_2}\otimes I_{L_1}.\end{aligned}$$

Let $\overline{Im}_1=\max\limits_{1\le q\le L_1L_2L_3}|Im(\overline{\delta}_q)|$, $\overline{Re}_{\max}=\max\limits_{1\le q\le L_1L_2L_3}\{Re(\overline{\delta}_q)\}$ and $\overline{Re}_{\min}=\min\limits_{1\le q\le L_1L_2L_3}\{Re(\overline{\delta}_q)\}$. Then the DGI algorithm is convergent provided that

$$0<\gamma<\begin{cases}\dfrac{12\overline{Re}_{\min}}{\overline{Re}_{\min}^2+\overline{Im}_1^2},\ as\ \overline{Im}_1\ge\sqrt{\overline{Re}_{\max}\overline{Re}_{\min}}\ ,\\ \dfrac{12\overline{Re}_{\max}}{\overline{Re}_{\max}^2+\overline{Im}_1^2},\ as\ \overline{Im}_1\le\sqrt{\overline{Re}_{\max}\overline{Re}_{\min}}\ .\end{cases}$$

Proof. Define the error tensor $\widetilde{\mathcal{Y}}^l=\mathcal{Y}^l-\mathcal{Y}^*$. From the framework of the DGI algorithm, it has

$$\begin{aligned}\mathcal{Y}^{l+1}=\ &\mathcal{Y}^l+\frac{\gamma}{6}(\mathcal{R}_1^l\times_1D_1^T+\mathcal{R}_1^l\times_2D_2^T+\mathcal{R}_1^l\times_3D_3^T+\mathcal{R}_2^l\times_1D_{11}^T\\ &+\mathcal{R}_2^l\times_2D_{22}^T+\mathcal{R}_2^l\times_3D_{33}^T).\end{aligned}\tag{4.5}$$

Applying straightening operator to (4.5) and making use of Lemma 2.1 yield that

$$\text{vec}(\mathcal{Y}^{l+1})=\text{vec}(\mathcal{Y}^l)+\frac{\gamma}{6}[(I_{L_3}\otimes I_{L_2}\otimes D_1^T+I_{L_3}\otimes D_2^T\otimes I_{L_1}+D_3^T\otimes I_{L_2}\otimes I_{L_1})\text{vec}(\mathcal{R}_1^l)$$

$$
\begin{aligned}
&+(I_{L_3}\otimes I_{L_2}\otimes D_{11}^T+I_{L_3}\otimes D_{22}^T\otimes I_{L_1}+D_{33}^T\otimes I_{L_2}\otimes I_{L_1})\mathrm{vec}(\mathcal{R}_2^l)]\\
&=\mathrm{vec}(\mathcal{Y}^l)+\frac{\gamma}{6}[\overline{M}_1^T\mathrm{vec}(\mathcal{R}_1^l)+\overline{N}_1^T\mathrm{vec}(\mathcal{R}_2^l)]. \qquad (4.6)
\end{aligned}
$$

Direct computations give

$$
\begin{aligned}
\mathrm{vec}(\mathcal{R}_1^l)&=-\mathrm{vec}(\widetilde{\mathcal{Y}}^l\times_1 V_1+\widetilde{\mathcal{Y}}^l\times_2 V_2+\widetilde{\mathcal{Y}}^l\times_3 V_3)\\
&=-(I_{L_3}\otimes I_{L_2}\otimes V_1+I_{L_3}\otimes V_2\otimes I_{L_1}+V_3\otimes I_{L_2}\otimes I_{L_1})\mathrm{vec}(\widetilde{\mathcal{Y}}^l)\\
&=-M_1\mathrm{vec}(\widetilde{\mathcal{Y}}^l),\\
\mathrm{vec}(\mathcal{R}_2^l)&=-\mathrm{vec}(\widetilde{\mathcal{Y}}^l\times_1 T_1+\widetilde{\mathcal{Y}}^l\times_2 T_2+\widetilde{\mathcal{Y}}^l\times_3 T_3)\\
&=-(I_{L_3}\otimes I_{L_2}\otimes T_1+I_{L_3}\otimes T_2\otimes I_{L_1}+T_3\otimes I_{L_2}\otimes I_{L_1})\mathrm{vec}(\widetilde{\mathcal{Y}}^l)\\
&=-N_1\mathrm{vec}(\widetilde{\mathcal{Y}}^l).
\end{aligned}
$$

Thus it follows from (4.6) that

$$
\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1})&=\mathrm{vec}(\widetilde{\mathcal{Y}}^l)+\frac{\gamma}{6}[\overline{M}_1^T\mathrm{vec}(\mathcal{R}_1^l)+\overline{N}_1^T\mathrm{vec}(\mathcal{R}_2^l)]\\
&=\mathrm{vec}(\widetilde{\mathcal{Y}}^l)-\frac{\gamma}{6}(\overline{M}_1^TM_1+\overline{N}_1^TN_1)\mathrm{vec}(\widetilde{\mathcal{Y}}^l)\\
&=\mathrm{vec}(\widetilde{\mathcal{Y}}^l)-\frac{\gamma}{6}\begin{pmatrix}\overline{M}_1^T & \overline{N}_1^T\end{pmatrix}\begin{pmatrix}M_1\\ N_1\end{pmatrix}\mathrm{vec}(\widetilde{\mathcal{Y}}^l)\\
&:=\mathrm{vec}(\widetilde{\mathcal{Y}}^l)-\frac{\gamma}{6}K_1^TK\mathrm{vec}(\widetilde{\mathcal{Y}}^l)\\
&=(I_{L_1L_2L_3}-\frac{\gamma}{6}K_1^TK)\mathrm{vec}(\widetilde{\mathcal{Y}}^l),
\end{aligned}
$$

where $K_1=\begin{pmatrix}\overline{M}_1\\ \overline{N}_1\end{pmatrix}$ and $K=\begin{pmatrix}M_1\\ N_1\end{pmatrix}$. Hence $I_{L_1L_2L_3}-\frac{\gamma}{6}K_1^TK$ is the iteration matrix of the DGI algorithm. Similar to the derivation of Theorem 3.2, we can obtain the conclusions of this theorem. $\square$

Theorem 4.3. *Assume that the assumptions in Theorem 4.2 are valid, then the quasi-optimal parameter γ_{opt} and the corresponding quasi-optimal convergence factor of the DGI algorithm are*
(1) If $\overline{Im}_1\geq\sqrt{\overline{Re}_{\max}\overline{Re}_{\min}}$, then

$$
\gamma_{opt}=\frac{6\overline{Re}_{\min}}{\overline{Re}_{\min}^2+\overline{Im}_1^2},\quad \rho_{opt}(I_{L_1L_2L_3}-\frac{1}{6}\gamma K_1^TK)=\frac{\overline{Im}_1}{\sqrt{\overline{Re}_{\min}^2+\overline{Im}_1^2}}.
$$

(2) If $\overline{Im}_1\leq\sqrt{\overline{Re}_{\max}\overline{Re}_{\min}}$, there are two cases:
(i) When $0<\gamma\leq\frac{12}{\overline{Re}_{\max}+\overline{Re}_{\min}}$, if $\overline{Im}_1^2<\frac{\overline{Re}_{\min}(\overline{Re}_{\max}-\overline{Re}_{\min})}{2}$, then $\gamma_{opt}=\frac{12}{\overline{Re}_{\max}+\overline{Re}_{\min}}$ and $\rho_{opt}(I_{L_1L_2L_3}-\frac{1}{6}\gamma K_1^TK)=\frac{\sqrt{(\overline{Re}_{\max}-\overline{Re}_{\min})^2+4\overline{Im}_1^2}}{\overline{Re}_{\max}+\overline{Re}_{\min}}$. If $\overline{Im}_1^2\geq\frac{\overline{Re}_{\min}(\overline{Re}_{\max}-\overline{Re}_{\min})}{2}$, then $\gamma_{opt}=\frac{6\overline{Re}_{\min}}{\overline{Re}_{\min}^2+\overline{Im}_1^2}$ and

$$
\rho_{opt}(I_{L_1L_2L_3}-\frac{1}{6}\gamma K_1^TK)=\frac{\overline{Im}_1}{\sqrt{\overline{Re}_{\min}^2+\overline{Im}_1^2}}.
$$

(ii) When $\frac{12}{\overline{Re}_{\max}+\overline{Re}_{\min}} \le \gamma < \frac{12\overline{Re}_{\max}}{\overline{Re}_{\max}+\overline{Im}_1^2}$, *then* $\gamma_{opt} = \frac{12}{\overline{Re}_{\max}+\overline{Re}_{\min}}$ *and*

$$\rho_{opt}(I_{L_1L_2L_3} - \frac{1}{6}\gamma K_1^T K) = \frac{\sqrt{(\overline{Re}_{\max} - \overline{Re}_{\min})^2 + 4\overline{Im}_1^2}}{\overline{Re}_{\max} + \overline{Re}_{\min}}.$$

Proof. The proof is similar to that of Theorem 3.3, hence it is omitted. □

By applying the new update strategy to the DGI algorithm, we can construct the new MGI (NMGI) algorithm for the coupled Sylvester tensor equation (4.1) as follows.

Algorithm 4.3. The new MGI (NMGI) algorithm:

Step 1. Given matrices $V_1, T_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3 \in \mathbb{R}^{L_3\times L_3}$, two tensors $\mathcal{W}_1, \mathcal{W}_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and positive constants γ, η. Choose the initial tensor $\mathcal{Y}^0$, and set $l = 0$.

Step 2. If $\tau_l = \frac{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^l\times_1 V_1-\mathcal{Y}^l\times_2 V_2-\mathcal{Y}^l\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^l\times_1 T_1-\mathcal{Y}^l\times_2 T_2-\mathcal{Y}^l\times_3 T_3\|^2}}{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^0\times_1 V_1-\mathcal{Y}^0\times_2 V_2-\mathcal{Y}^0\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^0\times_1 T_1-\mathcal{Y}^0\times_2 T_2-\mathcal{Y}^0\times_3 T_3\|^2}}$ $< \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ by the following procedure

$$\begin{aligned}
\mathcal{Y}_1^{l+1} &= \mathcal{Y}^l + \gamma\mathcal{R}_1^l\times_1 D_1^T,\\
\overrightarrow{\mathcal{Y}}^l &= \frac{\mathcal{Y}_1^{l+1} + 5\mathcal{Y}^l}{6},\\
\mathcal{Y}_2^{l+1} &= \overrightarrow{\mathcal{Y}}^l + \gamma\overrightarrow{\mathcal{R}}_1^l\times_2 D_2^T,\\
\overline{\mathcal{Y}}^l &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + 4\mathcal{Y}^l}{6},\\
\mathcal{Y}_3^{l+1} &= \overline{\mathcal{Y}}^l + \gamma\overline{\mathcal{R}}_1^l\times_3 D_3^T,\\
\hat{\mathcal{Y}}^l &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + 3\mathcal{Y}^l}{6},\\
\mathcal{Y}_4^{l+1} &= \hat{\mathcal{Y}}^l + \gamma\hat{\mathcal{R}}_2^l\times_1 D_{11}^T,\\
\check{\mathcal{Y}}^l &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + 2\mathcal{Y}^l}{6},\\
\mathcal{Y}_5^{l+1} &= \check{\mathcal{Y}}^l + \gamma\check{\mathcal{R}}_2^l\times_2 D_{22}^T,\\
\acute{\mathcal{Y}}^l &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}^l}{6},\\
\mathcal{Y}_6^{l+1} &= \acute{\mathcal{Y}}^l + \gamma\acute{\mathcal{R}}_2^l\times_3 D_{33}^T,\\
\mathcal{Y}^{l+1} &= \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}_6^{l+1}}{6}.
\end{aligned}$$

Step 4. Set $l := l + 1$ and return to Step 2.

In the following theorem, we give the convergence analysis of the NMGI algorithm.

Theorem 4.4. *The NMGI algorithm is convergent if the parameter* γ *is selected to satisfy* $w_1 + w_2u_2 + w_3u_3 + w_4u_4 + w_5u_5 + w_6u_6 < 6$, *where*

$$w_1 = \|I_{L_1} - \gamma D_1^T V_1\|_2 + \gamma\|D_1\|_2\|V_2\|_2 + \gamma\|D_1\|_2\|V_3\|_2,$$

$$
\begin{aligned}
&w_2 = \|I_{L_2} - \gamma D_2^T V_2\|_2 + \gamma\|V_1\|_2\|D_2\|_2 + \gamma\|D_2\|_2\|V_3\|_2,\\
&w_3 = \|I_{L_3} - \gamma D_3^T V_3\|_2 + \gamma\|V_1\|_2\|D_3\|_2 + \gamma\|V_2\|_2\|D_3\|_2,\\
&w_4 = \|I_{L_1} - \gamma D_{11}^T T_1\|_2 + \gamma\|D_{11}\|_2\| T_2\|_2 + \gamma\|D_{11}\|_2\| T_3\|_2,\\
&w_5 = \|I_{L_2} - \gamma D_{22}^T T_2\|_2 + \gamma\|T_1\|_2\|D_{22}\|_2 + \gamma\|D_{22}\|_2\|T_3\|_2,\\
&w_6 = \|I_{L_3} - \gamma D_{33}^T T_3\|_2 + \gamma\|T_1\|_2\|D_{33}\|_2 + \gamma\|T_2\|_2\|D_{33}\|_2,\\
&u_2 = \frac{w_1}{6} + \frac{5}{6},\ u_3 = \frac{w_1}{6} + \frac{w_2u_2}{6} + \frac{2}{3},\ u_4 = \frac{w_1}{6} + \frac{w_2u_2}{6} + \frac{w_3u_3}{6} + \frac{1}{2},\\
&u_5 = \frac{w_1}{6} + \frac{w_2u_2}{6} + \frac{w_3u_3}{6} + \frac{w_4u_4}{6} + \frac{1}{3},\ u_6 = \frac{w_1}{6} + \frac{w_2u_2}{6} + \frac{w_3u_3}{6} + \frac{w_4u_4}{6} + \frac{w_5u_5}{6} + \frac{1}{6}.
\end{aligned}
$$

Proof. We define the following error tensors

$$
\begin{aligned}
&\widetilde{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*,\ \widetilde{\overrightarrow{\mathcal{Y}}}^l = \overrightarrow{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\overline{\mathcal{Y}}}^l = \overline{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\hat{\mathcal{Y}}}^l = \hat{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\check{\mathcal{Y}}}^l = \check{\mathcal{Y}}^l - \mathcal{Y}^*,\\
&\widetilde{\acute{\mathcal{Y}}}^l = \acute{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\mathcal{Y}}_i^{l+1} = \mathcal{Y}_i^{l+1} - \mathcal{Y}^*\ (i = 1, \cdots, 6).
\end{aligned}
$$

According to Line 1 of Step 3 of the NMGI algorithm, it follows that

$$
\begin{aligned}
\widetilde{\mathcal{Y}}_1^{l+1} &= \widetilde{\mathcal{Y}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3) \times_1 D_1^T\\
&= \widetilde{\mathcal{Y}}^l \times_1 (I_{L_1} - \gamma D_1^T V_1) - \gamma\widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_2 V_2 - \gamma\widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_3 V_3. \qquad (4.7)
\end{aligned}
$$

Similarly, from the expressions of $\mathcal{Y}_2^{l+1}$, $\mathcal{Y}_3^{l+1}$, $\mathcal{Y}_4^{l+1}$, $\mathcal{Y}_5^{l+1}$ and $\mathcal{Y}_6^{l+1}$, we have

$$
\begin{aligned}
\widetilde{\mathcal{Y}}_2^{l+1} &= \widetilde{\overrightarrow{\mathcal{Y}}}^l \times_2 (I_{L_2} - \gamma D_2^T V_2) - \gamma\widetilde{\overrightarrow{\mathcal{Y}}}^l \times_1 V_1 \times_2 D_2^T - \gamma\widetilde{\overrightarrow{\mathcal{Y}}}^l \times_2 D_2^T \times_3 V_3,\\
\widetilde{\mathcal{Y}}_3^{l+1} &= \widetilde{\overline{\mathcal{Y}}}^l \times_3 (I_{L_3} - \gamma D_3^T V_3) - \gamma\widetilde{\overline{\mathcal{Y}}}^l \times_1 V_1 \times_3 D_3^T - \gamma\widetilde{\overline{\mathcal{Y}}}^l \times_2 V_2 \times_3 D_3^T,\\
\widetilde{\mathcal{Y}}_4^{l+1} &= \widetilde{\hat{\mathcal{Y}}}^l \times_1 (I_{L_1} - \gamma D_{11}^T T_1) - \gamma\widetilde{\hat{\mathcal{Y}}}^l \times_1 D_{11}^T \times_2 T_2 - \gamma\widetilde{\hat{\mathcal{Y}}}^l \times_1 D_{11}^T \times_3 T_3,\\
\widetilde{\mathcal{Y}}_5^{l+1} &= \widetilde{\check{\mathcal{Y}}}^l \times_2 (I_{L_2} - \gamma D_{22}^T T_2) - \gamma\widetilde{\check{\mathcal{Y}}}^l \times_1 T_1 \times_2 D_{22}^T - \gamma\widetilde{\check{\mathcal{Y}}}^l \times_2 D_{22}^T \times_3 T_3,\\
\widetilde{\mathcal{Y}}_6^{l+1} &= \widetilde{\acute{\mathcal{Y}}}^l \times_3 (I_{L_3} - \gamma D_{33}^T T_3) - \gamma\widetilde{\acute{\mathcal{Y}}}^l \times_1 T_1 \times_3 D_{33}^T - \gamma\widetilde{\acute{\mathcal{Y}}}^l \times_2 T_2 \times_3 D_{33}^T,
\end{aligned}
$$

from which one can deduce that

$$
\begin{aligned}
&\|\widetilde{\mathcal{Y}}_1^{l+1}\| \leq \|\widetilde{\mathcal{Y}}^l\|(\|I_{L_1} - \gamma D_1^T V_1\|_2 + \gamma\|D_1\|_2\| V_2\|_2 + \gamma\|D_1\|_2\| V_3\|_2) := w_1\|\widetilde{\mathcal{Y}}^l\|,\\
&\|\widetilde{\mathcal{Y}}_2^{l+1}\| \leq \|\widetilde{\overrightarrow{\mathcal{Y}}}^l\|(\|I_{L_2} - \gamma D_2^T V_2\|_2 + \gamma\|V_1\|_2\|D_2\|_2 + \gamma\|D_2\|_2\|V_3\|_2) := w_2\|\widetilde{\overrightarrow{\mathcal{Y}}}^l\|,\\
&\|\widetilde{\mathcal{Y}}_3^{l+1}\| \leq \|\widetilde{\overline{\mathcal{Y}}}^l\|(\|I_{L_3} - \gamma D_3^T V_3\|_2 + \gamma\|V_1\|_2\|D_3\|_2 + \gamma\|V_2\|_2\|D_3\|_2) := w_3\|\widetilde{\overline{\mathcal{Y}}}^l\|,\\
&\|\widetilde{\mathcal{Y}}_4^{l+1}\| \leq \|\widetilde{\hat{\mathcal{Y}}}^l\|(\|I_{L_1} - \gamma D_{11}^T T_1\|_2 + \gamma\|D_{11}\|_2\| T_2\|_2 + \gamma\|D_{11}\|_2\| T_3\|_2) := w_4\|\widetilde{\hat{\mathcal{Y}}}^l\|,\\
&\|\widetilde{\mathcal{Y}}_5^{l+1}\| \leq \|\widetilde{\check{\mathcal{Y}}}^l\|(\|I_{L_2} - \gamma D_{22}^T T_2\|_2 + \gamma\|T_1\|_2\|D_{22}\|_2 + \gamma\|D_{22}\|_2\|T_3\|_2) := w_5\|\widetilde{\check{\mathcal{Y}}}^l\|,\\
&\|\widetilde{\mathcal{Y}}_6^{l+1}\| \leq \|\widetilde{\acute{\mathcal{Y}}}^l\|(\|I_{L_3} - \gamma D_{33}^T T_3\|_2 + \gamma\|T_1\|_2\|D_{33}\|_2 + \gamma\|T_2\|_2\|D_{33}\|_2) := w_6\|\widetilde{\acute{\mathcal{Y}}}^l\|, \qquad (4.8)
\end{aligned}
$$

in terms of Lemma 2.1. By making use of Lemma 2.2, it holds that

$$
\|\widetilde{\overrightarrow{\mathcal{Y}}}^l\| \leq \frac{\|\widetilde{\mathcal{Y}}_1^{l+1}\| + 5\|\widetilde{\mathcal{Y}}^l\|}{6} \leq (\frac{w_1}{6} + \frac{5}{6})\|\widetilde{\mathcal{Y}}^l\| := u_2\|\widetilde{\mathcal{Y}}^l\|,
$$

$$
\begin{aligned}
\|\widetilde{\bar{\mathcal{Y}}}^{l}\| &\leq \frac{\|\widetilde{\mathcal{Y}}_1^{l+1}\| + \|\widetilde{\mathcal{Y}}_2^{l+1}\| + 4\|\widetilde{\mathcal{Y}}^{l}\|}{6} \leq (\frac{w_1}{6} + \frac{w_2 u_2}{6} + \frac{2}{3})\|\widetilde{\mathcal{Y}}^{l}\| := u_3\|\widetilde{\mathcal{Y}}^{l}\|,\\
\|\widetilde{\hat{\mathcal{Y}}}^{l}\| &\leq \frac{\|\widetilde{\mathcal{Y}}_1^{l+1}\| + \|\widetilde{\mathcal{Y}}_2^{l+1}\| + \|\widetilde{\mathcal{Y}}_3^{l+1}\| + 3\|\widetilde{\mathcal{Y}}^{l}\|}{6} \leq (\frac{w_1}{6} + \frac{w_2 u_2}{6} + \frac{w_3 u_3}{6} + \frac{1}{2})\|\widetilde{\mathcal{Y}}^{l}\| := u_4\|\widetilde{\mathcal{Y}}^{l}\|,\\
\|\widetilde{\check{\mathcal{Y}}}^{l}\| &\leq \frac{\|\widetilde{\mathcal{Y}}_1^{l+1}\| + \|\widetilde{\mathcal{Y}}_2^{l+1}\| + \|\widetilde{\mathcal{Y}}_3^{l+1}\| + \|\widetilde{\mathcal{Y}}_4^{l+1}\| + 2\|\widetilde{\mathcal{Y}}^{l}\|}{6}\\
&\leq (\frac{w_1}{6} + \frac{w_2 u_2}{6} + \frac{w_3 u_3}{6} + \frac{w_4 u_4}{6} + \frac{1}{3})\|\widetilde{\mathcal{Y}}^{l}\|\\
&:= u_5\|\widetilde{\mathcal{Y}}^{l}\|,\\
\|\widetilde{\acute{\mathcal{Y}}}^{l}\| &\leq \frac{\|\widetilde{\mathcal{Y}}_1^{l+1}\| + \|\widetilde{\mathcal{Y}}_2^{l+1}\| + \|\widetilde{\mathcal{Y}}_3^{l+1}\| + \|\widetilde{\mathcal{Y}}_4^{l+1}\| + \|\widetilde{\mathcal{Y}}_5^{l+1}\| + \|\widetilde{\mathcal{Y}}^{l}\|}{6}\\
&\leq (\frac{w_1}{6} + \frac{w_2 u_2}{6} + \frac{w_3 u_3}{6} + \frac{w_4 u_4}{6} + \frac{w_5 u_5}{6} + \frac{1}{6})\|\widetilde{\mathcal{Y}}^{l}\|\\
&:= u_6\|\widetilde{\mathcal{Y}}^{l}\|.
\end{aligned}
\tag{4.9}
$$

Therefore, it is obtained from (4.8) and (4.9) that

$$
\begin{aligned}
\|\widetilde{\mathcal{Y}}^{l+1}\| &\leq \frac{\|\widetilde{\mathcal{Y}}_1^{l+1}\| + \|\widetilde{\mathcal{Y}}_2^{l+1}\| + \|\widetilde{\mathcal{Y}}_3^{l+1}\| + \|\widetilde{\mathcal{Y}}_4^{l+1}\| + \|\widetilde{\mathcal{Y}}_5^{l+1}\| + \|\widetilde{\mathcal{Y}}_6^{l+1}\|}{6}\\
&\leq (\frac{w_1}{6} + \frac{w_2 u_2}{6} + \frac{w_3 u_3}{6} + \frac{w_4 u_4}{6} + \frac{w_5 u_5}{6} + \frac{w_6 u_6}{6})\|\widetilde{\mathcal{Y}}^{l}\|.
\end{aligned}
$$

This implies that if $\frac{w_1}{6} + \frac{w_2u_2}{6} + \frac{w_3u_3}{6} + \frac{w_4u_4}{6} + \frac{w_5u_5}{6} + \frac{w_6u_6}{6} < 1$, i.e., $w_1 + w_2u_2 + w_3u_3 + w_4u_4 + w_5u_5 + w_6u_6 < 6$, then the NMGI algorithm is convergent. $\square$

5. The DGI and NMGI algorithms for the coupled Sylvester tensor equations with two unknowns

In this section, we extend the DGI and NMGI algorithms to solve the coupled Sylvester tensor equations with two unknowns, whose form is as below

$$
\begin{cases}
\mathcal{Y} \times_1 V_1 + \mathcal{Y} \times_2 V_2 + \mathcal{Y} \times_3 V_3 + \mathcal{Z} \times_1 P_1 + \mathcal{Z} \times_2 P_2 + \mathcal{Z} \times_3 P_3 = \mathcal{W}_1,\\
\mathcal{Y} \times_1 T_1 + \mathcal{Y} \times_2 T_2 + \mathcal{Y} \times_3 T_3 + \mathcal{Z} \times_1 Q_1 + \mathcal{Z} \times_2 Q_2 + \mathcal{Z} \times_3 Q_3 = \mathcal{W}_2,
\end{cases}
\tag{5.1}
$$

where $V_1, T_1, P_1, Q_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2, P_2, Q_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3, P_3, Q_3 \in \mathbb{R}^{L_3\times L_3}$, $\mathcal{W}_1, W_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and the unknown tensors $\mathcal{Y}, \mathcal{Z} \in \mathbb{R}^{L_1\times L_2\times L_3}$ need to be computed.

By making use of the hierarchical identification principle, we can construct the GI algorithm for the coupled Sylvester tensor equation with two unknowns (5.1) as follows.

Algorithm 5.1. The gradient-based iterative (GI) algorithm:

Step 1. Given matrices $V_1, T_1, P_1, Q_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2, P_2, Q_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3, P_3, Q_3 \in \mathbb{R}^{L_3\times L_3}$, two tensors $\mathcal{W}_1, \mathcal{W}_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and positive constants γ, η. Choose the initial tensors $\mathcal{Y}^0, \mathcal{Z}^0$, and set $l = 0$.

Step 2. If $\tau_l = \frac{\sqrt{\|\mathcal{W}_1 - \sum_{j=1}^{3}\mathcal{Y}^l\times_j V_j - \sum_{j=1}^{3}\mathcal{Z}^l\times_j P_j\|^2 + \|\mathcal{W}_2 - \sum_{j=1}^{3}\mathcal{Y}^l\times_j T_j - \sum_{j=1}^{3}\mathcal{Z}^l\times_j Q_j\|^2}}{\sqrt{\|\mathcal{W}_1 - \sum_{j=1}^{3}\mathcal{Y}^0\times_j V_j - \sum_{j=1}^{3}\mathcal{Z}^0\times_j P_j\|^2 + \|\mathcal{W}_2 - \sum_{j=1}^{3}\mathcal{Y}^0\times_j T_j - \sum_{j=1}^{3}\mathcal{Z}^0\times_j Q_j\|^2}} < \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ and $\mathcal{Z}^{l+1}$ by the following procedure

$$\begin{aligned}
&\mathcal{R}_1^l = \mathcal{W}_1 - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3 - \mathcal{Z}^l \times_1 P_1 - \mathcal{Z}^l \times_2 P_2 - \mathcal{Z}^l \times_3 P_3,\\
&\mathcal{R}_2^l = \mathcal{W}_2 - \mathcal{Y}^l \times_1 T_1 - \mathcal{Y}^l \times_2 T_2 - \mathcal{Y}^l \times_3 T_3 - \mathcal{Z}^l \times_1 Q_1 - \mathcal{Z}^l \times_2 Q_2 - \mathcal{Z}^l \times_3 Q_3,\\
&\mathcal{Y}_1^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_1^l \times_1 V_1^T,\ \mathcal{Z}_1^{l+1} = \mathcal{Z}^l + \gamma \mathcal{R}_1^l \times_1 P_1^T,\\
&\mathcal{Y}_2^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_1^l \times_2 V_2^T,\ \mathcal{Z}_2^{l+1} = \mathcal{Z}^l + \gamma \mathcal{R}_1^l \times_2 P_2^T,\\
&\mathcal{Y}_3^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_1^l \times_3 V_3^T,\ \mathcal{Z}_3^{l+1} = \mathcal{Z}^l + \gamma \mathcal{R}_1^l \times_3 P_3^T,\\
&\mathcal{Y}_4^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_2^l \times_1 T_1^T,\ \mathcal{Z}_4^{l+1} = \mathcal{Z}^l + \gamma \mathcal{R}_2^l \times_1 Q_1^T,\\
&\mathcal{Y}_5^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_2^l \times_2 T_2^T,\ \mathcal{Z}_5^{l+1} = \mathcal{Z}^l + \gamma \mathcal{R}_2^l \times_2 Q_2^T,\\
&\mathcal{Y}_6^{l+1} = \mathcal{Y}^l + \gamma \mathcal{R}_2^l \times_3 T_3^T,\ \mathcal{Z}_6^{l+1} = \mathcal{Z}^l + \gamma \mathcal{R}_2^l \times_3 Q_3^T,\\
&\mathcal{Y}^{l+1} = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}_6^{l+1}}{6},\\
&\mathcal{Z}^{l+1} = \frac{\mathcal{Z}_1^{l+1} + \mathcal{Z}_2^{l+1} + \mathcal{Z}_3^{l+1} + \mathcal{Z}_4^{l+1} + \mathcal{Z}_5^{l+1} + \mathcal{Z}_6^{l+1}}{6}.
\end{aligned}$$

Step 4. Set $l := l + 1$ and return to Step 2.

We will establish the convergence theorem of the GI algorithm for the coupled Sylvester tensor equation with two unknowns (5.1).

Theorem 5.1. *Let* $H_1 = I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1}$, $H_2 = I_{L_3} \otimes I_{L_2} \otimes T_1 + I_{L_3} \otimes T_2 \otimes I_{L_1} + T_3 \otimes I_{L_2} \otimes I_{L_1}$, $H_3 = I_{L_3} \otimes I_{L_2} \otimes P_1 + I_{L_3} \otimes P_2 \otimes I_{L_1} + P_3 \otimes I_{L_2} \otimes I_{L_1}$, $H_4 = I_{L_3} \otimes I_{L_2} \otimes Q_1 + I_{L_3} \otimes Q_2 \otimes I_{L_1} + Q_3 \otimes I_{L_2} \otimes I_{L_1}$ *and* $U = \begin{pmatrix} H_1 & H_3 \\ H_2 & H_4 \end{pmatrix}$. *And let* $\widehat{\lambda}_{\max}$ *and* $\widehat{\lambda}_{\min}$ *be the maximum and minimum eigenvalues of the matrix* $U^T U$, *respectively. Assume that the coupled Sylvester tensor equation with two unknowns* (5.1) *has a unique solution* $(\mathcal{Y}^*, \mathcal{Z}^*)$. *Then the GI algorithm is convergent if and only if*

$$0 < \gamma < \frac{12}{\widehat{\lambda}_{\max}}.$$

And the optimal parameter γ_{opt} *and the corresponding optimal convergence factor* ρ_{opt} *of the GI algorithm are*

$$\gamma_{opt} = \frac{12}{\widehat{\lambda}_{\max} + \widehat{\lambda}_{\min}}, \quad \rho_{opt} = \frac{\widehat{\lambda}_{\max} - \widehat{\lambda}_{\min}}{\widehat{\lambda}_{\max} + \widehat{\lambda}_{\min}}.$$

Proof. We first define the error tensors $\widetilde{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*$ and $\widetilde{\mathcal{Z}}^l = \mathcal{Z}^l - \mathcal{Z}^*$. It follows from the framework of the GI algorithm that

$$\begin{aligned}
\widetilde{\mathcal{Y}}^{l+1} &= \widetilde{\mathcal{Y}}^l + \frac{\gamma}{6}(\mathcal{R}_1^l \times_1 V_1^T + \mathcal{R}_1^l \times_2 V_2^T + \mathcal{R}_1^l \times_3 V_3^T + \mathcal{R}_2^l \times_1 T_1^T\\
&\quad + \mathcal{R}_2^l \times_2 T_2^T + \mathcal{R}_2^l \times_3 T_3^T),\\
\widetilde{\mathcal{Z}}^{l+1} &= \widetilde{\mathcal{Z}}^l + \frac{\gamma}{6}(\mathcal{R}_1^l \times_1 P_1^T + \mathcal{R}_1^l \times_2 P_2^T + \mathcal{R}_1^l \times_3 P_3^T + \mathcal{R}_2^l \times_1 Q_1^T
\end{aligned}$$

$$+\mathcal{R}_2^l \times_2 Q_2^T + \mathcal{R}_2^l \times_3 Q_3^T). \tag{5.2}$$

Taking straightening operator into (5.2) yields that

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) &= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + \frac{\gamma}{6}[(I_{L_3} \otimes I_{L_2} \otimes V_1^T + I_{L_3} \otimes V_2^T \otimes I_{L_1} + V_3^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_1^l) \\
&\quad +(I_{L_3} \otimes I_{L_2} \otimes T_1^T + I_{L_3} \otimes T_2^T \otimes I_{L_1} + T_3^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_2^l)] \\
&= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + \frac{\gamma}{6}[H_1^T \mathrm{vec}(\mathcal{R}_1^l) + H_2^T \mathrm{vec}(\mathcal{R}_2^l)], \\
\mathrm{vec}(\widetilde{\mathcal{Z}}^{l+1}) &= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) + \frac{\gamma}{6}[(I_{L_3} \otimes I_{L_2} \otimes P_1^T + I_{L_3} \otimes P_2^T \otimes I_{L_1} + P_3^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_1^l) \\
&\quad +(I_{L_3} \otimes I_{L_2} \otimes Q_1^T + I_{L_3} \otimes Q_2^T \otimes I_{L_1} + Q_3^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_2^l)] \\
&= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) + \frac{\gamma}{6}[H_3^T \mathrm{vec}(\mathcal{R}_1^l) + H_4^T \mathrm{vec}(\mathcal{R}_2^l)].
\end{aligned} \tag{5.3}$$

By some computations, it holds that

$$\begin{aligned}
\mathrm{vec}(\mathcal{R}_1^l) &= -\mathrm{vec}(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3 + \widetilde{\mathcal{Z}}^l \times_1 P_1 + \widetilde{\mathcal{Z}}^l \times_2 P_2 + \widetilde{\mathcal{Z}}^l \times_3 P_3) \\
&= -(I_{L_3} \otimes I_{L_2} \otimes V_1 + I_{L_3} \otimes V_2 \otimes I_{L_1} + V_3 \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&\quad -(I_{L_3} \otimes I_{L_2} \otimes P_1 + I_{L_3} \otimes P_2 \otimes I_{L_1} + P_3 \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\widetilde{\mathcal{Z}}^l) \\
&= -H_1 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) - H_3 \mathrm{vec}(\widetilde{\mathcal{Z}}^l), \\
\mathrm{vec}(\mathcal{R}_2^l) &= -\mathrm{vec}(\widetilde{\mathcal{Y}}^l \times_1 T_1 + \widetilde{\mathcal{Y}}^l \times_2 T_2 + \widetilde{\mathcal{Y}}^l \times_3 T_3 + \widetilde{\mathcal{Z}}^l \times_1 Q_1 + \widetilde{\mathcal{Z}}^l \times_2 Q_2 + \widetilde{\mathcal{Z}}^l \times_3 Q_3) \\
&= -(I_{L_3} \otimes I_{L_2} \otimes T_1 + I_{L_3} \otimes T_2 \otimes I_{L_1} + T_3 \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\
&\quad -(I_{L_3} \otimes I_{L_2} \otimes Q_1 + I_{L_3} \otimes Q_2 \otimes I_{L_1} + Q_3 \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\widetilde{\mathcal{Z}}^l) \\
&= -H_2 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) - H_4 \mathrm{vec}(\widetilde{\mathcal{Z}}^l).
\end{aligned} \tag{5.4}$$

Substituting (5.4) into (5.3) results in

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) &= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + \frac{\gamma}{6}[H_1^T \mathrm{vec}(\mathcal{R}_1^l) + H_2^T \mathrm{vec}(\mathcal{R}_2^l)] \\
&= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}\{H_1^T[H_1 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_3 \mathrm{vec}(\widetilde{\mathcal{Z}}^l)] + H_2^T[H_2 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_4 \mathrm{vec}(\widetilde{\mathcal{Z}}^l)]\} \\
&= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}\{(H_1^T H_1 + H_2^T H_2)\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + (H_1^T H_3 + H_2^T H_4)\mathrm{vec}(\widetilde{\mathcal{Z}}^l)\}, \\
\mathrm{vec}(\widetilde{\mathcal{Z}}^{l+1}) &= \mathrm{vec}(\tilde{\mathcal{Z}}^l) + \frac{\gamma}{6}[H_3^T \mathrm{vec}(\mathcal{R}_1^l) + H_4^T \mathrm{vec}(\mathcal{R}_2^l)] \\
&= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) - \frac{\gamma}{6}\{H_3^T[H_1 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_3 \mathrm{vec}(\widetilde{\mathcal{Z}}^l)] + H_4^T[H_2 \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_4 \mathrm{vec}(\widetilde{\mathcal{Z}}^l)]\} \\
&= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) - \frac{\gamma}{6}\{(H_3^T H_1 + H_4^T H_2)\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + (H_3^T H_3 + H_4^T H_4)\mathrm{vec}(\widetilde{\mathcal{Z}}^l)\},
\end{aligned}$$

which leads to

$$\begin{aligned}
\begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) \\ \mathrm{vec}(\widetilde{\mathcal{Z}}^{l+1}) \end{pmatrix} &= \begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ \mathrm{vec}(\tilde{\mathcal{Z}}^l) \end{pmatrix} - \frac{\gamma}{6} \begin{pmatrix} H_1^T H_1 + H_2^T H_2 & H_1^T H_3 + H_2^T H_4 \\ H_3^T H_1 + H_4^T H_2 & H_3^T H_3 + H_4^T H_4 \end{pmatrix} \begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ \mathrm{vec}(\tilde{\mathcal{Z}}^l) \end{pmatrix} \\
&= \left[I_{2L_1L_2L_3} - \frac{\gamma}{6} \begin{pmatrix} H_1 & H_3 \\ H_2 & H_4 \end{pmatrix}^T \begin{pmatrix} H_1 & H_3 \\ H_2 & H_4 \end{pmatrix} \right] \begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ \mathrm{vec}(\tilde{\mathcal{Z}}^l) \end{pmatrix}
\end{aligned}$$

$$:= (I_{2L_1L_2L_3} - \frac{\gamma}{6}U^TU)\begin{pmatrix}\mathrm{vec}(\widetilde{\mathcal{Y}}^l)\\ \mathrm{vec}(\tilde{\mathcal{Z}}^l)\end{pmatrix},$$

where $U = \begin{pmatrix} H_1 & H_3 \\ H_2 & H_4 \end{pmatrix}$. Therefore $I_{2L_1L_2L_3} - \frac{\gamma}{6}U^TU$ is the iteration matrix of the GI algorithm. By assumptions, (5.1) has a unique solution, then the matrix U is of full column rank and hence U^TU is a symmetric positive-definite matrix. Let $\widehat{\lambda}_i > 0$ $(i = 1, 2, \cdots, 2L_1L_2L_3)$ be the eigenvalues of U^TU, then by utilizing the similar methods applied in Theorems 3.1 and 4.1, we can conclude that the GI algorithm is convergent if and only if $0 < \gamma < \frac{12}{\widehat{\lambda}_{\max}}$. And we can derive the optimal parameter and the corresponding optimal convergence factor of the GI algorithm are $\gamma_{\text{opt}} = \frac{12}{\widehat{\lambda}_{\min}+\widehat{\lambda}_{\max}}$ and $\rho_{\text{opt}} = \frac{\widehat{\lambda}_{\max}-\widehat{\lambda}_{\min}}{\widehat{\lambda}_{\min}+\widehat{\lambda}_{\max}}$, respectively. □

Let D_j, D_{jj}, D_{jjj} and D_{jjjj} be the diagonal parts of the matrices V_j, T_j, P_j and Q_j $(j = 1, 2, 3)$, respectively. In order to reduce the computation of the GI algorithm, we replace the system matrices in the GI algorithm by their diagonal parts, and establish the following diagonal GI (DGI) algorithm for the coupled Sylvester tensor equation with two unknowns (5.1).

Algorithm 5.2. The diagonal GI (DGI) algorithm:

Step 1. Given matrices $V_1, T_1, P_1, Q_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2, P_2, Q_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3, P_3, Q_3 \in \mathbb{R}^{L_3\times L_3}$, two tensors $\mathcal{W}_1, \mathcal{W}_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and positive constants γ, η. Choose the initial tensors $\mathcal{Y}^0$, $\mathcal{Z}^0$, and set $l = 0$.

Step 2. If $\tau_l = \frac{\sqrt{\|\mathcal{W}_1 - \sum_{j=1}^{3}\mathcal{Y}^l\times_j V_j - \sum_{j=1}^{3}\mathcal{Z}^l\times_j P_j\|^2 + \|\mathcal{W}_2 - \sum_{j=1}^{3}\mathcal{Y}^l\times_j T_j - \sum_{j=1}^{3}\mathcal{Z}^l\times_j Q_j\|^2}}{\sqrt{\|\mathcal{W}_1 - \sum_{j=1}^{3}\mathcal{Y}^0\times_j V_j - \sum_{j=1}^{3}\mathcal{Z}^0\times_j P_j\|^2 + \|\mathcal{W}_2 - \sum_{j=1}^{3}\mathcal{Y}^0\times_j T_j - \sum_{j=1}^{3}\mathcal{Z}^0\times_j Q_j\|^2}} < \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ and $\mathcal{Z}^{l+1}$ by the following procedure

$$\begin{aligned}
&\mathcal{R}_1^l = \mathcal{W}_1 - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3 - \mathcal{Z}^l \times_1 P_1 - \mathcal{Z}^l \times_2 P_2 - \mathcal{Z}^l \times_3 P_3,\\
&\mathcal{R}_2^l = \mathcal{W}_2 - \mathcal{Y}^l \times_1 T_1 - \mathcal{Y}^l \times_2 T_2 - \mathcal{Y}^l \times_3 T_3 - \mathcal{Z}^l \times_1 Q_1 - \mathcal{Z}^l \times_2 Q_2 - \mathcal{Z}^l \times_3 Q_3,\\
&\mathcal{Y}_1^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_1^l \times_1 D_1^T,\ \mathcal{Z}_1^{l+1} = \mathcal{Z}^l + \gamma\mathcal{R}_1^l \times_1 D_{111}^T,\\
&\mathcal{Y}_2^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_1^l \times_2 D_2^T,\ \mathcal{Z}_2^{l+1} = \mathcal{Z}^l + \gamma\mathcal{R}_1^l \times_2 D_{222}^T,\\
&\mathcal{Y}_3^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_1^l \times_3 D_3^T,\ \mathcal{Z}_3^{l+1} = \mathcal{Z}^l + \gamma\mathcal{R}_1^l \times_3 D_{333}^T,\\
&\mathcal{Y}_4^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_2^l \times_1 D_{11}^T,\ \mathcal{Z}_4^{l+1} = \mathcal{Z}^l + \gamma\mathcal{R}_2^l \times_1 D_{1111}^T,\\
&\mathcal{Y}_5^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_2^l \times_2 D_{22}^T,\ \mathcal{Z}_5^{l+1} = \mathcal{Z}^l + \gamma\mathcal{R}_2^l \times_2 D_{2222}^T,\\
&\mathcal{Y}_6^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_2^l \times_3 D_{33}^T,\ \mathcal{Z}_6^{l+1} = \mathcal{Z}^l + \gamma\mathcal{R}_2^l \times_3 D_{3333}^T,\\
&\mathcal{Y}^{l+1} = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}_6^{l+1}}{6},\\
&\mathcal{Z}^{l+1} = \frac{\mathcal{Z}_1^{l+1} + \mathcal{Z}_2^{l+1} + \mathcal{Z}_3^{l+1} + \mathcal{Z}_4^{l+1} + \mathcal{Z}_5^{l+1} + \mathcal{Z}_6^{l+1}}{6}.
\end{aligned}$$

Step 4. Set $l := l + 1$ and return to Step 2.

Below we investigate the convergence properties of the DGI algorithm for the coupled Sylvester tensor equation with two unknowns (5.1).

Theorem 5.2. *Suppose that the conditions of Theorem 5.1 are satisfied, and* $Re(\widehat{\delta}_q) > 0$ *with* $\widehat{\delta}_q$ $(q = 1, \cdots, 2L_1L_2L_3)$ *being the eigenvalues of* U_1^TU*, where* $U_1 = \begin{pmatrix} \bar{H}_1 & \bar{H}_3 \\ \bar{H}_2 & \bar{H}_4 \end{pmatrix}$, $U = \begin{pmatrix} H_1 & H_3 \\ H_2 & H_4 \end{pmatrix}$,

$$\begin{aligned}
\bar{H}_1 &= I_{L_3} \otimes I_{L_2} \otimes D_1 + I_{L_3} \otimes D_2 \otimes I_{L_1} + D_3 \otimes I_{L_2} \otimes I_{L_1}, \\
\bar{H}_2 &= I_{L_3} \otimes I_{L_2} \otimes D_{11} + I_{L_3} \otimes D_{22} \otimes I_{L_1} + D_{33} \otimes I_{L_2} \otimes I_{L_1}, \\
\bar{H}_3 &= I_{L_3} \otimes I_{L_2} \otimes D_{111} + I_{L_3} \otimes D_{222} \otimes I_{L_1} + D_{333} \otimes I_{L_2} \otimes I_{L_1}, \\
\bar{H}_4 &= I_{L_3} \otimes I_{L_2} \otimes D_{1111} + I_{L_3} \otimes D_{2222} \otimes I_{L_1} + D_{3333} \otimes I_{L_2} \otimes I_{L_1},
\end{aligned}$$

and H_i $(i = 1, 2, 3, 4)$ *are defined as in Theorem 5.1. Let* $\widehat{Im}_1 = \max\limits_{1\le q\le 2L_1L_2L_3} |Im(\widehat{\delta}_q)|$, $\widehat{Re}_{\max} = \max\limits_{1\le q\le 2L_1L_2L_3} \{Re(\widehat{\delta}_q)\}$ *and* $\widehat{Re}_{\min} = \min\limits_{1\le q\le 2L_1L_2L_3} \{Re(\widehat{\delta}_q)\}$*. Then the DGI algorithm is convergent if*

$$0 < \gamma < \begin{cases} \dfrac{12\widehat{Re}_{\min}}{\widehat{Re}_{\min}^2 + \widehat{Im}_1^2}, & as\ \widehat{Im}_1 \ge \sqrt{\widehat{Re}_{\max}\widehat{Re}_{\min}}\ , \\ \dfrac{12\widehat{Re}_{\max}}{\widehat{Re}_{\max}^2 + \widehat{Im}_1^2}, & as\ \widehat{Im}_1 \le \sqrt{\widehat{Re}_{\max}\widehat{Re}_{\min}}\ . \end{cases}$$

Proof. Define the error tensors $\widetilde{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*$ and $\widetilde{\mathcal{Z}}^l = \mathcal{Z}^l - \mathcal{Z}^*$. In view of the framework of the DGI algorithm, we deduce that

$$\begin{aligned}
\widetilde{\mathcal{Y}}^{l+1} &= \widetilde{\mathcal{Y}}^l + \frac{\gamma}{6}(\mathcal{R}_1^l \times_1 D_1^T + \mathcal{R}_1^l \times_2 D_2^T + \mathcal{R}_1^l \times_3 D_3^T \\
&\quad + \mathcal{R}_2^l \times_1 D_{11}^T + \mathcal{R}_2^l \times_2 D_{22}^T + \mathcal{R}_2^l \times_3 D_{33}^T), \\
\widetilde{\mathcal{Z}}^{l+1} &= \widetilde{\mathcal{Z}}^l + \frac{\gamma}{6}(\mathcal{R}_1^l \times_1 D_{111}^T + \mathcal{R}_1^l \times_2 D_{222}^T + \mathcal{R}_1^l \times_3 D_{333}^T \\
&\quad + \mathcal{R}_2^l \times_1 D_{1111}^T + \mathcal{R}_2^l \times_2 D_{2222}^T + \mathcal{R}_2^l \times_3 D_{3333}^T).
\end{aligned} \tag{5.5}$$

Applying straightening operator to (5.5) results in

$$\begin{aligned}
\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) &= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + \frac{\gamma}{6}[(I_{L_3} \otimes I_{L_2} \otimes D_1^T + I_{L_3} \otimes D_2^T \otimes I_{L_1} + D_3^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_1^l) \\
&\quad + (I_{L_3} \otimes I_{L_2} \otimes D_{11}^T + I_{L_3} \otimes D_{22}^T \otimes I_{L_1} + D_{33}^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_2^l)] \\
&= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) + \frac{\gamma}{6}[\bar{H}_1^T\mathrm{vec}(\mathcal{R}_1^l) + \bar{H}_2^T\mathrm{vec}(\mathcal{R}_2^l)], \\
\mathrm{vec}(\widetilde{\mathcal{Z}}^{l+1}) &= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) + \frac{\gamma}{6}[(I_{L_3} \otimes I_{L_2} \otimes D_{111}^T + I_{L_3} \otimes D_{222}^T \otimes I_{L_1} + D_{333}^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_1^l) \\
&\quad + (I_{L_3} \otimes I_{L_2} \otimes D_{1111}^T + I_{L_3} \otimes D_{2222}^T \otimes I_{L_1} + D_{3333}^T \otimes I_{L_2} \otimes I_{L_1})\mathrm{vec}(\mathcal{R}_2^l)] \\
&= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) + \frac{\gamma}{6}[\bar{H}_3^T\mathrm{vec}(\mathcal{R}_1^l) + \bar{H}_4^T\mathrm{vec}(\mathcal{R}_2^l)].
\end{aligned} \tag{5.6}$$

According to (5.4), we have $\mathrm{vec}(\mathcal{R}_1^l) = -H_1\mathrm{vec}(\widetilde{\mathcal{Y}}^l) - H_3\mathrm{vec}(\widetilde{\mathcal{Z}}^l)$ and $\mathrm{vec}(\mathcal{R}_2^l) = -H_2\mathrm{vec}(\widetilde{\mathcal{Y}}^l) - H_4\mathrm{vec}(\widetilde{\mathcal{Z}}^l)$, which together with (5.6) gives

$$\mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) = \mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}\{\bar{H}_1^T[H_1\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_3\mathrm{vec}(\widetilde{\mathcal{Z}}^l)] + \bar{H}_2^T[H_2\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_4\mathrm{vec}(\widetilde{\mathcal{Z}}^l)]\}$$

$$
\begin{aligned}
&= \mathrm{vec}(\widetilde{\mathcal{Y}}^l) - \frac{\gamma}{6}\{(\bar{H}_1^T H_1 + \bar{H}_2^T H_2)\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + (\bar{H}_1^T H_3 + \bar{H}_2^T H_4)\mathrm{vec}(\widetilde{\mathcal{Z}}^l)\},\\
\mathrm{vec}(\widetilde{\mathcal{Z}}^{l+1}) &= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) - \frac{\gamma}{6}\{\bar{H}_3^T[H_1\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_3\mathrm{vec}(\widetilde{\mathcal{Z}}^l)] + \bar{H}_4^T[H_2\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + H_4\mathrm{vec}(\widetilde{\mathcal{Z}}^l)]\}\\
&= \mathrm{vec}(\widetilde{\mathcal{Z}}^l) - \frac{\gamma}{6}\{(\bar{H}_3^T H_1 + \bar{H}_4^T H_2)\mathrm{vec}(\widetilde{\mathcal{Y}}^l) + (\bar{H}_3^T H_3 + \bar{H}_4^T H_4)\mathrm{vec}(\widetilde{\mathcal{Z}}^l)\}. \qquad (5.7)
\end{aligned}
$$

(5.7) can be equivalently transformed into the following equation

$$
\begin{aligned}
\begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^{l+1}) \\ \mathrm{vec}(\widetilde{\mathcal{Z}}^{l+1}) \end{pmatrix} &= \begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ \mathrm{vec}(\tilde{\mathcal{Z}}^l) \end{pmatrix} - \frac{\gamma}{6}\begin{pmatrix} \bar{H}_1^T H_1 + \bar{H}_2^T H_2 & \bar{H}_1^T H_3 + \bar{H}_2^T H_4 \\ \bar{H}_3^T H_1 + \bar{H}_4^T H_2 & \bar{H}_3^T H_3 + \bar{H}_4^T H_4 \end{pmatrix}\begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ \mathrm{vec}(\tilde{\mathcal{Z}}^l) \end{pmatrix}\\
&= \left[I_{2L_1L_2L_3} - \frac{\gamma}{6}\begin{pmatrix} \bar{H}_1 & \bar{H}_3 \\ \bar{H}_2 & \bar{H}_4 \end{pmatrix}^T \begin{pmatrix} H_1 & H_3 \\ H_2 & H_4 \end{pmatrix} \right]\begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ \mathrm{vec}(\tilde{\mathcal{Z}}^l) \end{pmatrix}\\
&:= (I_{2L_1L_2L_3} - \frac{\gamma}{6}U_1^T U)\begin{pmatrix} \mathrm{vec}(\widetilde{\mathcal{Y}}^l) \\ \mathrm{vec}(\tilde{\mathcal{Z}}^l) \end{pmatrix},
\end{aligned}
$$

where $U_1 = \begin{pmatrix} \bar{H}_1 & \bar{H}_3 \\ \bar{H}_2 & \bar{H}_4 \end{pmatrix}$ and $U = \begin{pmatrix} H_1 & H_3 \\ H_2 & H_4 \end{pmatrix}$. Thus $I_{2L_1L_2L_3} - \frac{\gamma}{6}U_1^T U$ is the iteration matrix of the DGI algorithm. Similar to the derivation of Theorem 3.2, the conclusions of this theorem are obtained. □

Theorem 5.3. *Assume that the conditions of Theorem 5.2 are satisfied, then the quasi-optimal parameter γ_{opt} and the corresponding quasi-optimal convergence factor of the DGI algorithm are*

(1) If $\widehat{Im}_1 \geq \sqrt{\widehat{Re}_{\max}\widehat{Re}_{\min}}$, then $\gamma_{opt} = \frac{6\widehat{Re}_{\min}}{\widehat{Re}_{\min}^2 + \widehat{Im}_1^2}$ and

$$\rho_{opt}(I_{2L_1L_2L_3} - \frac{1}{6}\gamma U_1^T U) = \frac{\widehat{Im}_1}{\sqrt{\widehat{Re}_{\min}^2 + \widehat{Im}_1^2}}.$$

(2) If $\widehat{Im}_1 \leq \sqrt{\widehat{Re}_{\max}\widehat{Re}_{\min}}$, there are two cases:

(i) When $0 < \gamma \leq \frac{12}{\widehat{Re}_{\max} + \widehat{Re}_{\min}}$, if $\widehat{Im}_1^2 < \frac{\widehat{Re}_{\min}(\widehat{Re}_{\max} - \widehat{Re}_{\min})}{2}$, then $\gamma_{opt} = \frac{12}{\widehat{Re}_{\max} + \widehat{Re}_{\min}}$ and

$$\rho_{opt}(I_{2L_1L_2L_3} - \frac{1}{6}\gamma U_1^T U) = \frac{\sqrt{(\widehat{Re}_{\max} - \widehat{Re}_{\min})^2 + 4\widehat{Im}_1^2}}{\widehat{Re}_{\max} + \widehat{Re}_{\min}}.$$

If $\widehat{Im}_1^2 \geq \frac{\widehat{Re}_{\min}(\widehat{Re}_{\max} - \widehat{Re}_{\min})}{2}$, then $\gamma_{opt} = \frac{6\widehat{Re}_{\min}}{\widehat{Re}_{\min}^2 + \widehat{Im}_1^2}$ and

$$\rho_{opt}(I_{2L_1L_2L_3} - \frac{1}{6}\gamma U_1^T U) = \frac{\widehat{Im}_1}{\sqrt{\widehat{Re}_{\min}^2 + \widehat{Im}_1^2}}.$$

(ii) When $\frac{12}{\widehat{Re}_{\max} + \widehat{Re}_{\min}} \leq \gamma < \frac{12\widehat{Re}_{\max}}{\widehat{Re}_{\max} + \widehat{Im}_1^2}$, then $\gamma_{opt} = \frac{12}{\widehat{Re}_{\max} + \widehat{Re}_{\min}}$ and

$$\rho_{opt}(I_{2L_1L_2L_3} - \frac{1}{6}\gamma U_1^T U) = \frac{\sqrt{(\widehat{Re}_{\max} - \widehat{Re}_{\min})^2 + 4\widehat{Im}_1^2}}{\widehat{Re}_{\max} + \widehat{Re}_{\min}}.$$

Proof. The proof is similar to that of Theorem 3.3, hence it is omitted. □

By applying the new update strategy to the DGI algorithm, we design the new MGI (NMGI) algorithm for the coupled Sylvester tensor equation with two unknowns (5.1) as below.

Algorithm 5.3. The new MGI (NMGI) algorithm:

Step 1. Given matrices $V_1, T_1, P_1, Q_1 \in \mathbb{R}^{L_1\times L_1}$, $V_2, T_2, P_2, Q_2 \in \mathbb{R}^{L_2\times L_2}$, $V_3, T_3, P_3, Q_3 \in \mathbb{R}^{L_3\times L_3}$, two tensors $\mathcal{W}_1, \mathcal{W}_2 \in \mathbb{R}^{L_1\times L_2\times L_3}$, and positive constants γ, η. Choose the initial tensors $\mathcal{Y}^0$, $\mathcal{Z}^0$, and set $l = 0$.

Step 2. If $\tau_l = \frac{\sqrt{\|\mathcal{W}_1-\sum\limits_{j=1}^{3}\mathcal{Y}^l\times_j V_j-\sum\limits_{j=1}^{3}\mathcal{Z}^l\times_j P_j\|^2+\|\mathcal{W}_2-\sum\limits_{j=1}^{3}\mathcal{Y}^l\times_j T_j-\sum\limits_{j=1}^{3}\mathcal{Z}^l\times_j Q_j\|^2}}{\sqrt{\|\mathcal{W}_1-\sum\limits_{j=1}^{3}\mathcal{Y}^0\times_j V_j-\sum\limits_{j=1}^{3}\mathcal{Z}^0\times_j P_j\|^2+\|\mathcal{W}_2-\sum\limits_{j=1}^{3}\mathcal{Y}^0\times_j T_j-\sum\limits_{j=1}^{3}\mathcal{Z}^0\times_j Q_j\|^2}} < \eta$, stop; otherwise, go to Step 3.

Step 3. Compute $\mathcal{Y}^{l+1}$ and $\mathcal{Z}^{l+1}$ by the following procedure

$$
\begin{aligned}
&\mathcal{Y}_1^{l+1} = \mathcal{Y}^l + \gamma\mathcal{R}_1^l\times_1 D_1^T,\ \mathcal{Z}_1^{l+1} = \mathcal{Z}^l + \gamma\mathcal{R}_1^l\times_1 D_{111}^T,\\
&\overrightarrow{\mathcal{Y}}^l = \frac{\mathcal{Y}_1^{l+1} + 5\mathcal{Y}^l}{6},\ \overrightarrow{\mathcal{Z}}^l = \frac{\mathcal{Z}_1^{l+1} + 5\mathcal{Z}^l}{6},\\
&\mathcal{Y}_2^{l+1} = \overrightarrow{\mathcal{Y}}^l + \gamma\overrightarrow{\mathcal{R}}_1^l\times_2 D_2^T,\ \mathcal{Z}_2^{l+1} = \overrightarrow{\mathcal{Z}}^l + \gamma\overrightarrow{\mathcal{R}}_1^l\times_2 D_{222}^T,\\
&\overline{\mathcal{Y}}^l = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + 4\mathcal{Y}^l}{6},\ \overline{\mathcal{Z}}^l = \frac{\mathcal{Z}_1^{l+1} + \mathcal{Z}_2^{l+1} + 4\mathcal{Z}^l}{6},\\
&\mathcal{Y}_3^{l+1} = \overline{\mathcal{Y}}^l + \gamma\overline{\mathcal{R}}_1^l\times_3 D_3^T,\ \mathcal{Z}_3^{l+1} = \overline{\mathcal{Z}}^l + \gamma\overline{\mathcal{R}}_1^l\times_3 D_{333}^T,\\
&\hat{\mathcal{Y}}^l = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + 3\mathcal{Y}^l}{6},\ \hat{\mathcal{Z}}^l = \frac{\mathcal{Z}_1^{l+1} + \mathcal{Z}_2^{l+1} + \mathcal{Z}_3^{l+1} + 3\mathcal{Z}^l}{6},\\
&\mathcal{Y}_4^{l+1} = \hat{\mathcal{Y}}^l + \gamma\hat{\mathcal{R}}_2^l\times_1 D_{11}^T,\ \mathcal{Z}_4^{l+1} = \hat{\mathcal{Z}}^l + \gamma\hat{\mathcal{R}}_2^l\times_1 D_{1111}^T,\\
&\check{\mathcal{Y}}^l = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + 2\mathcal{Y}^l}{6},\ \check{\mathcal{Z}}^l = \frac{\mathcal{Z}_1^{l+1} + \mathcal{Z}_2^{l+1} + \mathcal{Z}_3^{l+1} + \mathcal{Z}_4^{l+1} + 2\mathcal{Z}^l}{6},\\
&\mathcal{Y}_5^{l+1} = \check{\mathcal{Y}}^l + \gamma\check{\mathcal{R}}_2^l\times_2 D_{22}^T,\ \mathcal{Z}_5^{l+1} = \check{\mathcal{Z}}^l + \gamma\check{\mathcal{R}}_2^l\times_2 D_{2222}^T,\\
&\acute{\mathcal{Y}}^l = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}^l}{6},\ \acute{\mathcal{Z}}^l = \frac{\mathcal{Z}_1^{l+1} + \mathcal{Z}_2^{l+1} + \mathcal{Z}_3^{l+1} + \mathcal{Z}_4^{l+1} + \mathcal{Z}_5^{l+1} + \mathcal{Z}^l}{6},\\
&\mathcal{Y}_6^{l+1} = \acute{\mathcal{Y}}^l + \gamma\acute{\mathcal{R}}_2^l\times_3 D_{33}^T,\ \mathcal{Z}_6^{l+1} = \acute{\mathcal{Z}}^l + \gamma\acute{\mathcal{R}}_2^l\times_3 D_{3333}^T,\\
&\mathcal{Y}^{l+1} = \frac{\mathcal{Y}_1^{l+1} + \mathcal{Y}_2^{l+1} + \mathcal{Y}_3^{l+1} + \mathcal{Y}_4^{l+1} + \mathcal{Y}_5^{l+1} + \mathcal{Y}_6^{l+1}}{6},\\
&\mathcal{Z}^{l+1} = \frac{\mathcal{Z}_1^{l+1} + \mathcal{Z}_2^{l+1} + \mathcal{Z}_3^{l+1} + \mathcal{Z}_4^{l+1} + \mathcal{Z}_5^{l+1} + \mathcal{Z}_6^{l+1}}{6}.
\end{aligned}
$$

Step 4. Set $l := l + 1$ and return to Step 2.

In what follows, we discuss the convergence properties of the NMGI algorithm.

Theorem 5.4. *The NMGI algorithm is convergent if the parameter* γ *is selected to satisfy* $\rho(d_1 + d_2 f_1 + d_3 f_2 + d_4 f_3 + d_5 f_4 + d_6 f_5) < 6$, *where*

$$
\begin{aligned}
&p_1 = \|I_{L_1} - \gamma D_1^T V_1\|_2 + \gamma\|D_1\|_2\|V_2\|_2 + \gamma\|D_1\|_2\|V_3\|_2,\\
&\bar{p}_1 = \gamma(\|D_1^T P_1\|_2 + \|D_1\|_2\|P_2\|_2 + \|D_1\|_2\|P_3\|_2),
\end{aligned}
$$

$$
\begin{aligned}
p_2 &= \gamma(\|D_{111}^T V_1\|_2 + \|D_{111}\|_2\| V_2\|_2 + \|D_{111}\|_2\| V_3\|_2),\\
\bar{p}_2 &= \|I_{L_1} - \gamma D_{111}^T P_1\|_2 + \gamma\|D_{111}\|_2\|P_2\|_2 + \gamma\|D_{111}\|_2\|P_3\|_2,\\
p_3 &= \|I_{L_2} - \gamma D_2^T V_2\|_2 + \gamma\|V_1\|_2\|D_2\|_2 + \gamma\|D_2\|_2\|V_3\|_2,\\
\bar{p}_3 &= \gamma(\|P_1\|_2\|D_2\|_2 + \|D_2^T P_2\|_2 + \|D_2\|_2\|P_3\|_2),\\
p_4 &= \gamma(\|V_1\|_2\|D_{222}\|_2 + \|D_{222}^T V_2\|_2 + \|D_{222}\|_2\|V_3\|_2),\\
\bar{p}_4 &= \|I_{L_2} - \gamma D_{222}^T P_2\|_2 + \gamma\|P_1\|_2\|D_{222}\|_2 + \gamma\|D_{222}\|_2\|P_3\|_2,\\
p_5 &= \|I_{L_3} - \gamma D_3^T V_3\|_2 + \gamma\|V_1\|_2\|D_3\|_2 + \gamma\|V_2\|_2\|D_3\|_2,\\
\bar{p}_5 &= \gamma(\|P_1\|_2\|D_3\|_2 + \|P_2\|_2\|D_3\|_2 + \|D_3^T P_3\|_2),\\
p_6 &= \gamma(\|V_1\|_2\|D_{333}\|_2 + \|V_2\|_2\|D_{333}\|_2 + \|D_{333}^T V_3\|_2),\\
\bar{p}_6 &= \|I_{L_3} - \gamma D_{333}^T P_3\|_2 + \gamma\|P_1\|_2\|D_{333}\|_2 + \gamma\|P_2\|_2\|D_{333}\|_2,\\
p_7 &= \|I_{L_1} - \gamma D_{11}^T T_1\|_2 + \gamma\|D_{11}\|_2\| T_2\|_2 + \gamma\|D_{11}\|_2\| T_3\|_2,\\
\bar{p}_7 &= \gamma(\|D_{11}^T Q_1\|_2 + \|D_{11}\|_2\|Q_2\|_2 + \|D_{11}\|_2\|Q_3\|_2),\\
p_8 &= \gamma(\|D_{1111}^T T_1\|_2 + \|D_{1111}\|_2\|T_2\|_2 + \|D_{1111}\|_2\|T_3\|_2),\\
\bar{p}_8 &= \|I_{L_1} - \gamma D_{1111}^T Q_1\|_2 + \gamma\|D_{1111}\|_2\| Q_2\|_2 + \gamma\|D_{1111}\|_2\| Q_3\|_2,\\
p_9 &= \|I_{L_2} - \gamma D_{22}^T T_2\|_2 + \gamma\|T_1\|_2\|D_{22}\|_2 + \gamma\|D_{22}\|_2\|T_3\|_2,\\
\bar{p}_9 &= \gamma(\|Q_1\|_2\|D_{22}\|_2 + \|D_{22}^T Q_2\|_2 + \|D_{22}\|_2\|Q_3\|_2),\\
p_{10} &= \gamma(\|T_1\|_2\|D_{2222}\|_2 + \|D_{2222}^T T_2\|_2 + \|D_{2222}\|_2\|T_3\|_2),\\
\bar{p}_{10} &= \|I_{L_2} - \gamma D_{2222}^T Q_2\|_2 + \gamma\|Q_1\|_2\|D_{2222}\|_2 + \gamma\|D_{2222}\|_2\|Q_3\|_2,\\
p_{11} &= \|I_{L_3} - \gamma D_{33}^T T_3\|_2 + \gamma\|T_1\|_2\|D_{33}\|_2 + \gamma\|T_2\|_2\|D_{33}\|_2,\\
\bar{p}_{11} &= \gamma(\|D_{33}^T Q_3\|_2 + \|Q_1\|_2\|D_{33}\|_2 + \gamma\|Q_2\|_2\|D_{33}\|_2),\\
p_{12} &= \gamma(\|D_{3333}^T T_3\|_2 + \|T_1\|_2\|D_{3333}\|_2 + \gamma\|T_2\|_2\|D_{3333}\|_2),\\
\bar{p}_{12} &= \|I_{L_3} - \gamma D_{3333}^T Q_3\|_2 + \gamma\|Q_1\|_2\|D_{3333}\|_2 + \gamma\|Q_2\|_2\|D_{3333}\|_2,\\
d_1 &= \begin{pmatrix} p_1\, \bar{p}_1 \\ p_2\, \bar{p}_2 \end{pmatrix}, d_2 = \begin{pmatrix} p_3\, \bar{p}_3 \\ p_4\, \bar{p}_4 \end{pmatrix}, d_3 = \begin{pmatrix} p_5\, \bar{p}_5 \\ p_6\, \bar{p}_6 \end{pmatrix}, d_4 = \begin{pmatrix} p_7\, \bar{p}_7 \\ p_8\, \bar{p}_8 \end{pmatrix},\\
d_5 &= \begin{pmatrix} p_9 & \bar{p}_9 \\ p_{10} & \bar{p}_{10} \end{pmatrix}, d_6 = \begin{pmatrix} p_{11}\, \bar{p}_{11} \\ p_{12}\, \bar{p}_{12} \end{pmatrix},\\
f_1 &= \frac{d_1}{6} + \frac{5}{6}I_2,\ f_2 = \frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{2}{3}I_2,\ f_3 = \frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{1}{2}I_2,\\
f_4 &= \frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{d_4 f_3}{6} + \frac{1}{3}I_2, f_5 = \frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{d_4 f_3}{6} + \frac{d_5 f_4}{6} + \frac{1}{6}I_2,\\
f_6 &= \frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{d_4 f_3}{6} + \frac{d_5 f_4}{6} + \frac{d_6 f_5}{6}.
\end{aligned}
$$

Proof. We define the following error tensors

$$
\begin{aligned}
&\widetilde{\mathcal{Y}}^l = \mathcal{Y}^l - \mathcal{Y}^*,\ \widetilde{\overrightarrow{\mathcal{Y}}}^l = \overrightarrow{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\overline{\mathcal{Y}}}^l = \overline{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\hat{\mathcal{Y}}}^l = \hat{\mathcal{Y}}^l - \mathcal{Y}^*,\\
&\widetilde{\check{\mathcal{Y}}}^l = \check{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\acute{\mathcal{Y}}}^l = \acute{\mathcal{Y}}^l - \mathcal{Y}^*,\ \widetilde{\mathcal{Y}}_i^{l+1} = \mathcal{Y}_i^{l+1} - \mathcal{Y}^*,
\end{aligned}
$$

$$\widetilde{\mathcal{Z}}^l = \mathcal{Z}^l - \mathcal{Z}^*,\ \widetilde{\overrightarrow{\mathcal{Z}}}^l = \overrightarrow{\mathcal{Z}}^l - \mathcal{Z}^*,\ \widetilde{\overline{\mathcal{Z}}}^l = \overline{\mathcal{Z}}^l - \mathcal{Z}^*,\ \widetilde{\hat{\mathcal{Z}}}^l = \hat{\mathcal{Z}}^l - \mathcal{Z}^*,$$
$$\widetilde{\check{\mathcal{Z}}}^l = \check{\mathcal{Z}}^l - \mathcal{Z}^*,\ \widetilde{\acute{\mathcal{Z}}}^l = \acute{\mathcal{Z}}^l - \mathcal{Z}^*,\ \widetilde{\mathcal{Z}}_i^{l+1} = \mathcal{Z}_i^{l+1} - \mathcal{Z}^*\ (i = 1, \cdots, 6).$$

It follows from Line 1 of Step 3 of the NMGI algorithm that

$$\begin{aligned}
\widetilde{\mathcal{Y}}_1^{l+1} &= \widetilde{\mathcal{Y}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3 + \widetilde{\mathcal{Z}}^l \times_1 P_1 + \widetilde{\mathcal{Z}}^l \times_2 P_2 + \widetilde{\mathcal{Z}}^l \times_3 P_3) \times_1 D_1^T \\
&= \widetilde{\mathcal{Y}}^l \times_1 (I_{L_1} - \gamma D_1^T V_1) - \gamma \widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_2 V_2 - \gamma \widetilde{\mathcal{Y}}^l \times_1 D_1^T \times_3 V_3 \\
&\quad - \gamma \widetilde{\mathcal{Z}}^l \times_1 (D_1^T P_1) - \gamma \widetilde{\mathcal{Z}}^l \times_1 D_1^T \times_2 P_2 - \gamma \widetilde{\mathcal{Z}}^l \times_1 D_1^T \times_3 P_3, \\
\widetilde{\mathcal{Z}}_1^{l+1} &= \widetilde{\mathcal{Z}}^l - \gamma(\widetilde{\mathcal{Y}}^l \times_1 V_1 + \widetilde{\mathcal{Y}}^l \times_2 V_2 + \widetilde{\mathcal{Y}}^l \times_3 V_3 + \widetilde{\mathcal{Z}}^l \times_1 P_1 + \widetilde{\mathcal{Z}}^l \times_2 P_2 + \widetilde{\mathcal{Z}}^l \times_3 P_3) \times_1 D_{111}^T \\
&= \widetilde{\mathcal{Z}}^l \times_1 (I_{L_1} - \gamma D_{111}^T P_1) - \gamma \widetilde{\mathcal{Z}}^l \times_1 D_{111}^T \times_2 P_2 - \gamma \widetilde{\mathcal{Z}}^l \times_1 D_{111}^T \times_3 P_3 \\
&\quad - \gamma \widetilde{\mathcal{Y}}^l \times_1 (D_{111}^T V_1) - \gamma \widetilde{\mathcal{Y}}^l \times_1 D_{111}^T \times_2 V_2 - \gamma \widetilde{\mathcal{Y}}^l \times_1 D_{111}^T \times_3 V_3.
\end{aligned}$$

Similarly, in view of the expressions of $\mathcal{Y}_2^{l+1}$, $\mathcal{Z}_2^{l+1}$, $\mathcal{Y}_3^{l+1}$, $\mathcal{Z}_3^{l+1}$, $\mathcal{Y}_4^{l+1}$, $\mathcal{Z}_4^{l+1}$, $\mathcal{Y}_5^{l+1}$, $\mathcal{Z}_5^{l+1}$, $\mathcal{Y}_6^{l+1}$ and $\mathcal{Z}_6^{l+1}$, it has

$$\begin{aligned}
\widetilde{\mathcal{Y}}_2^{l+1} =& \widetilde{\overrightarrow{\mathcal{Y}}}^l \times_2 (I_{L_2} - \gamma D_2^T V_2) - \gamma \widetilde{\overrightarrow{\mathcal{Y}}}^l \times_1 V_1 \times_2 D_2^T - \gamma \widetilde{\overrightarrow{\mathcal{Y}}}^l \times_2 D_2^T \times_3 V_3 \\
&- \gamma \widetilde{\overrightarrow{\mathcal{Z}}}^l \times_1 P_1 \times_2 D_2^T - \gamma \widetilde{\overrightarrow{\mathcal{Z}}}^l \times_2 (D_2^T P_2) - \gamma \widetilde{\overrightarrow{\mathcal{Z}}}^l \times_2 D_2^T \times_3 P_3, \\
\widetilde{\mathcal{Z}}_2^{l+1} =& \widetilde{\overrightarrow{\mathcal{Z}}}^l \times_2 (I_{L_2} - \gamma D_{222}^T P_2) - \gamma \widetilde{\overrightarrow{\mathcal{Z}}}^l \times_1 P_1 \times_2 D_{222}^T - \gamma \widetilde{\overrightarrow{\mathcal{Z}}}^l \times_2 D_{222}^T \times_3 P_3 \\
&- \gamma \widetilde{\overrightarrow{\mathcal{Y}}}^l \times_1 V_1 \times_2 D_{222}^T - \gamma \widetilde{\overrightarrow{\mathcal{Y}}}^l \times_2 (D_{222}^T V_2) - \gamma \widetilde{\overrightarrow{\mathcal{Y}}}^l \times_2 D_{222}^T \times_3 V_3, \\
\widetilde{\mathcal{Y}}_3^{l+1} =& \widetilde{\overline{\mathcal{Y}}}^l \times_3 (I_{L_3} - \gamma D_3^T V_3) - \gamma \widetilde{\overline{\mathcal{Y}}}^l \times_1 V_1 \times_3 D_3^T - \gamma \widetilde{\overline{\mathcal{Y}}}^l \times_2 V_2 \times_3 D_3^T \\
&- \gamma \widetilde{\overline{\mathcal{Z}}}^l \times_1 P_1 \times_3 D_3^T - \gamma \widetilde{\overline{\mathcal{Z}}}^l \times_2 P_2 \times_3 D_3^T - \gamma \widetilde{\overline{\mathcal{Z}}}^l \times_3 (D_3^T P_3), \\
\widetilde{\mathcal{Z}}_3^{l+1} =& \widetilde{\overline{\mathcal{Z}}}^l \times_3 (I_{L_3} - \gamma D_{333}^T P_3) - \gamma \widetilde{\overline{\mathcal{Z}}}^l \times_1 P_1 \times_3 D_{333}^T - \gamma \widetilde{\overline{\mathcal{Z}}}^l \times_2 P_2 \times_3 D_{333}^T \\
&- \gamma \widetilde{\overline{\mathcal{Y}}}^l \times_1 V_1 \times_3 D_{333}^T - \gamma \widetilde{\overline{\mathcal{Y}}}^l \times_2 V_2 \times_3 D_{333}^T - \gamma \widetilde{\overline{\mathcal{Y}}}^l \times_3 (D_{333}^T V_3), \\
\widetilde{\mathcal{Y}}_4^{l+1} =& \widetilde{\hat{\mathcal{Y}}}^l \times_1 (I_{L_1} - \gamma D_{11}^T T_1) - \gamma \widetilde{\hat{\mathcal{Y}}}^l \times_1 D_{11}^T \times_2 T_2 - \gamma \widetilde{\hat{\mathcal{Y}}}^l \times_1 D_{11}^T \times_3 T_3 \\
&- \gamma \widetilde{\hat{\mathcal{Z}}}^l \times_1 (D_{11}^T Q_1) - \gamma \widetilde{\hat{\mathcal{Z}}}^l \times_1 D_{11}^T \times_2 Q_2 - \gamma \widetilde{\hat{\mathcal{Z}}}^l \times_1 D_{11}^T \times_3 Q_3, \\
\widetilde{\mathcal{Z}}_4^{l+1} =& \widetilde{\hat{\mathcal{Z}}}^l \times_1 (I_{L_1} - \gamma D_{1111}^T Q_1) - \gamma \widetilde{\hat{\mathcal{Z}}}^l \times_1 D_{1111}^T \times_2 Q_2 - \gamma \widetilde{\hat{\mathcal{Z}}}^l \times_1 D_{1111}^T \times_3 Q_3 \\
&- \gamma \widetilde{\hat{\mathcal{Y}}}^l \times_1 (D_{1111}^T T_1) - \gamma \widetilde{\hat{\mathcal{Y}}}^l \times_1 D_{1111}^T \times_2 T_2 - \gamma \widetilde{\hat{\mathcal{Y}}}^l \times_1 D_{1111}^T \times_3 T_3, \\
\widetilde{\mathcal{Y}}_5^{l+1} =& \widetilde{\check{\mathcal{Y}}}^l \times_2 (I_{L_2} - \gamma D_{22}^T T_2) - \gamma \widetilde{\check{\mathcal{Y}}}^l \times_1 T_1 \times_2 D_{22}^T - \gamma \widetilde{\check{\mathcal{Y}}}^l \times_2 D_{22}^T \times_3 T_3 - \gamma \widetilde{\check{\mathcal{Z}}}^l \times_2 (D_{22}^T Q_2) \\
&- \gamma \widetilde{\check{\mathcal{Z}}}^l \times_1 Q_1 \times_2 D_{22}^T - \gamma \widetilde{\check{\mathcal{Z}}}^l \times_2 D_{22}^T \times_3 Q_3, \\
\widetilde{\mathcal{Z}}_5^{l+1} =& \widetilde{\check{\mathcal{Z}}}^l \times_2 (I_{L_2} - \gamma D_{2222}^T Q_2) - \gamma \widetilde{\check{\mathcal{Z}}}^l \times_1 Q_1 \times_2 D_{2222}^T - \gamma \widetilde{\check{\mathcal{Z}}}^l \times_2 D_{2222}^T \times_3 Q_3 \\
&- \gamma \widetilde{\check{\mathcal{Y}}}^l \times_2 (D_{2222}^T T_2) - \gamma \widetilde{\check{\mathcal{Y}}}^l \times_1 T_1 \times_2 D_{2222}^T - \gamma \widetilde{\check{\mathcal{Y}}}^l \times_2 D_{2222}^T \times_3 T_3, \\
\widetilde{\mathcal{Y}}_6^{l+1} =& \widetilde{\acute{\mathcal{Y}}}^l \times_3 (I_{L_3} - \gamma D_{33}^T T_3) - \gamma \widetilde{\acute{\mathcal{Y}}}^l \times_1 T_1 \times_3 D_{33}^T - \gamma \widetilde{\acute{\mathcal{Y}}}^l \times_2 T_2 \times_3 D_{33}^T - \gamma \widetilde{\acute{\mathcal{Z}}}^l \times_3 (D_{33}^T Q_3)
\end{aligned}$$

$$
\begin{aligned}
& - \gamma \widetilde{\acute{\mathcal{Z}}}^{l} \times_1 Q_1 \times_3 D_{33}^T - \gamma \widetilde{\acute{\mathcal{Z}}}^{l} \times_2 Q_2 \times_3 D_{33}^T, \\
\widetilde{\mathcal{Z}}_6^{l+1} = & \widetilde{\acute{\mathcal{Z}}}^{l} \times_3 (I_{L_3} - \gamma D_{3333}^T Q_3) - \gamma \widetilde{\acute{\mathcal{Z}}}^{l} \times_1 Q_1 \times_3 D_{3333}^T - \gamma \widetilde{\acute{\mathcal{Z}}}^{l} \times_2 Q_2 \times_3 D_{3333}^T \\
& - \gamma \widetilde{\acute{\mathcal{Y}}}^{l} \times_3 (D_{3333}^T T_3) - \gamma \widetilde{\acute{\mathcal{Y}}}^{l} \times_1 T_1 \times_3 D_{3333}^T - \gamma \widetilde{\acute{\mathcal{Y}}}^{l} \times_2 T_2 \times_3 D_{3333}^T,
\end{aligned}
$$

from which one can deduce that

$$
\begin{aligned}
\|\widetilde{\mathcal{Y}}_1^{l+1}\| \leq\ & \|\widetilde{\mathcal{Y}}^l\|(\|I_{L_1} - \gamma D_1^T V_1\|_2 + \gamma\|D_1\|_2\|V_2\|_2 + \gamma\|D_1\|_2\|V_3\|_2) \\
& + \gamma\|\widetilde{\mathcal{Z}}^l\|(\|D_1^T P_1\|_2 + \|D_1\|_2\|P_2\|_2 + \|D_1\|_2\|P_3\|_2) \\
:=\ & p_1\|\widetilde{\mathcal{Y}}^l\| + \bar{p}_1\|\widetilde{\mathcal{Z}}^l\|, \\
\|\widetilde{\mathcal{Z}}_1^{l+1}\| \leq\ & \gamma\|\widetilde{\mathcal{Y}}^l\|(\|D_{111}^T V_1\|_2 + \|D_{111}\|_2\|V_2\|_2 + \|D_{111}\|_2\|V_3\|_2) \\
& + \|\widetilde{\mathcal{Z}}^l\|(\|I_{L_1} - \gamma D_{111}^T P_1\|_2 + \gamma\|D_{111}\|_2\|P_2\|_2 + \gamma\|D_{111}\|_2\|P_3\|_2) \\
:=\ & p_2\|\widetilde{\mathcal{Y}}^l\| + \bar{p}_2\|\widetilde{\mathcal{Z}}^l\|, \\
\|\widetilde{\mathcal{Y}}_2^{l+1}\| \leq\ & \|\widetilde{\overrightarrow{\mathcal{Y}}}^{l}\|(\|I_{L_2} - \gamma D_2^T V_2\|_2 + \gamma\|V_1\|_2\|D_2\|_2 + \gamma\|D_2\|_2\|V_3\|_2) \\
& + \gamma\|\widetilde{\overrightarrow{\mathcal{Z}}}^{l}\|(\|P_1\|_2\|D_2\|_2 + \|D_2^T P_2\|_2 + \|D_2\|_2\|P_3\|_2) \\
:=\ & p_3\|\widetilde{\overrightarrow{\mathcal{Y}}}^{l}\| + \bar{p}_3\|\widetilde{\overrightarrow{\mathcal{Z}}}^{l}\|, \\
\|\widetilde{\mathcal{Z}}_2^{l+1}\| \leq\ & \gamma\|\widetilde{\overrightarrow{\mathcal{Y}}}^{l}\|(\|V_1\|_2\|D_{222}\|_2 + \|D_{222}^T V_2\|_2 + \|D_{222}\|_2\|V_3\|_2) \\
& + \|\widetilde{\overrightarrow{\mathcal{Z}}}^{l}\|(\|I_{L_2} - \gamma D_{222}^T P_2\|_2 + \gamma\|P_1\|_2\|D_{222}\|_2 + \gamma\|D_{222}\|_2\|P_3\|_2) \\
:=\ & p_4\|\widetilde{\overrightarrow{\mathcal{Y}}}^{l}\| + \bar{p}_4\|\widetilde{\overrightarrow{\mathcal{Z}}}^{l}\|, \\
\|\widetilde{\mathcal{Y}}_3^{l+1}\| \leq\ & \|\widetilde{\overline{\mathcal{Y}}}^{l}\|(\|I_{L_3} - \gamma D_3^T V_3\|_2 + \gamma\|V_1\|_2\|D_3\|_2 + \gamma\|V_2\|_2\|D_3\|_2) \\
& + \gamma\|\widetilde{\overline{\mathcal{Z}}}^{l}\|(\|P_1\|_2\|D_3\|_2 + \|P_2\|_2\|D_3\|_2 + \|D_3^T P_3\|_2) \\
:=\ & p_5\|\widetilde{\overline{\mathcal{Y}}}^{l}\| + \bar{p}_5\|\widetilde{\overline{\mathcal{Z}}}^{l}\|, \\
\|\widetilde{\mathcal{Z}}_3^{l+1}\| \leq\ & \gamma\|\widetilde{\overline{\mathcal{Y}}}^{l}\|(\|V_1\|_2\|D_{333}\|_2 + \|V_2\|_2\|D_{333}\|_2 + \|D_{333}^T V_3\|_2) \\
& + \|\widetilde{\overline{\mathcal{Z}}}^{l}\|(\|I_{L_3} - \gamma D_{333}^T P_3\|_2 + \gamma\|P_1\|_2\|D_{333}\|_2 + \gamma\|P_2\|_2\|D_{333}\|_2) \\
:=\ & p_6\|\widetilde{\overline{\mathcal{Y}}}^{l}\| + \bar{p}_6\|\widetilde{\overline{\mathcal{Z}}}^{l}\|, \\
\|\widetilde{\mathcal{Y}}_4^{l+1}\| \leq\ & \|\widetilde{\hat{\mathcal{Y}}}^{l}\|(\|I_{L_1} - \gamma D_{11}^T T_1\|_2 + \gamma\|D_{11}\|_2\|T_2\|_2 + \gamma\|D_{11}\|_2\|T_3\|_2) \\
& + \gamma\|\widetilde{\hat{\mathcal{Z}}}^{l}\|(\|D_{11}^T Q_1\|_2 + \|D_{11}\|_2\|Q_2\|_2 + \|D_{11}\|_2\|Q_3\|_2) \\
:=\ & p_7\|\widetilde{\hat{\mathcal{Y}}}^{l}\| + \bar{p}_7\|\widetilde{\hat{\mathcal{Z}}}^{l}\|, \\
\|\widetilde{\mathcal{Z}}_4^{l+1}\| \leq\ & \gamma\|\widetilde{\hat{\mathcal{Y}}}^{l}\|(\|D_{1111}^T T_1\|_2 + \|D_{1111}\|_2\|T_2\|_2 + \|D_{1111}\|_2\|T_3\|_2) \\
& + \|\widetilde{\hat{\mathcal{Z}}}^{l}\|(\|I_{L_1} - \gamma D_{1111}^T Q_1\|_2 + \gamma\|D_{1111}\|_2\|Q_2\|_2 + \gamma\|D_{1111}\|_2\|Q_3\|_2) \\
:=\ & p_8\|\widetilde{\hat{\mathcal{Y}}}^{l}\| + \bar{p}_8\|\widetilde{\hat{\mathcal{Z}}}^{l}\|, \\
\|\widetilde{\mathcal{Y}}_5^{l+1}\| \leq\ & \|\widetilde{\check{\mathcal{Y}}}^{l}\|(\|I_{L_2} - \gamma D_{22}^T T_2\|_2 + \gamma\|T_1\|_2\|D_{22}\|_2 + \gamma\|D_{22}\|_2\|T_3\|_2)
\end{aligned}
$$

$$
\begin{aligned}
&\quad +\gamma\|\widetilde{\check{\mathcal{Z}}}^l\|(\|Q_1\|_2\|D_{22}\|_2+\|D_{22}^TQ_2\|_2+\|D_{22}\|_2\|Q_3\|_2)\\
&:= p_9\|\widetilde{\check{\mathcal{Y}}}^l\|+\bar{p}_9\|\widetilde{\check{\mathcal{Z}}}^l\|,\\
\|\widetilde{\mathcal{Z}}_5^{l+1}\| &\le \gamma\|\widetilde{\check{\mathcal{Y}}}^l\|(\|T_1\|_2\|D_{2222}\|_2+\|D_{2222}^TT_2\|_2+\|D_{2222}\|_2\|T_3\|_2)\\
&\quad +\|\widetilde{\check{\mathcal{Z}}}^l\|(\|I_{L_2}-\gamma D_{2222}^TQ_2\|_2+\gamma\|Q_1\|_2\|D_{2222}\|_2+\gamma\|D_{2222}\|_2\|Q_3\|_2)\\
&:= p_{10}\|\widetilde{\check{\mathcal{Y}}}^l\|+\bar{p}_{10}\|\widetilde{\check{\mathcal{Z}}}^l\|,\\
\|\widetilde{\mathcal{Y}}_6^{l+1}\| &\le \|\widetilde{\acute{\mathcal{Y}}}^l\|(\|I_{L_3}-\gamma D_{33}^TT_3\|_2+\gamma\|T_1\|_2\|D_{33}\|_2+\gamma\|T_2\|_2\|D_{33}\|_2)\\
&\quad +\gamma\|\widetilde{\acute{\mathcal{Z}}}^l\|(\|D_{33}^TQ_3\|_2+\|Q_1\|_2\|D_{33}\|_2+\gamma\|Q_2\|_2\|D_{33}\|_2)\\
&:= p_{11}\|\widetilde{\acute{\mathcal{Y}}}^l\|+\bar{p}_{11}\|\widetilde{\acute{\mathcal{Z}}}^l\|,\\
\|\widetilde{\mathcal{Z}}_6^{l+1}\| &\le \gamma\|\widetilde{\acute{\mathcal{Y}}}^l\|(\|D_{3333}^TT_3\|_2+\|T_1\|_2\|D_{3333}\|_2+\gamma\|T_2\|_2\|D_{3333}\|_2)\\
&\quad +\|\widetilde{\acute{\mathcal{Z}}}^l\|(\|I_{L_3}-\gamma D_{3333}^TQ_3\|_2+\gamma\|Q_1\|_2\|D_{3333}\|_2+\gamma\|Q_2\|_2\|D_{3333}\|_2)\\
&:= p_{12}\|\widetilde{\acute{\mathcal{Y}}}^l\|+\bar{p}_{12}\|\widetilde{\acute{\mathcal{Z}}}^l\|,
\end{aligned}
$$

in terms of Lemmas 2.1–2.2. In consequence, we have that

$$
\begin{aligned}
\begin{pmatrix}\|\widetilde{\mathcal{Y}}_1^{l+1}\|\\ \|\widetilde{\mathcal{Z}}_1^{l+1}\|\end{pmatrix} &\le \begin{pmatrix}p_1 & \bar{p}_1\\ p_2 & \bar{p}_2\end{pmatrix}\begin{pmatrix}\|\widetilde{\mathcal{Y}}^l\|\\ \|\widetilde{\mathcal{Z}}^l\|\end{pmatrix} := d_1\begin{pmatrix}\|\widetilde{\mathcal{Y}}^l\|\\ \|\widetilde{\mathcal{Z}}^l\|\end{pmatrix},\\
\begin{pmatrix}\|\widetilde{\mathcal{Y}}_2^{l+1}\|\\ \|\widetilde{\mathcal{Z}}_2^{l+1}\|\end{pmatrix} &\le \begin{pmatrix}p_3 & \bar{p}_3\\ p_4 & \bar{p}_4\end{pmatrix}\begin{pmatrix}\|\widetilde{\overrightarrow{\mathcal{Y}}}^l\|\\ \|\widetilde{\overrightarrow{\mathcal{Z}}}^l\|\end{pmatrix} := d_2\begin{pmatrix}\|\widetilde{\overrightarrow{\mathcal{Y}}}^l\|\\ \|\widetilde{\overrightarrow{\mathcal{Z}}}^l\|\end{pmatrix},\\
\begin{pmatrix}\|\widetilde{\mathcal{Y}}_3^{l+1}\|\\ \|\widetilde{\mathcal{Z}}_3^{l+1}\|\end{pmatrix} &\le \begin{pmatrix}p_5 & \bar{p}_5\\ p_6 & \bar{p}_6\end{pmatrix}\begin{pmatrix}\|\widetilde{\bar{\mathcal{Y}}}^l\|\\ \|\widetilde{\bar{\mathcal{Z}}}^l\|\end{pmatrix} := d_3\begin{pmatrix}\|\widetilde{\bar{\mathcal{Y}}}^l\|\\ \|\widetilde{\bar{\mathcal{Z}}}^l\|\end{pmatrix},\\
\begin{pmatrix}\|\widetilde{\mathcal{Y}}_4^{l+1}\|\\ \|\widetilde{\mathcal{Z}}_4^{l+1}\|\end{pmatrix} &\le \begin{pmatrix}p_7 & \bar{p}_7\\ p_8 & \bar{p}_8\end{pmatrix}\begin{pmatrix}\|\widetilde{\hat{\mathcal{Y}}}^l\|\\ \|\widetilde{\hat{\mathcal{Z}}}^l\|\end{pmatrix} := d_4\begin{pmatrix}\|\widetilde{\hat{\mathcal{Y}}}^l\|\\ \|\widetilde{\hat{\mathcal{Z}}}^l\|\end{pmatrix},\\
\begin{pmatrix}\|\widetilde{\mathcal{Y}}_5^{l+1}\|\\ \|\widetilde{\mathcal{Z}}_5^{l+1}\|\end{pmatrix} &\le \begin{pmatrix}p_9 & \bar{p}_9\\ p_{10} & \bar{p}_{10}\end{pmatrix}\begin{pmatrix}\|\widetilde{\check{\mathcal{Y}}}^l\|\\ \|\widetilde{\check{\mathcal{Z}}}^l\|\end{pmatrix} := d_5\begin{pmatrix}\|\widetilde{\check{\mathcal{Y}}}^l\|\\ \|\widetilde{\check{\mathcal{Z}}}^l\|\end{pmatrix},\\
\begin{pmatrix}\|\widetilde{\mathcal{Y}}_6^{l+1}\|\\ \|\widetilde{\mathcal{Z}}_6^{l+1}\|\end{pmatrix} &\le \begin{pmatrix}p_{11} & \bar{p}_{11}\\ p_{12} & \bar{p}_{12}\end{pmatrix}\begin{pmatrix}\|\widetilde{\acute{\mathcal{Y}}}^l\|\\ \|\widetilde{\acute{\mathcal{Z}}}^l\|\end{pmatrix} := d_6\begin{pmatrix}\|\widetilde{\acute{\mathcal{Y}}}^l\|\\ \|\widetilde{\acute{\mathcal{Z}}}^l\|\end{pmatrix}.
\end{aligned}
$$

By making use of Lemma 2.2, we deduce that

$$
\begin{pmatrix}\|\widetilde{\overrightarrow{\mathcal{Y}}}^l\|\\ \|\widetilde{\overrightarrow{\mathcal{Z}}}^l\|\end{pmatrix} \le \frac{1}{6}\begin{pmatrix}\|\widetilde{\mathcal{Y}}_1^{l+1}\|\\ \|\widetilde{\mathcal{Z}}_1^{l+1}\|\end{pmatrix}+\frac{5}{6}\begin{pmatrix}\|\widetilde{\mathcal{Y}}^l\|\\ \|\widetilde{\mathcal{Z}}^l\|\end{pmatrix} \le \left(\frac{d_1}{6}+\frac{5}{6}I_2\right)\begin{pmatrix}\|\widetilde{\mathcal{Y}}^l\|\\ \|\widetilde{\mathcal{Z}}^l\|\end{pmatrix} := f_1\begin{pmatrix}\|\widetilde{\mathcal{Y}}^l\|\\ \|\widetilde{\mathcal{Z}}^l\|\end{pmatrix},
$$

$$
\begin{aligned}
\begin{pmatrix} \|\widetilde{\bar{\mathcal{Y}}}^{l}\| \\ \|\widetilde{\bar{\mathcal{Z}}}^{l}\| \end{pmatrix} \leq & \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_1^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_1^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_2^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_2^{l+1}\| \end{pmatrix} + \frac{2}{3}\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
\leq & (\frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{2}{3} I_2)\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
:= & f_2 \begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix}, \\
\begin{pmatrix} \|\widetilde{\hat{\mathcal{Y}}}^{l}\| \\ \|\widetilde{\hat{\mathcal{Z}}}^{l}\| \end{pmatrix} \leq & \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_1^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_1^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_2^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_2^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_3^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_3^{l+1}\| \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
\leq & (\frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{1}{2} I_2)\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
:= & f_3 \begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix}, \\
\begin{pmatrix} \|\widetilde{\check{\mathcal{Y}}}^{l}\| \\ \|\widetilde{\check{\mathcal{Z}}}^{l}\| \end{pmatrix} \leq & \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_1^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_1^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_2^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_2^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_3^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_3^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_4^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_4^{l+1}\| \end{pmatrix} + \frac{1}{3}\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
\leq & (\frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{d_4 f_3}{6} + \frac{1}{3} I_2)\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
:= & f_4 \begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix}, \\
\begin{pmatrix} \|\widetilde{\acute{\mathcal{Y}}}^{l}\| \\ \|\widetilde{\acute{\mathcal{Z}}}^{l}\| \end{pmatrix} \leq & \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_1^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_1^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_2^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_2^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_3^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_3^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_4^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_4^{l+1}\| \end{pmatrix} \\
& + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_5^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_5^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
\leq & (\frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{d_4 f_3}{6} + \frac{d_5 f_4}{6} + \frac{1}{6} I_2)\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
:= & f_5 \begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix}.
\end{aligned}
$$

Thus it holds that

$$
\begin{aligned}
\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l+1}\| \\ \|\widetilde{\mathcal{Z}}^{l+1}\| \end{pmatrix} \leq & \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_1^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_1^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_2^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_2^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_3^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_3^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_4^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_4^{l+1}\| \end{pmatrix} \\
& + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_5^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_5^{l+1}\| \end{pmatrix} + \frac{1}{6}\begin{pmatrix} \|\widetilde{\mathcal{Y}}_6^{l+1}\| \\ \|\widetilde{\mathcal{Z}}_6^{l+1}\| \end{pmatrix} \\
\leq & \left(\frac{d_1}{6} + \frac{d_2 f_1}{6} + \frac{d_3 f_2}{6} + \frac{d_4 f_3}{6} + \frac{d_5 f_4}{6} + \frac{d_6 f_5}{6}\right)\begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix} \\
:= & f_6 \begin{pmatrix} \|\widetilde{\mathcal{Y}}^{l}\| \\ \|\widetilde{\mathcal{Z}}^{l}\| \end{pmatrix}.
\end{aligned}
$$

This shows that if $\rho(f_6) < 1$, i.e., $\rho(d_1 + d_2 f_1 + d_3 f_2 + d_4 f_3 + d_5 f_4 + d_6 f_5) < 6$, then the NMGI algorithm is convergent. □

6. Numerical experiments for Sylvester tensor equations

In this section, three numerical examples are performed to validate the correctness of theories, and the effectiveness, feasibilities and advantages of the optimal GI (OGI), DGI and NMGI algorithms for the Sylvester tensor equations, with respect to the number of iteration steps (IT) and the elapsed time in seconds (CPU). In Example 6.1, the specific matrices come from Example 5 in [6] and Example 4.1 in [43]. Using the same system matrices in [6, 43] makes the numerical comparisons more convincing. And the system matrices in Example 6.2 are taken as in Example 4.2 of [43] and the selection of them is random, then the results obtained by these matrices have certain universal applicable. Besides, by applying the proper finite difference discretization method to the convection-diffusion equation [4], it can be obtained the specific matrices involved in Example 6.3, which shows that the proposed methods can be utilized to solve the practical problems and have certain practicabilities.

All numerical experiments are computed in MATLAB (version R2018a) in a personal computer with Intel (R) Core (TM) i9-10900 2.81 GHz and 32.0 GB RAM. In our computations, the initial tensors are taken to be $\mathcal{Y}^0 = \mathcal{Y}_2^0 = \mathcal{Y}_3^0 = 10^{-6} \cdot tenones(L_1, L_2, L_3)$, whose all elements are 1. And all runs are terminated once

$$\mathrm{RES} = \frac{\|\mathcal{W} - \mathcal{Y}^l \times_1 V_1 - \mathcal{Y}^l \times_2 V_2 - \mathcal{Y}^l \times_3 V_3\|}{\|\mathcal{W} - \mathcal{Y}^0 \times_1 V_1 - \mathcal{Y}^0 \times_2 V_2 - \mathcal{Y}^0 \times_3 V_3\|} \leq \eta,$$

with $\eta > 0$, or the IT exceeds $l_{\max} = 10000$. We denote the latter case by "Invalid" and "–" in tables. According to [6, 43], Theorems 3.1 and 3.3, the parameters of the tested algorithms are taken as follows

$$
\begin{aligned}
\gamma_{\mathrm{GI}} &= \frac{1}{\|V_1\|_2^2 + \|V_2\|_2^2 + \|V_3\|_2^2}, \\
\gamma_{\mathrm{OGI}} &= \frac{6}{\lambda_{\max} + \lambda_{\min}},
\end{aligned}
$$

$$\gamma_{\text{RGI}} = \frac{1}{(\alpha-\beta)\beta\|V_1\|_2^2 + (1-\alpha)\beta\|V_2\|_2^2 + (1-\alpha)(\alpha-\beta)\|V_3\|_2^2},$$
$$\gamma_{\text{DGI}} = \frac{6}{Re_{\min} + Re_{\max}} \text{ or } \frac{3Re_{\min}}{Re_{\min}^2 + Im_1^2},$$

where $\lambda_{\max}, \lambda_{\min}$ and $Re_{\min}, Re_{\max}, Im_1$ are defined as in Theorem 3.1 and Theorem 3.2, respectively. Besides, the parameters adopted in the MGI, MRGI and NMGI algorithms are the experimental optimal ones that minimize their IT as $\eta = 10^{-6}$.

In Table 1, the comparisons for the number of arithmetic operations of the tested algorithms are reported. As observed in Table 1, the number of arithmetic operations for each iteration of the DGI algorithm is the least, and it is nearly a half of, a third of and a fourth of those for GI and RGI, NMGI, and MGI and MRGI ones, respectively. This indicates that the DGI algorithm may be less than the other tested ones if their IT are comparable. Although the computational complexity of the NMGI algorithm is more than the GI and RGI ones, the former has faster convergence speed for many cases, thus the CPU time of the former may be more than the latter ones. Moreover, the number of arithmetic operations for the MGI and MRGI algorithms are comparable, and they are more than the other ones. Specially, they are nearly 1.33, 2 and 4 times of those for NMGI, GI and RGI, DGI ones, respectively. Nevertheless, the MGI and MRGI algorithms are established by applying the update strategy to the GI and RGI algorithms, respectively, then the former ones have faster convergence speeds for almost all cases and their IT may be far less than the latter ones, which leads to less CPU time. From the perspective of the computing times for each iteration, the proposed DGI and NMGI algorithms outperform the MGI and MRGI ones. Specially, if their IT are not much different, then the former ones need less CPU time. These facts will be verified by numerical results in Tables 2–7.

Example 6.1. [6,43] We compute the approximate solutions of Sylvester tensor equation (1.1), where

$$V_1 = \begin{pmatrix} 3 & 1 \\ -1 & 2 \end{pmatrix},\ V_2 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},\ V_3 = \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix},$$
$$\mathcal{W}(:,:,1) = \begin{pmatrix} 10 & 13 \\ 15 & 11 \end{pmatrix},\ \mathcal{W}(:,:,2) = \begin{pmatrix} 14 & 3 \\ 3 & 0 \end{pmatrix}.$$

Table 1. The number of arithmetic operations of the tested algorithms.

Algorithm	The computing times for each iteration of the tested algorithms
GI	$L_1L_2L_3(4L_1+4L_2+4L_3+6)$
MGI	$L_1L_2L_3(8L_1+8L_2+8L_3+12)$
RGI	$L_1L_2L_3(4L_1+4L_2+4L_3+8)+9$
DGI	$L_1L_2L_3(2L_1+2L_2+2L_3+12)$
MRGI	$L_1L_2L_3(8L_1+8L_2+8L_3+18)+10$
NMGI	$L_1L_2L_3(6L_1+6L_2+6L_3+18)$

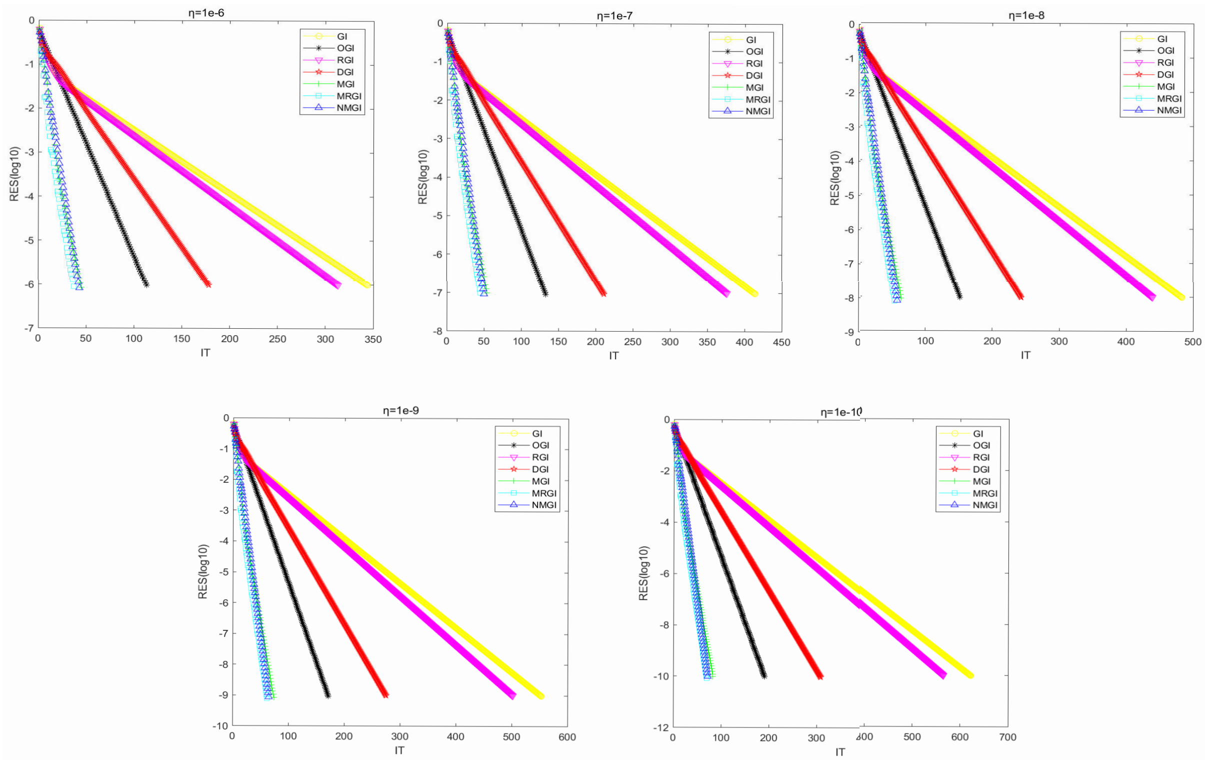

Figure 1. RES(log10) curves of seven algorithms for Example 6.1 with five values of η.

In Table 2, we report the parameters and numerical results of the seven tested algorithms with five different values of η. It can be observed from Table 2 that all algorithms can successfully compute approximate solutions satisfying the prescribed stopping criterion within $l_{\max}$ iteration steps, and the IT and CPU time of the tested methods increase with decreasing of η. The RGI algorithm is more efficient than the GI one with respect to IT and CPU time when the relaxation factor is chosen properly. Meanwhile, the OGI algorithm (GI algorithm with the optimal parameter γ_{opt} in Theorem 3.1) returns better numerical performance than the GI one, which further confirms the correctness of Theorem 3.1. In addition, the DGI algorithm outperforms the GI and RGI ones in terms of computational efficiency due to the fact that the former requires less IT and CPU time than the latter ones. Specially, the IT and CPU time of the DGI algorithm are nearly half of those for the GI and RGI ones. And the MGI algorithm performs better than DGI, RGI, GI and OGI ones. Although the IT of MGI algorithm is far less than the OGI one, their CPU time are close. The reason is that each step of the MGI algorithm needs less basic arithmetic operations than the OGI one, thus total computational loads of the MGI and OGI algorithms are comparable. This coincides with the results in Table 1. Also, numerical performances of the NMGI and MRGI algorithms are comparable, and they are superior to the other ones. And their advantages become more pronounced as the value of η becomes smaller, because the numerical performances gap between the NMGI, MRGI algorithms and the remaining tested ones are increasingly larger with the decreases of η. Last but not least, the proposed DGI algorithm is more stable than the GI and RGI ones as the changing scope of the IT for the former is smaller than the latter ones. And the stability of the proposed NMGI algorithm is the highest among all tested algorithms since the variation range of its IT is the smallest in all algorithms.

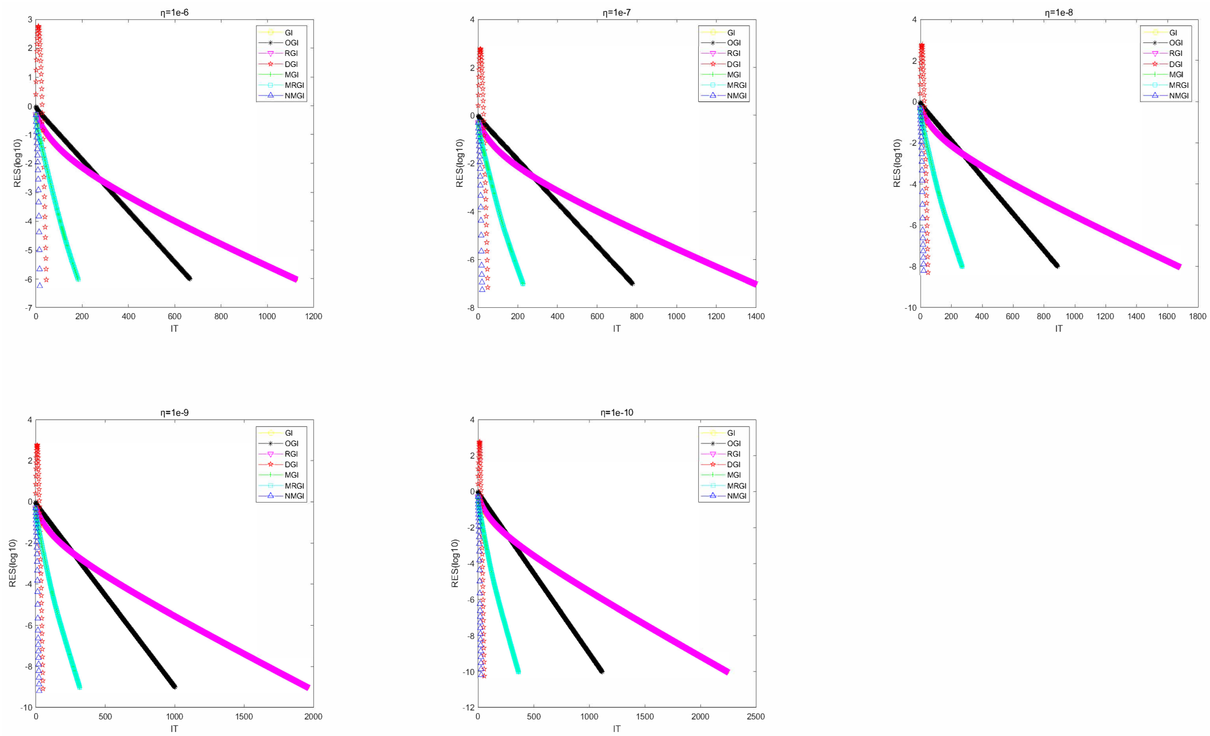

Figure 2. RES(log10) curves of seven algorithms for Example 6.2 with five values of η and $\rho = 2$.

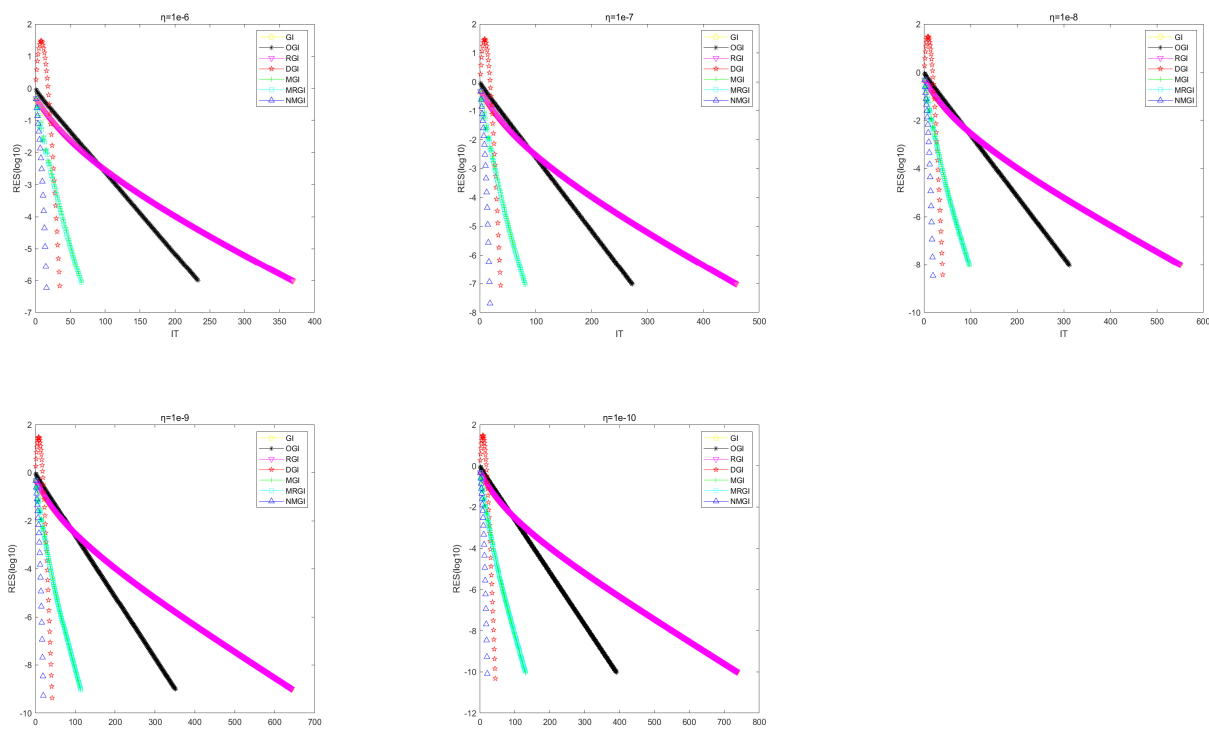

Figure 3. RES(log10) curves of seven algorithms for Example 6.2 with five values of η and $\rho = 3$.

To better validate the superiorities of the proposed OGI, DGI and NMGI algorithms, Figure 1 displays the curves of the RES in base-10 logarithm for GI, OGI, RGI, DGI, MGI, MRGI and NMGI algorithms with respect to five different values of η in Figure 1. As shown in Figure 1, all tested algorithms are convergent, and the OGI and DGI algorithm have faster convergence rates than GI and RGI ones as the former ones require less IT to achieve the termination criterion. In addition, the convergence speeds of the MGI, MRGI and NMGI algorithms are comparable because their IT are close. Also, the MRGI and NMGI algorithms are more efficient than the other tested ones from perspective of IT, and their advantages become more pronounced as the value of η decreases. Lastly, the NMGI algorithm requires slightly less IT then the MGI one, which coincides with the results listed in Table 2. These facts further confirm that the superiorities of the proposed algorithms for solving the Sylvester tensor equations.

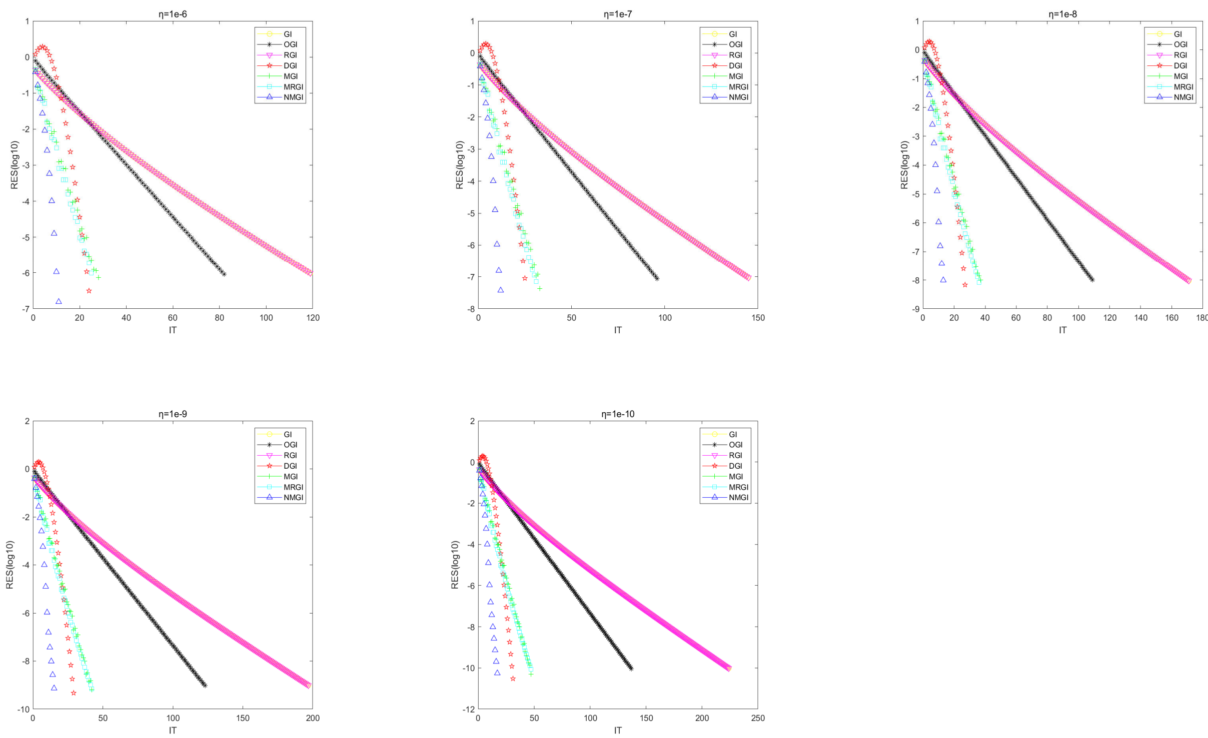

Figure 4. RES(log10) curves of seven algorithms for Example 6.2 with five values of η and $\rho = 5$.

Example 6.2. [6, 43] Consider the Sylvester tensor equation (1.1) with the following system matrices:

$$\begin{aligned}
&L_1 = L_2 = L_3 = 30,\\
&V_1 = \mathrm{triu}(\mathrm{rand}(L_1, L_1), 1) + \mathrm{diag}(\rho + \mathrm{diag}(\mathrm{rand}(L_1))),\\
&V_2 = \mathrm{triu}(\mathrm{rand}(L_2, L_2), 1) + \mathrm{diag}(\rho + \mathrm{diag}(\mathrm{rand}(L_2))),\\
&V_3 = \mathrm{triu}(\mathrm{rand}(L_3, L_3), 1) + \mathrm{diag}(\rho + \mathrm{diag}(\mathrm{rand}(L_3))),\\
&\mathcal{W} = \mathrm{tenrand}(L_1, L_2, L_3).
\end{aligned}$$

In Tables 3–5, we list the numerical results of the seven tested algorithms for solving Example 6.2 with five different values of η for $\rho = 2$, $\rho = 3$ and $\rho = 5$, respectively. The diagonal dominant

Table 2. Parameters and numerical results of seven tested algorithms for Example 6.1.

Method	η	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
	IT	345	415	484	554	623
GI	CPU	0.6205	0.7580	0.8855	1.0159	1.1721
$\gamma = 0.0574$	RES	9.96e-07	9.78e-08	9.93e-09	9.75e-10	9.89e-11
	IT	114	133	153	172	191
OGI	CPU	0.2017	0.2473	0.2832	0.3074	0.3499
$\gamma_{\text{opt}} = 0.1966$	RES	9.09e-07	9.57e-08	8.95e-09	9.43e-10	9.93e-11
	IT	314	377	440	503	566
RGI	CPU	0.6273	0.6771	0.7840	0.9219	1.0047
$(\alpha, \beta, \gamma) = (0.52, 0.32, 0.6839)$	RES	9.83e-07	9.87e-08	9.91e-09	9.95e-10	1.00e-10
	IT	179	212	244	276	309
DGI	CPU	0.3297	0.3835	0.4511	0.5080	0.5716
$\gamma_{\text{opt}} = 0.1268$	RES	9.82e-07	9.39e-08	9.64e-09	9.90e-10	9.47e-11
	IT	46	55	65	75	84
MGI	CPU	0.1809	0.2199	0.2524	0.2968	0.3301
$\gamma = 0.28$	RES	8.46e-07	9.86e-08	9.05e-09	8.30e-10	9.66e-11
	IT	38	46	55	63	71
MRGI	CPU	0.1466	0.1742	0.2094	0.2407	0.2703
$(\alpha, \beta, \gamma) = (0.48, 0.27, 3.6)$	RES	8.92e-07	9.72e-08	7.41e-09	7.93e-10	8.31e-11
	IT	44	51	59	66	73
NMGI	CPU	0.1702	0.1991	0.2299	0.2562	0.2719
$\gamma = 0.178$	RES	8.25e-07	9.19e-08	7.66e-09	8.70e-10	9.83e-11

degrees of the system matrices V_t $(t = 1, 2, 3)$ in this example become larger with ρ. According to the numerical results in Tables 3–5, we can conclude that all tested algorithms are convergent and feasible for all cases, and their IT and CPU time are increasing with the decreasing of η. In the meanwhile, the numerical performances of the GI and the RGI algorithms are almost the same, and this is also valid for the MGI and MRGI algorithms. And the GI algorithm with the optimal parameter (OGI) has advantages over the GI and RGI ones because the IT and CPU time of former are nearly half of those for latter ones. In addition, the proposed DGI and NMGI algorithms have better numerical performances than the other ones especially for IT, and their advantages become more evident with the decreases of η. Specially, the IT of the DGI algorithm is less than a seventh and a quarter of those for GI, RGI and OGI ones, respectively. And the IT of the NMGI algorithm is less than a half of those for the MGI and MRGI ones. Comparing the results in Tables 3–5, we also observe that numerical performances of all tested algorithms become better as the value of ρ becomes large, whereas the superiorities of the proposed DGI and NMGI algorithms to the other tested methods diminish. Also, the NMGI algorithm performs the best among the tested algorithms from the point of view of IT, while the CPU time of DGI and NMGI algorithms are close. The reason is that each iteration of the NMGI algorithm requires more basic arithmetic operations than the DGI one, which is accordance with the results in Table 1. Finally, The IT of GI, OGI, RGI and MGI, MRGI algorithms change sharply and change in moderate with respect to η, respectively, while those of the proposed DGI and NMGI algorithms change very little as η decreases. This reveals that

Table 3. Parameters and numerical results of seven tested algorithms for Example 6.2 with $\rho = 2$.

Method	η	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
	IT	1122	1394	1672	1953	2237
GI	CPU	3.8111	4.4887	5.4219	6.3403	7.1986
$\gamma = 0.0025$	RES	9.92e-07	9.98e-08	9.94e-09	9.95e-10	9.94e-11
	IT	667	779	892	1004	1117
OGI	CPU	2.1485	2.5084	2.8898	3.2480	3.6112
$\gamma_{\text{opt}} = 0.0061$	RES	9.92e-07	9.99e-08	9.87e-09	9.94e-10	9.82e-11
	IT	1123	1397	1675	1956	2240
RGI	CPU	3.7252	4.5817	5.4901	6.4073	7.3515
$(\alpha, \beta, \gamma) = (0.65, 0.34, 0.0227)$	RES	1.00e-06	9.92e-08	9.92e-09	9.96e-10	9.99e-11
	IT	47	50	53	55	58
DGI	CPU	0.1661	0.1789	0.1892	0.1894	0.2114
$\gamma_{\text{opt}} = 0.0512$	RES	9.41e-07	7.01e-08	5.00e-09	8.43e-10	5.22e-11
	IT	184	228	273	319	365
MGI	CPU	1.7456	2.1928	2.6830	3.0465	3.4893
$\gamma = 0.0102$	RES	9.82e-07	9.82e-08	9.89e-09	9.77e-10	9.87e-11
	IT	184	228	273	319	365
MRGI	CPU	1.2888	1.5902	1.8913	2.1898	2.5700
$(\alpha, \beta, \gamma) = (2/3, 1/3, 0.0918)$	RES	9.82e-07	9.82e-08	9.89e-09	9.77e-10	9.87e-11
	IT	19	22	25	28	31
NMGI	CPU	0.1434	0.1722	0.1790	0.2015	0.2223
$\gamma = 0.042$	RES	5.72e-07	5.54e-08	6.16e-09	6.48e-10	6.54e-11

the proposed DGI and NMGI algorithms are insensitive to η, and have higher stabilities than the other ones since the variation amplitudes of their IT are smaller.

To more clear illustrate the advantages of the OGI, DGI and NMGI algorithms, for five different values of η, we depict the RES(log10) curves of seven tested algorithms with $\rho = 2$, $\rho = 3$ and $\rho = 5$ in Figures 2–4, respectively. Here, we adopt the parameters in Tables 3–5 for the algorithms. By observation, we find that the curves for GI and RGI algorithms are almost coincident, which implies that their IT are almost the same and is in accordance with the results in Tables 3–5. And it also holds for MGI and MRGI ones. Additionally, the OGI algorithm requires less IT than the GI and RGI ones. Also, the proposed DGI and NMGI algorithms perform better than the other ones , and their advantages gradually become more pronounced when η becomes small. Lastly, the NMGI algorithm has the fastest convergence speed in all algorithms as it needs the least IT for all cases. It also can be seen that the curves of the DGI algorithm increase first, and then reduce. These results coincide with those in Tables 3–5.

Example 6.3. [4] We compute the approximate solutions of the Sylvester tensor equation (1.1) arises from the convection-diffusion equation

$$\begin{aligned} &-e\Phi h + f^T \Psi h = q, \quad \text{in } \Omega = [0,1] \times [0,1] \times [0,1], \\ &e = 0, \quad \text{on } \partial\Omega. \end{aligned}$$

According to a standard finite difference discretization on equidistant nodes and a second

Table 4. Parameters and numerical results of seven tested algorithms for Example 6.2 with $\rho = 3$.

Method	η	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
	IT	370	460	551	644	738
GI	CPU	1.2373	1.5353	1.8046	2.0976	2.4110
$\gamma = 0.0022$	RES	9.81e-07	9.78e-08	9.97e-09	9.94e-10	9.81e-11
	IT	234	274	313	352	392
OGI	CPU	0.8120	0.9414	1.0729	1.2680	1.4263
$\gamma_{\text{opt}} = 0.0051$	RES	9.98e-07	9.59e-08	9.77e-09	9.95e-10	9.56e-11
	IT	370	460	551	644	738
RGI	CPU	1.2357	1.5303	1.8648	2.2038	2.5286
$(\alpha, \beta, \gamma) = (2/3, 1/3, 0.0201)$	RES	9.81e-07	9.78e-08	9.97e-09	9.94e-10	9.81e-11
	IT	36	38	41	43	45
DGI	CPU	0.1360	0.1391	0.1555	0.1629	0.1667
$\gamma_{\text{opt}} = 0.0263$	RES	6.08e-07	8.72e-08	3.70e-09	4.28e-10	4.77e-11
	IT	67	82	98	114	131
MGI	CPU	0.4820	0.5797	0.6829	0.8253	0.9111
$\gamma = 0.00825$	RES	8.81e-07	9.38e-08	9.48e-09	9.90e-10	9.12e-11
	IT	67	82	98	115	131
MRGI	CPU	0.4931	0.6047	0.7224	0.8474	0.9413
$(\alpha, \beta, \gamma) = (2/3, 1/3, 0.074)$	RES	8.73e-07	9.69e-08	9.93e-09	9.10e-10	9.72e-11
	IT	17	19	20	21	22
NMGI	CPU	0.1299	0.1440	0.1496	0.1532	0.1557
$\gamma = 0.023$	RES	5.89e-07	2.07e-08	3.43e-09	5.36e-10	8.20e-11

order convergent scheme, we can derive a Sylvester tensor equation (1.1) with

$$V_i = \begin{pmatrix} 2vh^{-2} + \frac{3}{4}ch^{-1} & -vh^{-2} - \frac{5}{4}ch^{-1} & \frac{1}{4}ch^{-1} & & \\ -vh^{-2} + \frac{1}{4}ch^{-1} & 2vh^{-2} + \frac{3}{4}ch^{-1} & -ch^{-2} - \frac{5}{4}ch^{-1} & \frac{1}{4}ch^{-1} & \\ \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & -ch^{-2} - \frac{5}{4}ch^{-1} & \\ 0 & \cdots & -vh^{-2} + \frac{1}{4}ch^{-1} & 2vh^{-2} + \frac{3}{4}ch^{-1} & \end{pmatrix}_{n\times n}, \; i = 1, 2, 3.$$

We adopt $h = \frac{1}{n+1}$ as in [43], then

$$V_i = \frac{v}{h^2}\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} + \frac{c}{4h}\begin{pmatrix} 3 & -5 & 1 & & \\ 1 & 3 & -5 & \ddots & \\ & \ddots & \ddots & \ddots & 1 \\ & & 1 & 3 & -5 \\ & & & 1 & 3 \end{pmatrix}, \; i = 1, 2, 3,$$

Table 5. Parameters and numerical results of seven tested algorithms for Example 6.2 with $\rho = 5$.

Method	η	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
	IT	120	146	172	198	225
GI	CPU	0.4015	0.4740	0.5836	0.6739	0.7377
$\gamma = 0.0017$	RES	9.77e-07	9.35e-08	9.40e-09	9.80e-10	9.63e-11
	IT	83	97	110	124	138
OGI	CPU	0.2816	0.3293	0.3738	0.4094	0.4686
$\gamma_{\text{opt}} = 0.0036$	RES	9.08e-07	8.69e-08	9.84e-09	9.42e-10	9.02e-11
	IT	120	146	172	198	225
RGI	CPU	0.4057	0.4929	0.5794	0.6614	0.7560
$(\alpha, \beta, \gamma) = (2/3, 1/3, 0.0157)$	RES	9.77e-07	9.35e-08	9.40e-09	9.80e-10	9.63e-11
	IT	25	26	28	30	32
DGI	CPU	0.0904	0.0934	0.1041	0.1151	0.1221
$\gamma_{\text{opt}} = 0.0109$	RES	3.06e-07	8.95e-08	6.75e-09	4.68e-10	3.01e-11
	IT	29	34	38	43	48
MGI	CPU	0.2056	0.2434	0.2727	0.3103	0.3389
$\gamma = 0.00535$	RES	9.82e-07	4.26e-08	9.97e-09	6.53e-10	5.06e-11
	IT	26	32	37	43	48
MRGI	CPU	0.1909	0.2324	0.2707	0.3092	0.3631
$(\alpha, \beta, \gamma) = (0.66, 0.33, 0.0466)$	RES	9.89e-07	7.15e-08	8.28e-09	6.89e-10	8.80e-11
	IT	12	13	14	16	18
NMGI	CPU	0.0814	0.0960	0.1068	0.1185	0.1354
$\gamma = 0.009$	RES	3.10e-07	3.72e-08	9.74e-09	7.33e-10	5.51e-11

$$\mathcal{W} = \text{tenrand}(n, n, n).$$

Let $v = c = 1$. Numerical results of the seven tested algorithms for $n = 3$ and $n = 6$ are reported in Table 6 and Table 7, respectively. By comparing the results in Tables 6–7, it is observed that all tested algorithms are convergent for all cases. When n becomes large, the IT and CPU time of all tested algorithms increase. And apart from the proposed DGI and NMGI algorithms, those of the remaining ones grow tenfold from $n = 3$ to $n = 6$, i.e., they change violently with n. While the convergence speeds of the DGI and NMGI algorithms change in moderate and mild style, respectively, with n. This indicates that the DGI and NMGI algorithms are more stable than the other ones. Meanwhile, numerical performances of the GI and RGI algorithms are very close, and it also holds for MGI and MRGI ones. Moreover, GI and RGI algorithms are less efficient than the OGI one, and IT and CPU time of the latter are nearly half of those for the former ones, which reveals that the GI algorithm with the optimal parameter can improve its computational efficiency. Also, the DGI and NMGI algorithms are much better than the other ones, because their IT and CPU time are far less than the other ones. And the NMGI algorithm is the most efficient among the tested algorithms as it requires the least IT and CPU time. Specially, IT of the DGI algorithm is nearly a seventh of, a twentieth of and a fortieth of those for MGI and MRGI, OGI, and GI and RGI ones, respectively, for $n = 6$.

Table 6. Parameters and numerical results of seven tested algorithms for Example 6.3 with $n = 3$.

Method	η	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
	IT	513	599	686	772	858
GI	CPU	0.9649	1.1263	1.2878	1.4423	1.6166
$\gamma = 9.0100e-05$	RES	9.93e-07	9.98e-08	9.76e-09	9.80e-10	9.85e-11
	IT	260	304	347	391	434
OGI	CPU	0.4938	0.5578	0.6563	0.7363	0.8194
$\gamma_{\text{opt}} = 1.7604e-04$	RES	9.83e-07	9.59e-08	9.86e-09	9.62e-10	9.89e-11
	IT	513	600	686	772	858
RGI	CPU	0.9283	1.0888	1.2324	1.3985	1.5492
$(\alpha, \beta, \gamma) = (0.66, 0.33, 8.1098e-04)$	RES	9.96e-07	9.75e-08	9.79e-09	9.84e-10	9.89e-11
	IT	42	49	56	63	69
DGI	CPU	0.0779	0.0916	0.1091	0.1175	0.1338
$\gamma = 2.7073e-04$	RES	8.87e-07	8.34e-08	7.85e-09	7.38e-10	9.74e-11
	IT	94	111	126	154	171
MGI	CPU	0.3716	0.4303	0.4962	0.6086	0.6715
$\gamma = 3.2525e-04$	RES	8.73e-07	8.23e-08	8.86e-09	8.60e-10	7.28e-11
	IT	94	111	126	156	173
MRGI	CPU	0.3569	0.4242	0.4828	0.6000	0.6624
$(\alpha, \beta, \gamma) = (2/3, 1/3, 0.00293)$	RES	9.03e-07	7.13e-08	9.77e-09	6.50e-10	5.24e-11
	IT	17	20	23	26	29
NMGI	CPU	0.0654	0.0759	0.0891	0.1022	0.1118
$\gamma = 0.0005$	RES	7.98e-07	7.17e-08	6.55e-09	6.03e-10	5.59e-11

Finally, although IT of the DGI algorithm is 2.5 times of that of the NMGI one for $n = 3$, their CPU time are close, and the latter consumes slight less CPU time. The reason is that the basic arithmetic operation for each iteration of the DGI algorithm is less than that of the NMGI one as reported in Table 1.

Figures 5 and 6 depict the curves of the RES in base-10 logarithm versus the IT with five different values of η for $n = 3$ and $n = 6$, respectively. Here, we adopt the parameters in Tables 6–7 for the algorithms. The comparisons in Figures 5–6 show that the DGI and NMGI algorithms have higher computational efficiencies than the other ones because they require less IT to achieve the stopping criterion, and their advantages become evident when n becomes large or η becomes small. Also, the NMGI algorithm outperforms the DGI one in terms of convergence rate. Additionally, apart from the MGI and MRGI algorithms, the remaining tested algorithms have the global linear convergence rate behaviors with respect to RES(log10) of $\{\mathcal{Y}^l\}$. Lastly, the curves of the GI and RGI algorithms are almost overlap, which implies that their convergence speeds are almost the same. And this also holds for the MGI and MRGI ones.

In summary, diagonal substitution and new update techniques applied in the DGI and NMGI algorithms can ameliorate the numerical performances of the GI and MGI ones in [6]. Furthermore, the proposed DGI and NMGI algorithms are superior to the other ones from the point of view of computing efficiency.

Table 7. Parameters and numerical results of seven tested algorithms for Example 6.3 with $n = 6$.

Method	η	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
	IT	5429	6354	7280	8205	9131
GI	CPU	9.8370	11.6008	13.1316	15.0639	16.4730
$\gamma = 8.4088e-06$	RES	9.98e-07	9.99e-08	9.98e-09	1.00e-09	9.99e-11
	IT	2717	3180	3644	4107	4570
OGI	CPU	4.9146	5.6784	6.7422	7.3072	8.4561
$\gamma_{\text{opt}} = 1.6783e-05$	RES	9.99e-07	1.00e-07	9.96e-09	9.96e-10	9.97e-11
	IT	5430	6356	7281	8207	9133
RGI	CPU	9.7881	11.2032	12.9285	14.6789	16.4045
$(\alpha,\beta,\gamma) = (0.66, 0.33, 7.5686e-05)$	RES	9.98e-07	9.98e-08	9.99e-09	9.99e-10	9.98e-11
	IT	134	157	179	202	225
DGI	CPU	0.2397	0.2815	0.3227	0.3639	0.4060
$\gamma_{\text{opt}} = 3.0977e-05$	RES	9.77e-07	9.34e-08	9.89e-09	9.46e-10	9.04e-11
	IT	914	1069	1226	1383	1540
MGI	CPU	3.5202	4.1023	4.7366	5.3219	5.9675
$\mu = 3.329e-05$	RES	9.80e-07	9.92e-08	9.79e-09	9.72e-10	9.63e-11
	IT	915	1071	1227	1383	1538
MRGI	CPU	3.5106	4.0786	4.6462	5.2533	5.9009
$(\alpha,\beta,\gamma) = (2/3, 1/3, 2.99e-04)$	RES	9.89e-07	9.90e-08	9.88e-09	9.81e-10	9.88e-11
	IT	29	34	38	43	47
NMGI	CPU	0.1131	0.1329	0.1567	0.1685	0.1855
$\gamma = 0.0001$	RES	7.35e-07	6.19e-08	8.55e-09	7.19e-10	9.93e-11

7. Numerical experiments for coupled Sylvester tensor equations

This section provides a numerical example to validate the feasibilities and advantages of the proposed DGI and NMGI algorithms for the coupled Sylvester tensor equation (4.1) with respect to IT and CPU time.

All numerical experiments are computed in MATLAB (version R2018a) on a personal computer with Intel (R) Core (TM) i9-10900 2.81 GHz and 32.0 GB RAM. The initial tensors are taken to be $\mathcal{Y}^0 = \mathcal{Y}_2^0 = \mathcal{Y}_3^0 = \mathcal{Y}_4^0 = \mathcal{Y}_5^0 = \mathcal{Y}_6^0 = 10^{-6} \cdot tenones(L_1, L_2, L_3)$, whose all elements are 1. All iterations are terminated once the current residual (RES) satisfies

$$\text{RES} = \frac{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^l\times_1 V_1-\mathcal{Y}^l\times_2 V_2-\mathcal{Y}^l\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^l\times_1 T_1-\mathcal{Y}^l\times_2 T_2-\mathcal{Y}^l\times_3 T_3\|^2}}{\sqrt{\|\mathcal{W}_1-\mathcal{Y}^0\times_1 V_1-\mathcal{Y}^0\times_2 V_2-\mathcal{Y}^0\times_3 V_3\|^2+\|\mathcal{W}_2-\mathcal{Y}^0\times_1 T_1-\mathcal{Y}^0\times_2 T_2-\mathcal{Y}^0\times_3 T_3\|^2}} \\ \leq \eta,$$

with $\eta > 0$, or the IT reaches the maximal number of iteration steps $l_{\max} = 10000$. According to Theorems 4.1 and 4.3, the parameters of the GI and DGI algorithms are taken as

$$\gamma_{OGI} = \frac{12}{\bar{\lambda}_{\max} + \bar{\lambda}_{\min}},\ \gamma_{DGI} = \frac{12}{\overline{Re}_{\min} + \overline{Re}_{\max}} \text{ or } \frac{6\overline{Re}_{\min}}{\overline{Re}_{\min}^2 + \overline{Im}_1^2},$$

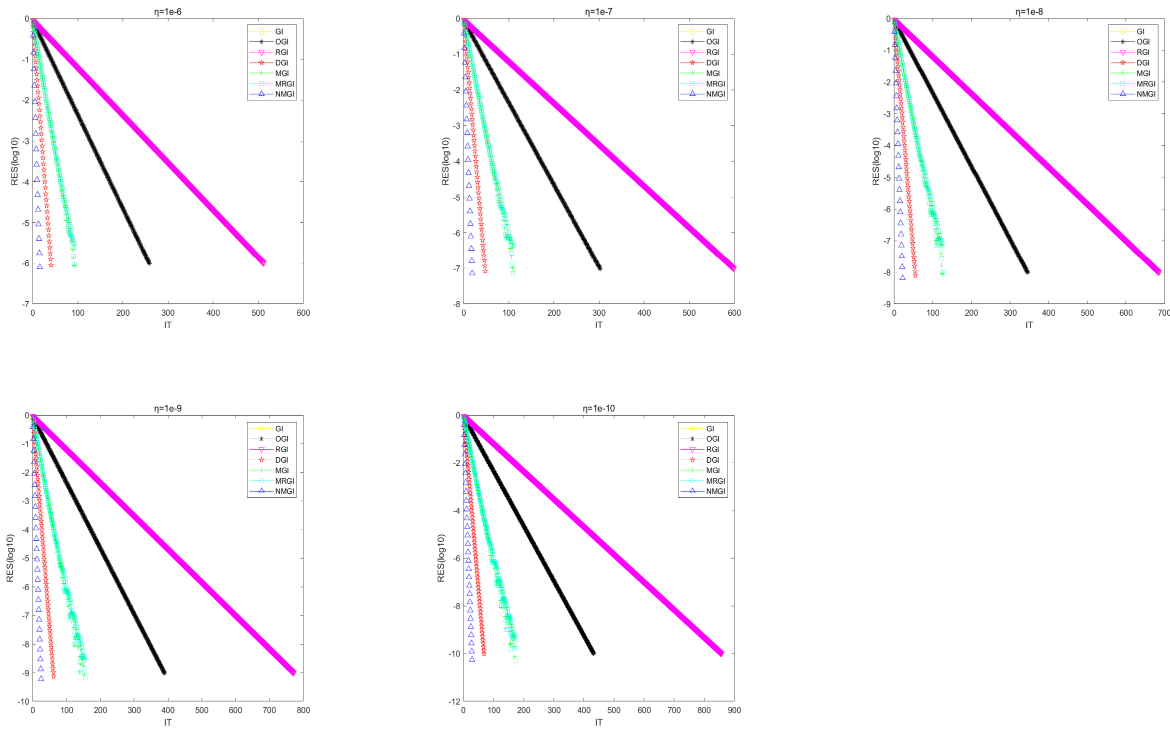

Figure 5. RES(log10) curves of seven algorithms for Example 6.3 with five values of η and $n = 3$.

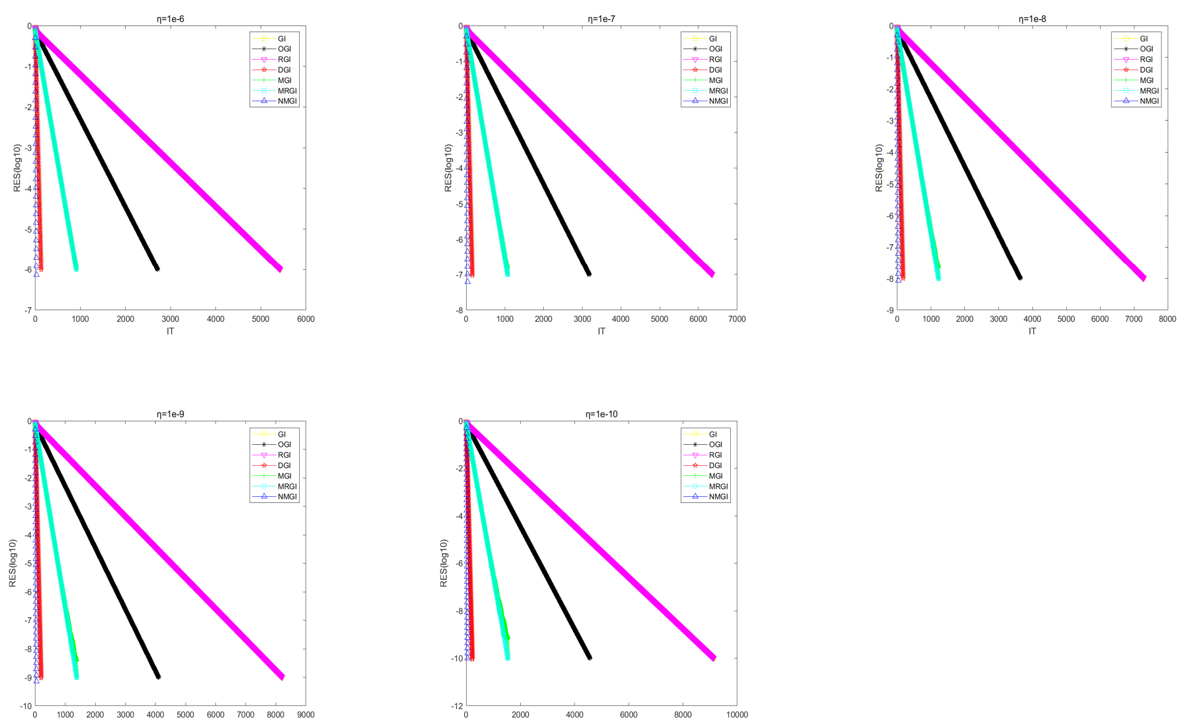

Figure 6. RES(log10) curves of seven algorithms for Example 6.3 with five values of η and $n = 6$.

where $\overline{\lambda}_{\max}, \overline{\lambda}_{\min}$ and $\overline{Re}_{\min}, \overline{Re}_{\max}, \overline{Im}_1$ are defined as in Theorem 4.1 and Theorem 4.2, respectively. In addition, the parameters adopted in the MGI and NMGI algorithms are the experimental optimal ones that minimize their IT for $\eta = 10^{-2}$.

Example 7.1. Consider the coupled Sylvester tensor equation (4.1), where

$$V_i = \begin{pmatrix} \frac{2v}{h^2}+\frac{3c}{4h} & -(\frac{v}{h^2}+\frac{5c}{4h}) & \frac{c}{4h} & & & \\ \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{4h} & -(\frac{v}{h^2}+\frac{5c}{4h}) & \ddots & & \\ & \ddots & \ddots & \ddots & \frac{c}{4h} \\ & & \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{4h} & -(\frac{v}{h^2}+\frac{5c}{4h}) \\ & & & \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{4h} \end{pmatrix}_{n\times n},$$

$$T_i = \begin{pmatrix} \frac{2v}{h^2}+\frac{c}{4h} & -(\frac{v}{h^2}+\frac{3c}{4h}) & \frac{c}{8h} & & & \\ \frac{c}{5h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{c}{4h} & -(\frac{v}{h^2}+\frac{3c}{4h}) & \ddots & & \\ & \ddots & \ddots & \ddots & \frac{c}{8h} \\ & & \frac{c}{5h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{c}{4h} & -(\frac{v}{h^2}+\frac{3c}{4h}) \\ & & & \frac{c}{5h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{c}{4h} \end{pmatrix}_{n\times n}, \quad i = 1,2,3,$$

$$\mathcal{W}_1 = \text{tenrand}(n,n,n),\ \mathcal{W}_2 = \mathcal{W}_1,$$

with $h = \frac{1}{n+1}$, $v = 5$, $c = 1$ and $n = 6$. The system matrices V_i and T_i $(i = 1,2,3)$ stem from Example 6.3 with few modifications.

The parameters, IT, CPU time and RES of the GI, DGI, MGI and NMGI algorithms for Example 7.1 with respect to four different values of η are listed in Table 8. Comparing the numerical results of Table 8, we see that all tested algorithms can successfully compute approximate solutions satisfying the prescribed stopping criterion, and their IT and CPU time increase gradually with decreasing of η. Meanwhile, the proposed NMGI algorithm performs far better than the GI, DGI and MGI ones in terms of both the IT and CPU time. And the IT and CPU time of the DGI and NMGI algorithms are nearly a twentieth of those for the GI and MGI ones, respectively, as $\eta = 10^{-2}$. Additionally, the proposed DGI algorithm is more stable than the GI one, due to the fact that the variational range of the IT of the former is smaller than that of the latter. And the stability of the proposed NMGI algorithm is the highest among the tested algorithms in view of IT. Finally, the numerical results in Table 8 show that the diagonal substitution technique and new update strategy applied in the DGI the NMGI algorithms, respectively, can improve the convergence speeds of the GI and MGI ones effectively, and the DGI and NMGI algorithms have higher computational efficiencies than the GI and MGI ones.

Table 8. Parameters and numerical results of the GI, DGI, MGI and NMGI algorithms for Example 7.1.

Method	η	5×10^{-1}	10^{-1}	5×10^{-2}	10^{-2}
	IT	83	390	525	1047
GI	CPU	0.3101	1.4367	1.8905	3.9371
$\gamma_{\rm opt}=5.2346e-07$	RES	0.4988	9.99e-02	4.99e-02	1.00e-02
	IT	6	21	28	55
DGI	CPU	0.0214	0.0765	0.0997	0.2056
$\gamma_{\rm opt}=9.5133e-07$	RES	0.4633	9.90e-02	4.89e-02	9.90e-03
	IT	15	67	90	178
MGI	CPU	0.1529	0.7172	0.9557	1.8489
$\gamma=1.80e-06$	RES	0.4988	9.94e-02	4.96e-02	1.00e-02
	IT	2	3	4	7
NMGI	CPU	0.0125	0.0226	0.0306	0.0596
$\gamma=6.00e-06$	RES	0.2776	8.17e-02	2.77e-02	9.80e-03

8. Numerical experiments for coupled Sylvester tensor equations with two unknowns

This section presents a numerical example to illustrate the feasibilities and superiorities of the proposed DGI and NMGI algorithms for the coupled Sylvester tensor equation with two unknowns (5.1) in terms of both IT and CPU time.

All tests are run by applying MATLAB (version R2018a) in a personal computer with Intel (R) Core (TM) i9-10900 2.81 GHz and 32.0 GB RAM. We take the initial tensors as $\mathcal{Y}^0=\mathcal{Y}_2^0=\mathcal{Y}_3^0=\mathcal{Y}_4^0=\mathcal{Y}_5^0=\mathcal{Y}_6^0=\mathcal{Z}^0=\mathcal{Z}_2^0=\mathcal{Z}_3^0=\mathcal{Z}_4^0=\mathcal{Z}_5^0=\mathcal{Z}_6^0=10^{-6}\cdot tenones(L_1,L_2,L_3)$, whose all elements are 1. All iterations are stopped once the RES satisfies

$$\mathrm{RES}=\frac{\sqrt{\|\mathcal{W}_1-\sum\limits_{j=1}^{3}\mathcal{Y}^l\times_j V_j-\sum\limits_{j=1}^{3}\mathcal{Z}^l\times_j P_j\|^2+\|\mathcal{W}_2-\sum\limits_{j=1}^{3}\mathcal{Y}^l\times_j T_j-\sum\limits_{j=1}^{3}\mathcal{Z}^l\times_j Q_j\|^2}}{\sqrt{\|\mathcal{W}_1-\sum\limits_{j=1}^{3}\mathcal{Y}^0\times_j V_j-\sum\limits_{j=1}^{3}\mathcal{Z}^0\times_j P_j\|^2+\|\mathcal{W}_2-\sum\limits_{j=1}^{3}\mathcal{Y}^0\times_j T_j-\sum\limits_{j=1}^{3}\mathcal{Z}^0\times_j Q_j\|^2}}\le\eta,$$

with $\eta>0$, or the IT exceeds $l_{\max}=10000$. We denote the latter case by "Invalid" and "–" in tables. According to Theorems 5.1 and 5.3, the parameters of the GI and DGI algorithms are taken as

$$\gamma_{\rm OGI}=\frac{12}{\widehat{\lambda}_{\max}+\widehat{\lambda}_{\min}},\ \gamma_{\rm DGI}=\frac{12}{\widehat{Re}_{\min}+\widehat{Re}_{\max}}\ \text{or}\ \frac{6\widehat{Re}_{\min}}{\widehat{Re}_{\min}^2+\widehat{Im}_1^2},$$

where $\widehat{\lambda}_{\max},\widehat{\lambda}_{\min}$ and $\widehat{Re}_{\min},\widehat{Re}_{\max},\widehat{Im}_1$ are defined as in Theorem 5.1 and Theorem 5.2, respectively. For the MGI and NMGI algorithms, their parameters are the experimentally found optimal ones which minimize their IT as $\eta=10^{-10}$.

Example 8.1. Consider the coupled Sylvester tensor equation with two unknowns (5.1), where

$$V_j = \begin{pmatrix} \frac{3v}{h^2}+\frac{5c}{4h} & -(\frac{v}{h^2}+\frac{5c}{4h}) & \frac{c}{4h} & & \\ \frac{c}{4h}-\frac{v}{h^2} & \frac{3v}{h^2}+\frac{5c}{4h} & -(\frac{v}{h^2}+\frac{5c}{4h}) & \ddots & \\ & \ddots & \ddots & \ddots & \frac{c}{4h} \\ & & \frac{c}{4h}-\frac{v}{h^2} & \frac{3v}{h^2}+\frac{5c}{4h} & -(\frac{v}{h^2}+\frac{5c}{4h}) \\ & & & \frac{c}{4h}-\frac{v}{h^2} & \frac{3v}{h^2}+\frac{5c}{4h} \end{pmatrix}_{n\times n},$$

$$P_j = \begin{pmatrix} \frac{2v}{h^2}+\frac{3c}{2h} & -(\frac{3v}{2h^2}+\frac{5c}{4h}) & \frac{c}{4h} & & \\ \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{2h} & -(\frac{3v}{2h^2}+\frac{5c}{4h}) & \ddots & \\ & \ddots & \ddots & \ddots & \frac{c}{4h} \\ & & \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{2h} & -(\frac{3v}{2h^2}+\frac{5c}{4h}) \\ & & & \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{2h} \end{pmatrix}_{n\times n},$$

$$T_j = \begin{pmatrix} \frac{2v}{h^2}+\frac{3c}{4h} & -(\frac{v}{h^2}+\frac{3c}{4h}) & \frac{c}{4h} & & \\ \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{4h} & -(\frac{v}{h^2}+\frac{3c}{4h}) & \ddots & \\ & \ddots & \ddots & \ddots & \frac{c}{4h} \\ & & \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{4h} & -(\frac{v}{h^2}+\frac{3c}{4h}) \\ & & & \frac{c}{4h}-\frac{v}{h^2} & \frac{2v}{h^2}+\frac{3c}{4h} \end{pmatrix}_{n\times n},$$

$$Q_j = \begin{pmatrix} \frac{3v}{h^2}+\frac{5c}{8h} & -(\frac{v}{h^2}+\frac{c}{h}) & \frac{c}{4h} & & \\ \frac{c}{8h}-\frac{v}{h^2} & \frac{3v}{h^2}+\frac{5c}{8h} & -(\frac{v}{h^2}+\frac{c}{h}) & \ddots & \\ & \ddots & \ddots & \ddots & \frac{c}{4h} \\ & & \frac{c}{8}-\frac{v}{h^2} & \frac{3v}{h^2}+\frac{5c}{8h} & -(\frac{v}{h^2}+\frac{c}{h}) \\ & & & \frac{c}{8h}-\frac{v}{h^2} & \frac{3v}{h^2}+\frac{5c}{8h} \end{pmatrix}_{n\times n}, \ j=1,2,3,$$

$$\mathcal{W}_1 = tenrand(n,n,n), \ \mathcal{W}_2 = \mathcal{W}_1,$$

with $h = \frac{1}{n+1}$, $v = c = 1$ and $n = 10$. The system matrices V_j, P_j, T_j and Q_j $(j = 1,2,3)$ come from Example 6.3 with few modifications as in Example 7.1.

In Table 9, we list the parameters, IT, CPU time and RES of the tested algorithms for

Table 9. Parameters and numerical results of the GI, DGI, MGI and NMGI algorithms for Example 8.1.

Method	η	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}
	IT	593	723	880	1045	1211
GI	CPU	6.4832	7.8923	9.4013	11.2406	12.9087
$\gamma_{\text{opt}} = 1.0108e-06$	RES	9.81e-07	9.96e-08	9.96e-09	9.88e-10	9.96e-11
	IT	265	330	398	468	538
DGI	CPU	2.9474	3.6393	4.2425	5.2666	5.7717
$\gamma_{\text{opt}} = 1.8293e-06$	RES	9.65e-07	9.95e-08	1.00e-08	9.77e-10	9.81e-11
	IT	127	158	193	226	262
MGI	CPU	4.1425	4.9977	5.8496	7.3428	8.6443
$\gamma = 2.96e-06$	RES	9.20e-07	1.00e-07	9.02e-09	9.73e-10	8.80e-11
	IT	68	85	103	120	138
NMGI	CPU	1.8510	2.4194	2.7994	3.4515	4.0439
$\gamma = 5.40e-06$	RES	9.77e-07	9.98e-08	9.27e-09	9.96e-10	9.42e-11

Example 8.1 with five different values of η. According to the numerical results in Table 9, we can conclude some observations: Firstly, all tested algorithms are convergent and feasible for all cases, and their IT and CPU time are increasing with the decreasing of η. Secondly, the IT of the GI algorithm change sharply with respect to η, which illustrates that its convergence behaviors are very sensitive to η. While the IT of the remaining algorithms change in moderate as η decreases, and the proposed NMGI algorithm is the most stable among the tested algorithms, because the variational range of IT of the NMGI algorithm is the smallest compared with other tested ones. Thirdly, the DGI algorithm has faster convergence speed than the GI one, and the IT and CPU time of the former are less than half of those for the latter. Fourthly, the proposed NMGI algorithm has better numerical performances and higher stability than the MGI one in terms of IT and CPU time, and its advantages become more evident with the decreases of η. Fifthly, the IT of the NMGI algorithm is far less than the DGI one, but there is no big difference on their CPU time. The reason is that each iteration of the NMGI algorithm requires more basic arithmetic operations than the DGI one, which is accordance with the results in Table 1. Lastly, the performance of the NMGI algorithm is optimal in all tested algorithms in view of IT and CPU time.

9. Conclusions

In this work, we establish two new GI-like algorithms to solve the (coupled) Sylvester tensor equations and their convergence properties. We first derive the optimal parameter and the corresponding optimal convergence factor of the GI algorithm [6] to further perfect its theories. To reduce the computational complexity of the GI algorithm, based on the hierarchical identification principle, we replace the system matrices in the GI algorithm by their diagonal parts and design the diagonal GI (DGI) algorithm for the (coupled) Sylvester tensor equations. By using the tensor stretching operator, classification and analysis, we derive the convergence condition and quasi-optimal parameter of the DGI algorithm. In addition, we apply the diagonal substitution technique and the new update strategy to the MGI algorithm, then construct the new MGI (NMGI) algorithm for the (coupled) Sylvester tensor equations. Meanwhile, the sufficient convergence condition of the NMGI algorithm is established by making use of the properties of the Frobenious norm of a tensor and n-mode product of a tensor and a matrix. Also, we compare the number of arithmetic operations of the tested algorithms. Finally, numerical experiments are performed to validate that the proposed algorithms are efficient, and outperform the GI, RGI, MGI and MRGI ones for the (coupled) Sylvester tensor equations in view of both IT and CPU time.

It is noteworthy that the convergent interval of the parameter γ and the (quasi-)optimal parameter of the NMGI algorithm have not been obtained at present, which are significant for the implementation of the NMGI algorithm. We will investigate these problems in our future work. Aside from that, convergence conditions of the DGI algorithm derived in this work are sufficient, then the necessary and sufficient conditions for the convergence of the DGI algorithm deserve to be further discussed.

Acknowledgments. We would like to express our sincere thanks to editor and anonymous reviewer for their valuable suggestions and constructive comments which greatly improved the presentation of this paper.

Author contributions. Fengqing Li: Writing-original draft; perform the numerical experiments. Zhengge Huang: Supervision; funding acquisition; methodology. Jingjing Cui: writing-review and editing. Xiuwen Zheng: writing-review and editing.

Data availability. Data will be made available on reasonable request.

Conflict of interest. The authors declare that they have no conflict of interests in this article.

References

[1] Z.-Z. Bai, *On Hermitian and skew-Hermitian splitting iteration methods for continuous Sylvester equations*, Journal of Computational Mathematics, 2011, 185–198.

[2] U. Baur and P. Benner, *Cross-Gramian based model reduction for data-sparse systems*, Electronic Transactions on Numerical Analysis, 2008, 31, 256–270.

[3] F. P. A. Beik, M. Najafi-Kalyani and L. Reichel, *Iterative Tikhonov regularization of tensor equations based on the Arnoldi process and some of its generalizations*, Applied Numerical Mathematics, 2020, 151, 425–447.

[4] F. P. A. Beik and F. S. Movahed and S. Ahmadi-Asl, *On the Krylov subspace methods based on tensor format for positive definite Sylvester tensor equations*, Numerical Linear Algebra with Applications, 2016, 23, 444–466.

[5] D. Calvetti and L. Reichel, *Application of ADI iterative methods to the restoration of noisy images*, SIAM Journal on Matrix Analysis and Applications, 1996, 17, 165–186.

[6] Z. Chen and L.-Z. Lu, *A gradient based iterative solutions for Sylvester tensor equations*, Mathematical Problems in Engineering, 2013, 2013, 819479.

[7] D.-L. Chu and V. Mehrmann, *CDisturbance decoupling for descriptor systems by state feedback*, SIAM Journal on Control and Optimization, 2000, 38, 1830–1858.

[8] M. Dehghan and M. Hajarian, *An iterative method for solving the generalized coupled Sylvester matrix equations over generalized bisymmetric matrices*, Applied Mathematical Modelling, 2010, 34, 639–654.

[9] F. Ding and T.-W. Chen, *Iterative least-squares solutions of coupled Sylvester matrix equations*, Systems & Control Letters, 2005, 54, 95–107.

[10] F. Ding and T.-W. Chen, *Hierarchical identification of lifted state-space models for general dual-rate systems*, IEEE Transactions on Circuits and Systems I: Regular Papers, 2005, 52, 1179–1187.

[11] F. Ding and T.-W. Chen, *On iterative solutions of general coupled matrix equations*, SIAM Journal on Control and Optimizatio, 2006, 44, 2269–2284.

[12] F. Ding and T.-W. Chen, *Hierarchical least squares identification methods for multivariable systems*, IEEE Transactions on Automatic control, 2005, 50, 397–402.

[13] F. Ding and T.-W. Chen, *Gradient based iterative algorithms for solving a class of matrix equations*, IEEE Transactions on Automatic Control, 2005, 50, 1216–1221.

[14] F. Ding, P.-X. Liu and J. Ding, *Iterative solutions of the generalized Sylvester matrix equations by using the hierarchical identification principle*, Applied Mathematics and Computation, 2008, 197, 41–50.

[15] G. H. Gloub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Universtiy Press, 3rd edtion, 1996.

[16] G. H. Golub, S. Nash and C. V. Loan, *A Hessenberg-Schur method for the problem $AX + XB = C$*, IEEE Transactions on Automatic Control, 1979, 24, 909–913.

[17] J. Heinen, *A technique for solving the extended discrete Lyapunov matrix equation*, IEEE Transactions on Automatic Control, 1972, 17, 156–157.

[18] Z.-G. Huang and J.-J. Cui, *Modified and accelerated relaxed gradient based iterative algorithms for the complex conjugate and transpose matrix equations*, Numerical Algorithms, 2024, 97, 1955–2009.

[19] A. Kaabi, A. Kerayechian and F. Toutounian, *A new version of successive approximations method for solving Sylvester matrix equations*, Applied Mathematics and Computation, 2007, 186, 638–645.

[20] A. Kaabi, F. Toutounian and A. Kerayechian, *Preconditioned Galerkin and minimal residual methods for solving Sylvester equations*, Applied Mathematics and Computation, 2006, 181, 1208–1214.

[21] Y.-F. Ke, *Finite iterative algorithm for the complex generalized Sylvester tensor equations*, Journal of Applied Analysis and Computation, 2020, 10, 972–985.

[22] T. G. Kolda, *Multilinear Operators for Higher-Order Decompositions*, Tech. rep., Sandia National Laboratories, Albuquerque, NM, and Livermore, CA, 2006.

[23] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Review, 2009, 51, 455–500.

[24] B.-W. Li, S. Tian and Y.-S. Sun, *Schur-decomposition for 3D matrix equations and its application in solving radiative discrete ordinates equations discretized by Chebyshev collocation spectral method*, Journal of Computational Physics, 2010, 229, 1198–1212.

[25] C.-X. Li and S.-L. Wu, *The SHSS preconditioner for saddle point problems*, Journal of Applied Analysis and Computation, 2023, 13, 3221–3230.

[26] H.-K. Li and R.-R. Li, *A note on the inversion of Sylvester matrices in control systems*, Mathematical Problems in Engineering, 2011, 2011, 609863.

[27] J.-H. Li, F. Ding and G.-W. Yang, *Maximum likelihood least squares identification method for input nonlinear finite impulse response moving average systems*, Mathematical and Computer Modelling, 2012, 55, 442–450.

[28] T. Li, Q.-W. Wang and X.-F. Zhang, *Gradient based iterative methods for solving symmetric tensor equations*, Numerical Linear Algebra with Applications, 2022, 29, e2414.

[29] Y.-Q. Lin, *Implicitly restarted global FOM and GMRES for nonsymmetric matrix equations and Sylvester equations*, Applied Mathematics and Computation, 2005, 167, 1004–1025.

[30] F. Saberi-Movahed, A. Tajaddini, M. Heyouni and L. Elbouyahyaoui, *Some iterative approaches for Sylvester tensor equations, Part I: A tensor format of truncated Loose Simpler GMRES*, Applied Numerical Mathematics, 2022, 172, 428–445.

[31] H. S. Najafi and A. Refahi Sheikhani, *Refinement methods for state estimation via Sylvester-Observer equation*, Advances in Numerical Analysis, 2011, 2011, 184314.

[32] Q. Niu, X. Wang and L.-Z. Lu, *A relaxed gradient based algorithm for solving Sylvester equations*, Asian Journal of Control, 2011, 13, 461–464.

[33] L.-Q. Qi, H.-B. Chen and Y.-N. Chen, *Tensor eigenvalues and their applications*, Springer, 2018, 39.

[34] D. C. Sorensen and Y.-K. Zhou, *Direct methods for matrix Sylvester and Lyapunov equations*, Journal of Applied Mathematics, 2023, 2023, 277–303.

[35] Q.-W. Wang, X.-J. Xu and X.-F. Duan, *Least squares solution of the quaternion Sylvester tensor equation*, Linear and Multilinear Algebra, 2021, 69, 104–130.

[36] W. Wang, F. Ding and J.-Y. Dai, *Maximum likelihood least squares identification for systems with autoregressive moving average noise*, Applied Mathematical Modelling, 2012, 36, 1842–1853.

[37] W.-L. Wang and C.-Q. Song, *Iterative algorithms for discrete-time periodic Sylvester matrix equations and its application in antilinear periodic system*, Applied Numerical Mathematics, 2021, 168, 251–273.

[38] W.-L. Wang, C.-Q. Song and W.-L. Wang, *A new BCR method for coupled operator equations with submatrix constraint*, Journal of Applied Analysis and Computation, 2024, 14, 2002–2036.

[39] Y.-N. Wu and M.-L. Zeng, *On ADMM-based methods for solving the nearness symmetric solution of the system of matrix equations*, Journal of Applied Analysis and Computation, 2021, 11, 227–241.

[40] Y.-J. Xie and C.-F. Ma, *The accelerated gradient based iterative algorithm for solving a class of generalized Sylvester-transpose matrix equation*, Applied Mathematics and Computation, 2016, 273, 1257–1269.

[41] T.-X. Yan and C.-F. Ma, *An iterative algorithm for solving a class of generalized coupled Sylvester-transpose matrix equations over bisymmetric or skew-anti-symmetric matrices*, Journal of Applied Analysis and Computation, 2020, 10, 1282–1310.

[42] H.-M. Zhang, *Tensor Analysis: Spectral Theory and Special Tensors*, SIAM, 2017.

[43] X.-F. Zhang and Q.-W. Wang, *On RGI algorithm for solving Sylvester tensor equations*, Taiwanese Journal of Mathematics, 2022, 26, 501–519.

[44] B. Zhou, J. Lam and G.-R. Duan, *Gradient-based maximal convergence rate iterative method for solving linear matrix equations*, International Journal of Computer Mathematics, 2010, 87, 515–527.

[45] D.-M. Zhou, G.-L. Chena and Q.-Y. Cai, *On modified HSS iteration methods for continuous Sylvester equations*, Applied Mathematics and Computation, 2015, 263, 84–93.

Received December 2024; Accepted August 2025; Available online September 2025.