Computing the extremal nonnegative solutions of the M-tensor equation with a nonnegative right-side vector

Chun-Hua Guo

Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

---

Abstract

We consider the tensor equation whose coefficient tensor is a nonsingular M-tensor and whose right-side vector is nonnegative. Such a tensor equation may have a large number of nonnegative solutions. It is already known that the tensor equation has a maximal nonnegative solution and a minimal nonnegative solution (called extremal solutions collectively). However, the existing proofs do not show how the extremal solutions can be computed. The existing numerical methods can find one of the nonnegative solutions, without knowing whether the computed solution is an extremal solution. In this paper, we present new proofs for the existence of extremal solutions. Our proofs are much shorter than existing ones and more importantly they give numerical methods that can compute the extremal solutions. Linear convergence of these numerical methods is also proved under mild assumptions. Some of our discussions also allow the coefficient tensor to be a Z-tensor or allow the right-side vector to have some negative elements.

MSC: 65F10; 65H10; 15A69

Keywords: Z-tensor; M-tensor; Maximal nonnegative solution; Minimal nonnegative solution; Iterative methods; Linear convergence

---

1. Introduction

We consider the tensor equation

$$\mathcal{A}x^{m-1} = b, \tag{1}$$

where $x, b \in \mathbb{R}^n$ and $\mathcal{A}$ is a real-valued mth-order n-dimensional tensor that has the form

$$\mathcal{A} = (a_{i_1 i_2 \ldots i_m}), \quad a_{i_1 i_2 \ldots i_m} \in \mathbb{R}, \quad 1 \leq i_1, i_2, \ldots, i_m \leq n,$$

Email address: chun-hua.guo@uregina.ca (Chun-Hua Guo)

and $\mathcal{A}x^{m-1} \in \mathbb{R}^n$ has elements

$$(\mathcal{A}x^{m-1})_i = \sum_{i_2,\dots,i_m=1}^{n} a_{ii_2\dots i_m} x_{i_2} \cdots x_{i_m}, \quad i = 1, 2, \dots, n.$$

Later on, we will also use $\{x_k\}$ to denote a sequence of vectors. The ith element of a vector x will then be denoted by $(x)_i$ (instead of x_i) to avoid confusion. The element $a_{i_1 i_2 \dots i_m}$ of the tensor $\mathcal{A}$ will also be denoted by $\mathcal{A}_{i_1 i_2 \dots i_m}$ or $\mathcal{A}(i_1, i_2, \dots, i_m)$. The tensor equation (1) is also called a multilinear system of equations. It appears in many applications including data mining, numerical partial differential equations, and tensor complementarity problems; see [11] and the references cited therein.

We denote the set of all real-valued mth-order n-dimensional tensors by $\mathbb{R}^{[m,n]}$. A tensor $\mathcal{A} = (a_{i_1 i_2 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is called semi-symmetric [16] if $a_{ij_2\dots j_m} = a_{ii_2\dots i_m}$, $1 \le i \le n$, $j_2 \dots j_m$ is any permutation of $i_2 \dots i_m$, $1 \le i_2, \dots, i_m \le n$. For any tensor $\mathcal{A} = (a_{i_1 i_2 \dots i_m})$, we have a semi-symmetric tensor $\bar{\mathcal{A}} = (\bar{a}_{i_1 i_2 \dots i_m})$ defined by

$$\bar{a}_{i_1 i_2 \dots i_m} = \frac{1}{(m-1)!} \sum_{j_2 \dots j_m} a_{i_1 j_2 \dots j_m},$$

where $j_2 \dots j_m$ is any permutation of $i_2 \dots i_m$. Then $\mathcal{A}x^{m-1} = \bar{\mathcal{A}}x^{m-1}$ for all $x \in \mathbb{R}^n$, and $(\mathcal{A}x^{m-1})' = (m-1)\bar{\mathcal{A}}x^{m-2}$ (see [12, 16]), where for any $\mathcal{A} \in \mathbb{R}^{[m,n]}$, the (i,j) element of $\mathcal{A}x^{m-2} \in \mathbb{R}^{n\times n}$ is defined by

$$(\mathcal{A}x^{m-2})_{ij} = \sum_{i_3,\dots,i_m=1}^{n} a_{iji_3\dots i_m} x_{i_3} \cdots x_{i_m}.$$

We call $\lambda \in \mathbb{R}$ an eigenvalue and $x \in \mathbb{R}^n \setminus \{0\}$ a corresponding eigenvector of $\mathcal{A}$ if [18]

$$\mathcal{A}x^{m-1} = \lambda x^{[m-1]},$$

where for any positive real number r, $x^{[r]} \in \mathbb{R}^n$ is given by $(x^{[r]})_i = (x_i)^r$, $i = 1, 2, \dots, n$. The spectral radius of $\mathcal{A}$ is the maximum modulus of its eigenvalues, and is denoted by $\rho(\mathcal{A})$.

Let $[n] = \{1, 2, \dots, n\}$. For $x, y \in \mathbb{R}^n$, we write $x \ge y$ if $x_i \ge y_i$ for all $i \in [n]$, and write $x > y$ if $x_i > y_i$ for all $i \in [n]$. If $x \ge 0$, we say x is nonnegative; if $x > 0$ we say x is positive. The set $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \ge 0\}$ will be used frequently. A solution x of (1) is said to be a maximal nonnegative solution if $x \ge y$ for every solution $y \ge 0$; a solution x of (1) is said to be a minimal nonnegative solution if $0 \le x \le y$ for every solution $y \ge 0$. A tensor $\mathcal{A}$ is said to be nonnegative, denoted by $\mathcal{A} \ge 0$, if all its elements are nonnegative. The identity tensor $\mathcal{I} \in \mathbb{R}^{[m,n]}$ is such that its diagonal elements are all ones and its off-diagonal elements are all zeros, i.e., $\mathcal{I}_{ii\dots i} = 1$ for $i \in [n]$ and $\mathcal{I}_{i_1 i_2 \dots i_m} = 0$ elsewhere.

Definition 1. [3] A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called a Z-tensor if its off-diagonal elements are nonpositive. If $\mathcal{A}$ can be written as $\mathcal{A} = s\mathcal{I} - \mathcal{B}$ with $\mathcal{B} \geq 0$ and $s > \rho(\mathcal{B})$, then the tensor $\mathcal{A}$ is called a nonsingular M-tensor.

For $\mathcal{B} \geq 0$, $\rho(\mathcal{B})$ can be found by the power method in [15], the Newton–Noda iteration in [12], and the iterative method in [20]. Thus, one can determine whether a Z-tensor is a nonsingular M-tensor using the definition. It is known [3, 19] that a Z-tensor $\mathcal{A}$ is a nonsingular M-tensor if and only if $\mathcal{A}x^{m-1} > 0$ for some $x > 0$. It follows that the diagonal elements of a nonsingular M-tensor are all positive.

Suppose $\mathcal{A}$ is a nonsingular M-tensor. When b is positive, equation (1) has a unique positive solution [4]. When b is nonnegative, equation (1) has nonnegative solutions and moreover it has a minimal nonnegative solution [14] and has a maximal nonnegative solution [9].

The following simple example is an extension of Example 1.1 in [1]. It shows that the number of nonnegative solutions could be huge when $b \geq 0$.

Example 1. Let $k \geq 1$ be an integer and let $\mathcal{A} = (a_{i_1 i_2 i_3 i_4}) \in \mathbb{R}^{[4,2k]}$, where $a_{iiii} = 1$ for $i \in [2k]$, $a_{2i-1,2i-1,2i-1,2i} = -2$ for $i \in [k]$, and all other elements are zero. It is clear that $\mathcal{A}$ is a Z-tensor with $\mathcal{A}x^3 > 0$ for $x = [3, 1, \ldots, 3, 1]^T$. So $\mathcal{A}$ is a nonsingular M-tensor. For $b = [0, 1, \ldots, 0, 1]^T$, we find that the equation $\mathcal{A}x^3 = b$ has 2^k solutions $x = [x_1, x_2, \ldots, x_{2k-1}, x_{2k}]^T$, where for $i \in [k]$, $[x_{2i-1}, x_{2i}] = [0, 1]$ or $[2, 1]$. The minimal nonnegative solution is $[0, 1, \ldots, 0, 1]^T$ and the maximal nonnegative solution is $[2, 1, \ldots, 2, 1]^T$.

Since equation (1) may have a large number of nonnegative solutions, it is unlikely that each of these solutions is of practical interest. Intuitively, the extremal nonnegative solutions are of particular interest. By computing the maximal nonnegative solution, we can answer the question whether the equation has a positive solution. By computing the extremal nonnegative solutions, we get bounds for all other nonnegative solutions.

In [14], the tensor complementarity problem (TCP):

$$x \geq 0, \quad \mathcal{A}x^{m-1} - b \geq 0, \quad x^T(\mathcal{A}x^{m-1} - b) = 0$$

is considered. When $\mathcal{A}$ is a Z-tensor and $b \in \mathbb{R}^n_+$, it is shown in [14] that the solution set of the TCP is the same as the set of all nonnegative solutions of equation (1). So a sparsest solution to the TCP is the minimal nonnegative solution of equation (1). The problem of finding a sparsest solution to the TCP is of practical interest [14]. Note that, for this application, the case $b \geq 0$ with some zero elements is much more interesting than the case $b > 0$ (where the equation has a unique positive solution and there are no other nonnegative solutions).

When $\mathcal{A}$ is a nonsingular M-tensor and $b > 0$, the unique positive solution $x(\mathcal{A}, b)$ of $\mathcal{A}x^{m-1} = b$ is the solution of particular interest. But some elements of b may be very tiny. What happens to $x(\mathcal{A}, b)$ if b decreases monotonically towards a vector $b^{(0)}$ with some zero elements? We will see in the next section that $x(\mathcal{A}, b)$ converges to the maximal nonnegative solution of the equation $\mathcal{A}x^{m-1} = b^{(0)}$.

In [9], the authors state in the introduction that their purpose is to find the largest (maximal) nonnegative solution. But their algorithms there only find a nonnegative solution, which is usually not maximal. Numerical methods have also been presented in [1, 10, 11]. Those methods can find a nonnegative solution, but the solution is usually not maximal and cannot be guaranteed to be minimal.

While our main interest is on equation (1) with $\mathcal{A}$ being a nonsingular M-tensor and $b \geq 0$, some of our discussions also allow $\mathcal{A}$ to be a Z-tensor or allow b to have some negative elements. In Section 2, we present new proofs for the existence of extremal nonnegative solutions by using simple iterative methods. These iterative methods can actually compute the extremal solutions. Some other results also follow directly from the existence theorems and their proofs. In Section 3, we show how the simple iterations used in Section 2 can be generalized to have faster convergence. Linear convergence of these iterative methods is also proved under mild assumptions. Numerical experiments are performed in Section 4 to illustrate our theoretical results. Some concluding remarks are given in Section 5.

2. Existence of extremal nonnegative solutions

The following theorem is a main result in [14] (stated differently in [14]; see [14, Theorem 3] and its proof). The proof there is based on several other results and does not show how the minimal nonnegative solution can be computed. Our new proof here is very short and it provides a way to compute the minimal solution. The approach we take here is similar to the approach we used many years ago for determining the existence of the minimal nonnegative solution of M-matrix algebraic Riccati equations [5, 6].

Theorem 1. *Let $\mathcal{A}$ be a Z-tensor and $b \in \mathbb{R}^n_+$. Suppose that $\mathcal{S}_g = \{x \in \mathbb{R}^n_+ \mid \mathcal{A}x^{m-1} \geq b\} \neq \emptyset$. Then $\mathcal{S}_g$ has a minimal element that is also the minimal nonnegative solution of $\mathcal{A}x^{m-1} = b$.*

Proof. We write $\mathcal{A} = \mathcal{D} - \mathcal{B}$, where $\mathcal{D}$ is a diagonal tensor with positive diagonal elements and $\mathcal{B} \geq 0$. Equation (1) becomes $\mathcal{D}x^{m-1} = \mathcal{B}x^{m-1} + b$. Then we have the fixed-point iteration, given implicitly as follows:

$$\mathcal{D}x_{k+1}^{m-1} = \mathcal{B}x_k^{m-1} + b. \tag{2}$$

(This is the Jacobi iteration when the diagonal elements of $\mathcal{A}$ are positive and $\mathcal{D}$ is the diagonal part of $\mathcal{A}$.) Note that $\mathcal{D}x_{k+1}^{m-1} = Dx_{k+1}^{[m-1]}$, where D is the diagonal matrix having the diagonal elements of the tensor $\mathcal{D}$ on the diagonal. When $\mathcal{B}x_k^{m-1} + b \geq 0$, iteration (2) can be given explicitly as

$$x_{k+1} = \left(D^{-1}(\mathcal{B}x_k^{m-1} + b)\right)^{[1/(m-1)]},$$

but the implicit form will be more convenient for our discussions.

Let x be any element in $\mathcal{S}_g$. Then $x \geq 0$ and $\mathcal{B}x^{m-1}+b \leq \mathcal{D}x^{m-1}$. Take $x_0 = 0$. We can generate a nonnegative sequence $\{x_k\}$ by iteration (2). It is clear that $x_0 \leq x_1$. Suppose $x_{k-1} \leq x_k$ ($k \geq 1$). Then $\mathcal{D}x_k^{m-1} = \mathcal{B}x_{k-1}^{m-1}+b \leq \mathcal{B}x_k^{m-1}+b = \mathcal{D}x_{k+1}^{m-1}$. Thus $x_k \leq x_{k+1}$. Therefore, $x_k \leq x_{k+1}$ for all $k \geq 0$. Also, $x_0 \leq x$. Suppose $x_k \leq x$ ($k \geq 0$). Then $\mathcal{D}x_{k+1}^{m-1} = \mathcal{B}x_k^{m-1} + b \leq \mathcal{B}x^{m-1} + b \leq \mathcal{D}x^{m-1}$. Thus $x_{k+1} \leq x$. Therefore, $x_k \leq x$ for all $k \geq 0$. Now, $\{x_k\}$ is monotonically increasing and bounded above by x. Thus, $\lim_{k\to\infty} x_k = x_*$ exists and $x_* \leq x$. Letting $k \to \infty$ in (2), we see that x_* is a nonnegative solution of (1) and $x_* \leq x$ for every x in $\mathcal{S}_g$. In particular, $x_* \leq x$ for every nonnegative solution x of (1), so x_* is the minimal nonnegative solution of (1), and also the minimal element in $\mathcal{S}_g$. □

Corollary 2. *Let $\mathcal{A}$ be a nonsingular M-tensor and $b \geq 0$. Then equation* (1) *has a minimal nonnegative solution.*

Proof. Since $\mathcal{A}$ be a nonsingular M-tensor, we have $\mathcal{A}\hat{x}^{m-1} > 0$ for some $\hat{x} > 0$. Then for scalar $t > 0$ sufficiently large, $\tilde{x} = t\hat{x} > 0$ is such that $\mathcal{A}\tilde{x}^{m-1} = t^{m-1}\mathcal{A}\hat{x}^{m-1} \geq b$, so $\mathcal{S}_g \neq \emptyset$. □

The next result is already known in [10]. It also follows quickly from our proof of Theorem 1.

Corollary 3. *Let $\mathcal{A}$ be a nonsingular M-tensor and $b > 0$. Then the unique positive solution of* (1) *is the only nonnegative solution.*

Proof. From our proof of Theorem 1, we see that the minimal nonnegative solution is positive in this case. □

Corollary 4. *Let $\mathcal{A}$ be a Z-tensor and $b \in \mathbb{R}^n_+$. Suppose that $\mathcal{S}_g = \{x \in \mathbb{R}^n_+ \mid \mathcal{A}x^{m-1} \geq b\} \neq \emptyset$ and let $x_{\min}(\mathcal{A}, b)$ be the minimal nonnegative solution of $\mathcal{A}x^{m-1} = b$. If any element of b decreases but remains nonnegative, or if any diagonal element of $\mathcal{A}$ increases, or if any off-diagonal element of $\mathcal{A}$ increases but remains nonpositive, then the new equation $\tilde{\mathcal{A}}x^{m-1} = \tilde{b}$ also has a minimal nonnegative solution $x_{\min}(\tilde{\mathcal{A}}, \tilde{b})$. Moreover, $x_{\min}(\tilde{\mathcal{A}}, \tilde{b}) \leq x_{\min}(\mathcal{A}, b)$.*

Proof. Under any of those changes, $\tilde{\mathcal{A}}$ is a Z-tensor and $\tilde{b} \in \mathbb{R}^n_+$. Let $\tilde{\mathcal{S}}_g = \{x \in \mathbb{R}^n_+ \mid \tilde{\mathcal{A}}x^{m-1} \geq \tilde{b}\}$. Then $\mathcal{S}_g \subseteq \tilde{\mathcal{S}}_g$ and the conclusions follow immediately. □

The following theorem has been proved in [9]. The proof there is somewhat complicated and does not show how the maximal solution can be computed. Here we present a short proof and also a way to compute the maximal solution.

Theorem 5. *Let $\mathcal{A}$ be a nonsingular M-tensor and suppose that $\mathcal{S}_l = \{x \in \mathbb{R}^n_+ \mid \mathcal{A}x^{m-1} \leq b\} \neq \emptyset$. Then $\mathcal{S}_l$ has a maximal element that is also the maximal nonnegative solution of $\mathcal{A}x^{m-1} = b$.*

PROOF. Let $\mathcal{D}$ be the diagonal part of $\mathcal{A}$ and write $\mathcal{A} = \mathcal{D} - \mathcal{B}$. Then the diagonal elements of $\mathcal{D}$ are positive and $\mathcal{B} \geq 0$. Equation (1) becomes $\mathcal{D}x^{m-1} = \mathcal{B}x^{m-1} + b$. Then we have the fixed-point iteration (Jacobi iteration), given implicitly as follows:

$$\mathcal{D}x_{k+1}^{m-1} = \mathcal{B}x_k^{m-1} + b. \tag{3}$$

Let x be any element in $\mathcal{S}_l$. Then $x \geq 0$ and $\mathcal{B}x^{m-1} + b \geq \mathcal{D}x^{m-1}$. Take $x_0 \geq x$ such that $\mathcal{A}x_0^{m-1} \geq b$, so $\mathcal{B}x_0^{m-1} + b \leq \mathcal{D}x_0^{m-1}$. Note that

$$\mathcal{B}x_0^{m-1} + b \geq \mathcal{B}x^{m-1} + b \geq \mathcal{D}x^{m-1} \geq 0.$$

Thus x_1 is determined by iteration (3) and $\mathcal{D}x_1^{m-1} \geq \mathcal{D}x^{m-1}$, so $x_1 \geq x$. Also, $\mathcal{D}x_1^{m-1} \leq \mathcal{D}x_0^{m-1}$, so $x_1 \leq x_0$. By induction, we can show that $x_{k+1} \leq x_k$ and $x_k \geq x$ for all $k \geq 0$. Therefore, $\lim_{k\to\infty} x_k = x_*$ exists and $x_* \geq x$. Then x_* is a nonnegative solution of (1).

We still need to show that we can choose one fixed x_0 such that $\mathcal{A}x_0^{m-1} \geq b$ and $x_0 \geq x$ for all $x \in \mathcal{S}_l$. To this end, we take any $\hat{b} > 0$ such that $\hat{b} \geq b$, and apply the above argument to the equation $\mathcal{A}x^{m-1} = \hat{b}$. We can conclude that $\mathcal{A}x^{m-1} = \hat{b}$ has a nonnegative solution $\hat{x}_*$ (actually the unique positive solution) with $\hat{x}_* \geq x$ for every element x in $\{x \in \mathbb{R}_+^n \mid \mathcal{A}x^{m-1} \leq \hat{b}\}$ and thus for every element x in $\mathcal{S}_l$.

We now return to the equation $\mathcal{A}x^{m-1} = b$ and take $x_0 \geq \hat{x}_*$ such that $\mathcal{A}x_0^{m-1} \geq b$. Then the sequence $\{x_k\}$ from the Jacobi iteration converges to a nonnegative solution x_* of (1) and $x_* \geq x$ for all elements x in $\mathcal{S}_l$. This x_* is then the maximal element in $\mathcal{S}_l$ and also the maximal nonnegative solution of (1). □

Remark 1. From the proof, we see that the maximal nonnegative solution (when exists) can be found by iteration (3) using any x_0 such that

$$x_0 > 0, \quad \mathcal{A}x_0^{m-1} > 0, \quad \mathcal{A}x_0^{m-1} \geq b.$$

In other words, we can take x_0 to be the unique positive solution of $\mathcal{A}x_0^{m-1} = \hat{b}$, where $\hat{b}$ is any vector such that $\hat{b} > 0$ and $\hat{b} \geq b$.

Remark 2. If $\mathcal{A}$ is a Z-tensor, but not a nonsingular M-tensor, then equation (1) may not have a maximal nonnegative solution when it has nonnegative solutions. One example is the equation (1) with $\mathcal{A} \in \mathbb{R}^{[4,3]}$ given by:

$$a_{1111} = 0, a_{2222} = a_{3333} = 1, a_{1112} = a_{3111} = -1, \text{ and } a_{i_1i_2i_3i_4} = 0 \text{ elsewhere},$$

and $b = [0, 0, 1]^T$. The equation has infinitely many nonnegative solutions, given by $[c, 0, (1 + c^3)^{1/3}]^T$ for any $c \geq 0$. The minimal nonnegative solution is $[0, 0, 1]^T$, but the maximal nonnegative solution does not exist.

Corollary 6. *Let $\mathcal{A}$ be a nonsingular M-tensor and $b \geq 0$. Then equation* (1) *has a maximal nonnegative solution.*

Remark 3. When $\mathcal{A}$ is a nonsingular M-tensor and $b \geq 0$, we can use Jacobi iteration with $x_0 = 0$ to get the minimal solution, and use $x_0 > 0$ with $\mathcal{A}x_0^{m-1} > 0$ and $\mathcal{A}x_0^{m-1} \geq b$ to get the maximal solution. When $b > 0$, we can use either of these two choices to get the unique positive solution. For the case $b > 0$, the Jacobi iteration has been studied in [4] with $x_0 > 0$ satisfying $0 < \mathcal{A}x_0^{m-1} \leq b$. More general tensor splitting methods have been studied in [13], again with this requirement on x_0 (see [13, Theorem 5.4]). It is interesting to note that both our choices are excluded by this requirement (unless x_0 is already the solution).

Corollary 7. *Let $\mathcal{A}$ be a nonsingular M-tensor. Suppose that $\mathcal{S}_l = \{x \in \mathbb{R}_+^n \mid \mathcal{A}x^{m-1} \leq b\} \neq \emptyset$ and let $x_{\max}(\mathcal{A}, b)$ be the maximal nonnegative solution of $\mathcal{A}x^{m-1} = b$. If any element of b increases, or if any entry of $\mathcal{A}$ decreases, then the new equation $\tilde{\mathcal{A}}x^{m-1} = \tilde{b}$ also has a maximal nonnegative solution $x_{\max}(\tilde{\mathcal{A}}, \tilde{b})$ provided that $\tilde{\mathcal{A}}$ is still a nonsingular M-tensor. Moreover, $x_{\max}(\tilde{\mathcal{A}}, \tilde{b}) \geq x_{\max}(\mathcal{A}, b)$.*

Proof. Let $\tilde{\mathcal{S}}_l = \{x \in \mathbb{R}_+^n \mid \tilde{\mathcal{A}}x^{m-1} \leq \tilde{b}\}$. Then $\mathcal{S}_g \subseteq \tilde{\mathcal{S}}_g$ and the conclusions follow immediately. □

The next result partially explain why the maximal nonnegative solution is of particular interest for equation (1) with $b \geq 0$.

Corollary 8. *Let $\mathcal{A}$ be a nonsingular M-tensor and $b \geq 0$. Suppose $x^{(k)}$ is the unique positive solution of $\mathcal{A}x^{m-1} = b^{(k)}$, where $b^{(k)} > 0$ ($k = 1, 2, \ldots$) and $b^{(k)}$ is monotonically decreasing and converges to b as $k \to \infty$. Then, as $k \to \infty$, $x^{(k)}$ converges to the maximal nonnegative solution $x_{\max}(\mathcal{A}, b)$ of $\mathcal{A}x^{m-1} = b$.*

Proof. By Corollary 7, we have $x^{(k)} \geq x^{(k+1)} \geq x_{\max}(\mathcal{A}, b)$ for all $k \geq 0$. Thus $\lim_{k\to\infty} x^{(k)} = x_*$ exists. Letting $k \to \infty$ in $\mathcal{A}(x^{(k)})^{m-1} = b^{(k)}$, we get $\mathcal{A}x_*^{m-1} = b$ and $x_* \geq x_{\max}(\mathcal{A}, b)$. Thus $x_* = x_{\max}(\mathcal{A}, b)$. □

3. Iterative methods for extremal nonnegative solutions

In the previous section, we have presented new proofs for the existence of extremal nonnegative solutions of the tensor equation $\mathcal{A}x^{m-1} = b$, where $\mathcal{A}$ is a nonsingular M-tensor and b is a nonnegative vector, by using the Jacobi iteration with suitable initial guesses. We have also presented some results when $\mathcal{A}$ is a Z-tensor or b is a general real vector. Now, we would like to present some iterative methods that may be more efficient than the Jacobi iteration for actual computation of the extremal solutions, determine and compare the rates of convergence of these methods.

The Jacobi iteration is just a very simple fixed-point iteration. In the Jacobi iteration, we keep the term involving x_i^{m-1} in the ith equation ($i \in [n]$) on the left and move all other terms to the right. But in the ith equation, we may also have terms x_j^{m-1} with $j \neq i$. We may consider keeping all unmixed terms $(x_1^{m-1}, \ldots, x_n^{m-1})$ on the left, regardless which equation they are from. In other words, we have a splitting of the tensor $\mathcal{A}$: $\mathcal{A} = \mathcal{M} - \mathcal{N}$, where $\mathcal{M} = (m_{i_1 i_2 \ldots i_m})$

with $m_{ij\dots j} = a_{ij\dots j}$ for $i, j \in [n]$ and $m_{i_1 i_2 \dots i_m} = 0$ elsewhere. We may call this splitting a level-1 splitting. If we keep all unmixed terms on the left, we then have the equation

$$\mathcal{M}x^{m-1} = \mathcal{N}x^{m-1} + b.$$

Note that $\mathcal{M}x^{m-1} = Mx^{[m-1]}$, where the $n \times n$ matrix M has (i, j) element $a_{ij\dots j}$ for $i, j \in [n]$, and is called [17] the majorization matrix associated with $\mathcal{A}$. It is easy to see [13] that M is a nonsingular M-matrix when $\mathcal{A}$ is a nonsingular M-tensor. When $\mathcal{A}$ is a Z-tensor, M is obviously a Z-matrix.

We then have a splitting of the matrix M: $M = P - Q$, where P is a nonsingular M-matrix and $Q \geq 0$. This may be called a level-2 splitting. For example, the following splitting of a Z-matrix is permitted, although not a good one.

$$\begin{bmatrix} -1 & 0 & 0 \\ -3 & 2 & -2 \\ 0 & -3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 3 & -1 \\ 0 & -2 & 5 \end{bmatrix} - \begin{bmatrix} 3 & 0 & 0 \\ 2 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

We can now rewrite $\mathcal{A}x^{m-1} = b$ as

$$Px^{[m-1]} = Qx^{[m-1]} + \mathcal{N}x^{m-1} + b$$

and get fixed-point iteration in implicit form

$$Px_{k+1}^{[m-1]} = Qx_k^{[m-1]} + \mathcal{N}x_k^{m-1} + b \tag{4}$$

or in explicit form

$$x_{k+1} = \left(P^{-1}\left(Qx_k^{[m-1]} + \mathcal{N}x_k^{m-1} + b\right)\right)^{[1/(m-1)]},$$

assuming $P^{-1}\left(Qx_k^{[m-1]} + \mathcal{N}x_k^{m-1} + b\right) \geq 0$ for each k. In [13], fixed-point iterations like this are called tensor splitting iterative methods and studied for finding the unique positive solution of the tensor equation $\mathcal{A}x^{m-1} = b$ with $\mathcal{A}$ being a nonsingular M-tensor and b a positive vector.

We first study iteration (4) for finding the minimal nonnegative solution.

Theorem 9. *Let $\mathcal{A}$ be a Z-tensor and $b \in \mathbb{R}^n_+$. Suppose that $\mathcal{S}_g = \{x \in \mathbb{R}^n_+ \mid \mathcal{A}x^{m-1} \geq b\} \neq \emptyset$. Then for $x_0 = 0$, the sequence $\{x_k\}$ from iteration (4) is monotonically increasing and converges to the minimal nonnegative solution of $\mathcal{A}x^{m-1} = b$.*

Proof. The proof is almost the same as the proof of Theorem 1. In that proof, we use the obvious fact that $\mathcal{D}x^{m-1} \geq \mathcal{D}y^{m-1}$ (with $x, y \geq 0$) implies $x \geq y$. We now need $Px^{[m-1]} \geq Py^{[m-1]}$ implies $x \geq y$, which is true since $P^{-1} \geq 0$ for the nonsingular M-matrix P. □

Remark 4. When $\mathcal{A}$ is a Z-tensor but not a nonsingular M-tensor, it may not be easy to determine whether $\mathcal{S}_g \neq \emptyset$. We may apply iteration (4) without checking this condition. In this case, the sequence $\{x_k\}$ is still well-defined and monotonically increasing. The sequence is bounded above if and only if $\mathcal{S}_g \neq \emptyset$.

In iteration (4), $\mathcal{N}$ is uniquely determined by $\mathcal{A}$, but we have the freedom to choose Q (P is uniquely determined by Q). The next result gives a comparison of convergence rate by examining the iterates right from the beginning (so we are not talking about asymptotic rate of convergence here).

Theorem 10. *Under the conditions of Theorem 9, let $\{x_k\}$ and $\{\hat{x}_k\}$ be the sequences generated by iteration (4) with two splittings of M: $M = P - Q$ and $M = \hat{P} - \hat{Q}$, respectively, and with $x_0 = \hat{x}_0 = 0$. If $\hat{Q} \le Q$, then $x_k \le \hat{x}_k$ for all $k \ge 0$. In other words, a smaller matrix Q gives faster termwise convergence for finding the minimal nonnegative solution.*

Proof. We need to show that $x_k \le \hat{x}_k$ implies $x_{k+1} \le \hat{x}_{k+1}$ for each $k \ge 0$. We have (4) and

$$\hat{P}\hat{x}_{k+1}^{[m-1]} = \hat{Q}\hat{x}_k^{[m-1]} + \mathcal{N}\hat{x}_k^{m-1} + b. \tag{5}$$

From $P - Q = \hat{P} - \hat{Q}$, we get $\hat{P} = P - (Q - \hat{Q})$ and get from (5) that

$$\begin{aligned} P\hat{x}_{k+1}^{[m-1]} &= (Q - \hat{Q})\hat{x}_{k+1}^{[m-1]} + \hat{Q}\hat{x}_k^{[m-1]} + \mathcal{N}\hat{x}_k^{m-1} + b \\ &\ge (Q - \hat{Q})\hat{x}_k^{[m-1]} + \hat{Q}\hat{x}_k^{[m-1]} + \mathcal{N}\hat{x}_k^{m-1} + b \\ &= Q\hat{x}_k^{[m-1]} + \mathcal{N}\hat{x}_k^{m-1} + b \\ &\ge Qx_k^{[m-1]} + \mathcal{N}x_k^{m-1} + b \\ &= Px_{k+1}^{[m-1]}. \end{aligned}$$

Thus $x_{k+1} \le \hat{x}_{k+1}$. □

We now show that iteration (4) has linear convergence under suitable assumptions. By linear convergence we actually mean at least linear convergence. For example, $x_0 = 0$ is already a solution when $b = 0$. Let $\mathcal{L}$ be the nonnegative tensor such that $\mathcal{L}x^{m-1} = Qx^{[m-1]} + \mathcal{N}x^{m-1}$. Then iteration (4) becomes

$$Px_{k+1}^{[m-1]} = \mathcal{L}x_k^{m-1} + b. \tag{6}$$

For $x \in \mathbb{R}^n$ and index set $I \subseteq [n]$, we denote by x_I the subvector of x, whose elements are x_i, $i \in I$. For tensor $\mathcal{A} = (a_{i_1 \ldots i_m}) \in \mathbb{R}^{[m,n]}$, we denote by $\mathcal{A}_I$ the subtensor of $\mathcal{A}$ with elements $a_{i_1 \ldots i_m}$, $i_1, \ldots, i_m \in I$.

Theorem 11. *Let $\mathcal{A}$ be a Z-tensor and $b \in \mathbb{R}_+^n$ be nonzero. Let $I_0 = \{i \mid b_i = 0\}$ and suppose that $\mathcal{S}_g = \{x \in \mathbb{R}_+^n \mid \mathcal{A}x^{m-1} \ge b\} \ne \emptyset$. Let $\{x_k\}$ be the sequence from iteration (6) with $x_0 = 0$. If k_0 is the smallest integer such that x_{k_0} and x_{k_0+1} have the same zero pattern, then $1 \le k_0 \le n$ and x_k have the same zero pattern for all $k \ge k_0$, which is also the zero pattern of the minimal nonnegative solution $x_{\min}$. Let $I = \{i \mid (x_{k_0})_i = 0\}$. Then $I \subseteq I_0$. Let $I_c = [n] \setminus I$. Then the iteration (6) for $k \ge k_0$ is reduced to an iteration for the lower-dimensional tensor equation:*

$$\mathcal{A}_{I_c}\hat{x}^{m-1} = b_{I_c}. \tag{7}$$

For $k \geq k_0$, $\hat{x}_k$ from the reduced iteration is the same as $(x_k)_{I_c}$. The minimal nonnegative solution $\hat{x}_{\min}$ of (7) is positive and is the same as $(x_{\min})_{I_c}$. Thus x_k converges to $x_{\min}$ linearly if and only if $\hat{x}_k$ converges to $\hat{x}_{\min}$ linearly. Assume $\mathcal{A}x^{m-1} = b$ is already the reduced equation for notation convenience (all diagonal elements of $\mathcal{A}$ are now positive). Then x_k converges to $x_{\min}$ linearly under any of the following four conditions:

(a) $P^{-1}b > 0$.

(b) $P^{-1}\bar{\mathcal{L}}e^{m-2}$ *is irreducible, where e is the vector of ones and $\bar{\mathcal{L}}$ is the semi-symmetric tensor from $\mathcal{L}$.*

(c) *With all diagonal elements of $\mathcal{A}$ included entirely in P and $\mathcal{B} = \mathcal{D} - \mathcal{A}$ ($\mathcal{D}$ is the diagonal part of $\mathcal{A}$), the matrix $\bar{\mathcal{B}}e^{m-2}$ is strictly upper or lower triangular.*

(d) *For each $i \in I_0$, there is an element $a_{ii_2\ldots i_m} \neq 0$, where $i_j \notin I_0$ for at least one j ($2 \leq j \leq m$).*

Proof. With $x_0 = 0$ for iteration (6), we have $x_k \leq x_{k+1}$ for each $k \geq 0$. Let k_0 be the smallest integer such that x_{k_0} and x_{k_0+1} have the same zero pattern. It is clear that $1 \leq k_0 \leq n$ and that the zero pattern will not change afterwards and the minimal solution $x_{\min}$ has the same zero pattern. Let $I = \{i \mid (x_{k_0})_i = 0\}$. Since $b_i > 0$ implies $(x_{k_0})_i \geq (x_1)_i > 0$, we have $I \subseteq I_0$. For $k \geq k_0$, we have $(x_k)_I = 0$. Let $I_c = [n] \setminus I$. Since $(x_k)_{i_2}(x_k)_{i_3}\cdots(x_k)_{i_m} = 0$ for all $k \geq k_0$ when $i_j \in I$ for at least one j ($j = 2, \ldots, m$), we only need the elements in $\mathcal{A}_{I_c}$ to continue the iteration (6) for $k \geq k_0$. In other words, the iteration (6) for $k \geq k_0$ is reduced to an iteration for the lower-dimensional tensor equation with indices running through I_c only:

$$\mathcal{A}_{I_c}\hat{x}^{m-1} = b_{I_c}, \tag{8}$$

with $\hat{x}_{k_0}$ obtained from x_{k_0} by deleting all zero elements. Note that the level-1 and level-2 splittings of $\mathcal{A}_{I_c}$ are the ones inherited from those for $\mathcal{A}$. The minimal nonnegative solution $\hat{x}_{\min}$ of (8) is positive (obtained from $x_{\min}$ by deleting all zero elements), but b_{I_c} will have some zero elements when $I \neq I_0$. For $k \geq k_0$, $\hat{x}_k$ from the reduced iteration can also be obtained from x_k by deleting all zero elements. Thus x_k converges to $x_{\min}$ linearly if and only if $\hat{x}_k$ converges to $\hat{x}_{\min}$ linearly. We now assume that $\mathcal{A}x^{m-1} = b$ is already the reduced equation for notation convenience, and denote its positive minimal solution by $\bar{x}$ for short. Note that all diagonal elements of the reduced tensor are positive.

By Theorem 10, we may assume P is a diagonal matrix (where slower convergence happens). So we will assume P is diagonal when needed.

The iteration (6) can be written explicitly as $x_{k+1} = \phi(x_k)$, with

$$\phi(x) = \left(P^{-1}(\mathcal{L}x^{m-1} + b)\right)^{[1/(m-1)]}.$$

From

$$P\phi(x)^{[m-1]} = \mathcal{L}x^{m-1} + b = \bar{\mathcal{L}}x^{m-1} + b,$$

where $\bar{\mathcal{L}}$ is the semi-symmetric tensor obtained from $\mathcal{L}$, we take derivative on both sides to obtain

$$P(m-1)\mathrm{diag}(\phi(x)^{[m-2]})\phi'(x) = (m-1)\bar{\mathcal{L}}x^{m-2},$$

where for $x \in \mathbb{R}^n$, $\mathrm{diag}(x)$ is the diagonal matrix with the elements of x on the diagonal. We need to show $\rho(\phi'(\bar{x})) < 1$ for

$$\phi'(\bar{x}) = \mathrm{diag}(\bar{x}^{[-(m-2)]})P^{-1}\bar{\mathcal{L}}\bar{x}^{m-2}. \tag{9}$$

Note that the nonnegative matrix $\phi'(\bar{x})$ is such that

$$\begin{aligned}
\phi'(\bar{x})\bar{x} &= \mathrm{diag}(\bar{x}^{[-(m-2)]})P^{-1}\bar{\mathcal{L}}\bar{x}^{m-2}\bar{x} \\
&= \mathrm{diag}(\bar{x}^{[-(m-2)]})P^{-1}\bar{\mathcal{L}}\bar{x}^{m-1} \\
&= \mathrm{diag}(\bar{x}^{[-(m-2)]})P^{-1}\mathcal{L}\bar{x}^{m-1} \\
&= \mathrm{diag}(\bar{x}^{[-(m-2)]})P^{-1}(P\bar{x}^{[m-1]} - \mathcal{A}\bar{x}^{m-1}) \\
&= \bar{x} - \mathrm{diag}(\bar{x}^{[-(m-2)]})P^{-1}b.
\end{aligned}$$

If $P^{-1}b > 0$, then we have $\phi'(\bar{x})\bar{x} < \bar{x}$ and thus [2] $\rho(\phi'(\bar{x})) < 1$. This shows that condition (a) is sufficient for linear convergence. Conditions (b), (c) and (d) are needed only when $I_0 \neq \emptyset$ for the reduced equation.

Since $P^{-1}b \geq 0$ and $P^{-1}b \neq 0$, we have $\phi'(\bar{x})\bar{x} \leq \bar{x}$ and $\phi'(\bar{x})\bar{x} \neq \bar{x}$. Thus we still have [2] $\rho(\phi'(\bar{x})) < 1$ if $\phi'(\bar{x})$ is irreducible. From (9), we see that $\phi'(\bar{x})$ is irreducible if and only if $P^{-1}\bar{\mathcal{L}}e^{m-2}$ is irreducible. Thus, condition (b) is also sufficient for linear convergence.

Under condition (c), we just need to show linear convergence for the Jacobi iteration. Now,

$$\phi'(\bar{x}) = \mathrm{diag}(\bar{x}^{[-(m-2)]})D^{-1}\bar{\mathcal{B}}\bar{x}^{m-2}.$$

Since $\bar{\mathcal{B}}e^{m-2}$ is strictly upper or lower triangular, so is $\phi'(\bar{x})$. Thus $\rho(\phi'(\bar{x})) = 0$ and linear convergence follows.

Finally, we show that condition (d) is also sufficient for linear convergence. We now assume P is diagonal. We modify $\bar{x}$ to $\hat{x}$ by changing $(\bar{x})_i$ to $(\bar{x})_i - \epsilon > 0$ for each $i \notin I_0$, where $0 < \epsilon < \min_{i \notin I_0}(\bar{x})_i^{-(m-2)}(P^{-1}b)_i$. Now for $i \notin I_0$,

$$(\phi'(\bar{x})\hat{x})_i \leq (\phi'(\bar{x})\bar{x})_i = (\bar{x})_i - (\bar{x})_i^{-(m-2)}(P^{-1}b)_i < (\hat{x})_i.$$

For $i \in I_0$,

$$(\phi'(\bar{x})\hat{x})_i \leq (\phi'(\bar{x})\bar{x})_i = (\bar{x})_i = (\hat{x})_i.$$

Since P is diagonal, $(\phi'(\bar{x})\hat{x})_i = (\phi'(\bar{x})\bar{x})_i$ if and only if $(\bar{\mathcal{L}}\bar{x}^{m-2}\hat{x})_i = (\bar{\mathcal{L}}\bar{x}^{m-2}\bar{x})_i$, i.e.,

$$\begin{aligned}
&\sum_{j=1}^{n}\left(\sum_{i_3,\ldots,i_m=1}^{n}\bar{\mathcal{L}}_{iji_3\ldots i_m}\bar{x}_{i_3}\cdots\bar{x}_{i_m}\right)\hat{x}_j \\
= &\sum_{j=1}^{n}\left(\sum_{i_3,\ldots,i_m=1}^{n}\bar{\mathcal{L}}_{iji_3\ldots i_m}\bar{x}_{i_3}\cdots\bar{x}_{i_m}\right)\bar{x}_j.
\end{aligned}$$

This holds if and only if $\bar{\mathcal{L}}_{iji_3\ldots i_m} = 0$ for all $j \notin I_0$ and for all $i_3, \ldots, i_m \in [n]$, i.e., $\mathcal{L}_{ii_2i_3\ldots i_m} = 0$ for all $i_2, \ldots, i_m \in [n]$ such that $i_j \notin I_0$ for at least one j $(2 \le j \le m)$. Therefore, when condition (d) holds, $(\phi'(\bar{x})\hat{x})_i < (\phi'(\bar{x})\bar{x})_i = (\hat{x})_i$ for each $i \in I_0$. Linear convergence follows since we again have $\rho(\phi'(\bar{x})) < 1$. $\square$

We now study iteration (4) for finding the maximal nonnegative solution.

Theorem 12. *Let $\mathcal{A}$ be a nonsingular M-tensor and suppose that $\mathcal{S}_l = \{x \in \mathbb{R}^n_+ \mid \mathcal{A}x^{m-1} \le b\} \ne \emptyset$. Then for any $x_0 > 0$ such that $\mathcal{A}x_0^{m-1} > 0$ and $\mathcal{A}x_0^{m-1} \ge b$. The sequence from iteration (4) is monotonically decreasing and converges to the maximal nonnegative solution of $\mathcal{A}x^{m-1} = b$.*

Proof. The proof is almost the same as the proof of Theorem 5. In that proof, we use the fact that $\mathcal{D}x^{m-1} \ge \mathcal{D}y^{m-1}$ (with $x, y \ge 0$) implies $x \ge y$. We now use the fact $Px^{[m-1]} \ge Py^{[m-1]}$ implies $x \ge y$. $\square$

Remark 5. When $b \notin \mathbb{R}^n_+$, it may not be easy to determine whether $\mathcal{S}_l \ne \emptyset$. We may apply iteration (4) without checking this condition. In this case, $\mathcal{S}_l \ne \emptyset$ if and only if $P^{-1}\left(Qx_k^{[m-1]} + \mathcal{N}x_k^{m-1} + b\right) \ge 0$ for each $k \ge 0$. Indeed, if $P^{-1}\left(Qx_k^{[m-1]} + \mathcal{N}x_k^{m-1} + b\right) \ge 0$ for each $k \ge 0$, then we see from a proof similar to that of Theorem 5 that $x_k \ge x_{k+1} \ge 0$ for all $k \ge 0$. In this case, $\lim_{k\to\infty} x_k = x_*$ exists and x_* is a solution of $\mathcal{A}x^{m-1} = b$ and thus $\mathcal{S}_l \ne \emptyset$.

Remark 6. To obtain a suitable x_0 in Theorem 12, we can always take a vector $\hat{b} > 0$ with $\hat{b} \ge b$, and use the methods in [7, 8] to get the unique positive solution x_* of the equation $\mathcal{A}x^{m-1} = \hat{b}$ and then take $x_0 = x_*$. The Jacobi iteration that we have discussed here will also work very well for this purpose. Specifically, we can take $\hat{b}$ with $\hat{b}_i = b_i + 1$ if $b_i \ge 0$ and $\hat{b}_i = 1$ if $b_i < 0$ (we assume that the vector b has been scaled so that its largest element is around 1), and apply Jacobi iteration with $x_0 = 0$ to the equation $\mathcal{A}x^{m-1} = \hat{b}$. We can determine the smallest $k \ge 1$ such that have $\mathcal{A}x_k^{m-1} > 0$ and $\mathcal{A}x_k^{m-1} \ge b$ (we have $x_k > 0$ for all $k \ge 1$). Then this x_k can be used as x_0 in Theorem 12. Note that the exact solution of $\mathcal{A}x^{m-1} = \hat{b}$ will be a worse x_0 for Theorem 12.

Theorem 13. *Under the conditions of Theorem 12, let $\{x_k\}$ and $\{\hat{x}_k\}$ be the sequences generated by iteration (4) with two splittings of M: $M = P - Q$ and $M = \hat{P} - \hat{Q}$, respectively, and with $x_0 \ge \hat{x}_0 > 0$ satisfying*

$$\mathcal{A}x_0^{m-1} > 0, \quad \mathcal{A}x_0^{m-1} \ge b, \quad \mathcal{A}\hat{x}_0^{m-1} > 0, \quad \mathcal{A}\hat{x}_0^{m-1} \ge b.$$

If $\hat{Q} \le Q$, then $x_k \ge \hat{x}_k$ for all $k \ge 0$. In other words, a smaller matrix Q gives faster termwise convergence for finding the maximal solution.

Proof. We need to show that $x_k \ge \hat{x}_k$ implies $x_{k+1} \ge \hat{x}_{k+1}$ for each $k \ge 0$.

We have (4) and (5). Since $\hat{P} = P - (Q - \hat{Q})$, we get from (5) that

$$\begin{aligned} P\hat{x}_{k+1}^{[m-1]} &= (Q-\hat{Q})\hat{x}_{k+1}^{[m-1]} + \hat{Q}\hat{x}_k^{[m-1]} + \mathcal{N}\hat{x}_k^{m-1} + b \\ &\leq (Q-\hat{Q})\hat{x}_k^{[m-1]} + \hat{Q}\hat{x}_k^{[m-1]} + \mathcal{N}\hat{x}_k^{m-1} + b \\ &= Q\hat{x}_k^{[m-1]} + \mathcal{N}\hat{x}_k^{m-1} + b \\ &\leq Qx_k^{[m-1]} + \mathcal{N}x_k^{m-1} + b \\ &= Px_{k+1}^{[m-1]}. \end{aligned}$$

Thus $x_{k+1} \geq \hat{x}_{k+1}$. □

We now study the convergence rate of iteration (4) for finding the maximal nonnegative solution of (1), under the conditions of Theorem 12. In our first result we assume that the maximal nonnegative solution is positive. We know from [10, Theorem 2.4] that every nonnegative solution of (1) is positive if $\mathcal{A}$ is irreducible and $b \geq 0$ is nonzero. We also know from Corollary 7 that if the maximal nonnegative solution of $\mathcal{A}x^{m-1} = b$ is positive, then the maximal nonnegative solution of $\tilde{\mathcal{A}}x^{m-1} = \tilde{b}$ will also be positive if $\tilde{\mathcal{A}}$ and $\tilde{b}$ are obtained from $\mathcal{A}$ and b in ways described there.

Theorem 14. *Let $\mathcal{A}$ be a nonsingular M-tensor and suppose that $\mathcal{S}_l = \{x \in \mathbb{R}_+^n \mid \mathcal{A}x^{m-1} \leq b\} \neq \emptyset$. Let $x_{\max}$ be the maximal nonnegative solution of $\mathcal{A}x^{m-1} = b$ and assume that $x_{\max} > 0$. Let $I_0 = \{i \mid b_i = 0\}$. Then for any $x_0 > 0$ such that $\mathcal{A}x_0^{m-1} > 0$ and $\mathcal{A}x_0^{m-1} \geq b$, the sequence $\{x_k\}$ from iteration* (4) *converges to x_{max} linearly under any of the following four conditions:*

(a) $P^{-1}b > 0$.
(b) *$P^{-1}\bar{\mathcal{L}}e^{m-2}$ is irreducible, where e is the vector of ones and $\bar{\mathcal{L}}$ is the semi-symmetric tensor from $\mathcal{L}$.*
(c) *With all diagonal elements of $\mathcal{A}$ included entirely in P and $\mathcal{B} = \mathcal{D} - \mathcal{A}$, the matrix $\bar{\mathcal{B}}e^{m-2}$ is strictly upper or lower triangular.*
(d) *For each $i \in I_0$, there is an element $a_{ii_2\ldots i_m} \neq 0$, where $i_j \notin I_0$ for at least one j $(2 \leq j \leq m)$.*

Proof. When $b = 0$, $\mathcal{A}x^{m-1} = b$ has a unique solution $x = 0$. Thus $b \neq 0$ when $x_{\max} > 0$. Exactly as in the proof of Theorem 11 (we do not need the reduction process there since we assume $x_{\max} > 0$), we can show that the iteration map ϕ is such that $\rho(\phi'(x_{\max})) < 1$ under any of the four conditions in the theorem. In proving linear convergence under condition (c) or (d), we may assume that P is diagonal. □

We now assume $b \geq 0$, but allow $x_{\max}$ to have some zero elements.

Theorem 15. *Let $\mathcal{A}$ be a nonsingular M-tensor and $b \in \mathbb{R}_+^n$ be nonzero. Let $x_{\max}$ be the maximal nonnegative solution of $\mathcal{A}x^{m-1} = b$. Let $I_0 = \{i \mid b_i = 0\}$. Then for iteration* (4) *with any $x_0 > 0$ such that $\mathcal{A}x_0^{m-1} > 0$ and $\mathcal{A}x_0^{m-1} \geq b$,* *there is a smallest integer* k_0 *(*$0 \le k_0 \le n-1$*) such that* x_{k_0} *and* x_{k_0+1} *have the same zero pattern (including the case with no zero elements), and* x_k *have the same zero pattern for all* $k \ge k_0$*. Let* $I = \{i \mid (x_{k_0})_i = 0\}$ *(which may be empty). Then* $I \subseteq I_0$*. Let* $I_c = [n] \setminus I$*. Then the iteration* (4) *for* $k \ge k_0$ *is reduced to an iteration for the lower-dimensional tensor equation:*

$$\mathcal{A}_{I_c} \hat{x}^{m-1} = b_{I_c}. \tag{10}$$

For $k \ge k_0$*,* $\hat{x}_k$ *from the reduced iteration is the same as* $(x_k)_{I_c}$*. The maximal nonnegative solution* $\hat{x}_{\max}$ *of* (10) *is the same as* $(x_{\max})_{I_c}$ *(which is not necessarily positive). Thus* x_k *converges to* $x_{\max}$ *linearly if and only if* $\hat{x}_k$ *converges to* $\hat{x}_{\max}$ *linearly. Assume* $\mathcal{A}x^{m-1} = b$ *is already the reduced equation and assume that its maximal solution* $x_{\max}$ *is positive. Then* x_k *converges to* $x_{\max}$ *linearly under any of the following four conditions:*

(a) $P^{-1}b > 0$.

(b) $P^{-1}\bar{\mathcal{L}}e^{m-2}$ *is irreducible, where* e *is the vector of ones and* $\bar{\mathcal{L}}$ *is the semi-symmetric tensor from* $\mathcal{L}$*.*

(c) *With all diagonal elements of* $\mathcal{A}$ *included entirely in* P *and* $\mathcal{B} = \mathcal{D} - \mathcal{A}$*, the matrix* $\bar{\mathcal{B}}e^{m-2}$ *is strictly upper or lower triangular.*

(d) *For each* $i \in I_0$*, there is an element* $a_{ii_2\ldots i_m} \ne 0$*, where* $i_j \notin I_0$ *for at least one* j *(*$2 \le j \le m$*).*

Proof. We have $x_0 > 0$ and $x_k \ge x_{k+1}$ for all $k \ge 0$. Each x_k has at least one nonzero element since $x_{\max} \ne 0$. Thus, there is a smallest integer k_0 ($0 \le k_0 \le n-1$) such that x_{k_0} and x_{k_0+1} have the same zero pattern. Since $b \ge 0$, x_k have the same zero pattern for all $k \ge k_0$. It is clear that $I \subseteq I_0$. The reduced equation (10) is then obtained as in the proof of Theorem 11. However, its maximal solution $x_{\max}$ (as the limit of a positive sequence) may have some zero elements. With the additional assumption that $x_{\max} > 0$ for the reduced equation, the proof of linear convergence is exactly the same as in the proof of Theorem 11. □

Remark 7. The reduction in Theorem 11 is obtained from the iteration for finding the minimal solution and the reduction in Theorem 15 is obtained from the iteration for finding the maximal solution. A reduction process has also been described in [10], without mentioning maximal and minimal solutions. We will explain that one can only find a nonnegative solution with the same number of nonzero elements as the minimal solution by finding a positive solution of the reduced equation in [10].

To describe the approach in [10], we recall the following definition.

Definition 2. A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called reducible with respect to $I \subset [n]$ if its elements satisfy

$$a_{i_1 i_2 \ldots i_m} = 0, \forall i_1 \in I, \forall i_2, \ldots, i_m \notin I.$$

The next result has been presented in [10] (see Corollary 2.8 there).

Theorem 16. *Suppose that $\mathcal{A}$ is a nonsingular M-tensor and $b \in \mathbb{R}^n_+$ is nonzero. Then there is an index set $I \subseteq I_0$ (which could be empty) such that every nonnegative solution to the following lower dimensional tensor equation with $I_c = [n]\backslash I$*

$$\mathcal{A}_{I_c} x_{I_c}^{m-1} = b_{I_c}$$

is positive. Moreover, every positive solution x_{I_c} of the last equation together with $x_I = 0$ forms a nonnegative solution to equation (1).

By comparing the discussions in [10] and in this paper, we can see that the index set $I \subseteq I_0$ in Theorem 16 is the largest set such that $\mathcal{A}$ is reducible with respect to I. We also see that the index set I can be determined automatically by iteration (4) for computing the minimal nonnegative solution, with $x_0 = 0$. Indeed, $I = \{i \mid (x_k)_i = 0\}$, where k is the smallest integer such that the vectors x_k and x_{k+1} from iteration (4) have the same zero pattern (see Theorem 11; $I = \emptyset$ when $b > 0$). This is a very easy way to determine the set I. It would be much more expensive to determine I by using the definition of reducibility to find the largest index set $I \subseteq I_0$ such that $\mathcal{A}$ is reducible with respect to I.

The numerical methods in [10, 11] for computing a nonnegative solution of (1) are based on Theorem 16. They have quadratic convergence under suitable assumptions. For example, assuming equation (1) has already been reduced, the assumption needed for a Newton method in [11] is that for each $i \in I_0$, there is an element $a_{ii_2\ldots i_m} \neq 0$ with all $i_j \notin I_0$ $(j = 2, \ldots, m)$. So the assumption is stronger than our condition (d) in Theorem 11 for the linear convergence of our simple iteration for finding the minimal solution. From our comments on Theorem 16, we know that the methods in [10, 11] can only find one of the nonnegative solutions that has the same number of zero elements as the minimal solution. In particular, they will never find the maximal solution if it have more nonzero elements than the minimal solution. Moreover, to use the methods in [10, 11], the tensor $\mathcal{A}$ should be semi-symmetrized first, which increases computational work and makes the tensor much less sparse.

Example 2. We consider Example 1 with $k = 1$. Then equation (1) has two solutions: $[0, 1]^T$ and $[2, 1]^T$. Note that $I_0 = \{1\}$. We have $I = \{1\}$ for Theorem 16 since $\mathcal{A}$ is reducible with respect to $\{1\}$. The reduced equation is $x_2^3 = 1$ with nonnegative solution $x_2 = 1$, which is positive. A nonnegative solution of the original equation is then $[0, 1]^T$ by Theorem 16, but the other solution is lost in the reduction. We now apply Theorem 11. With $x_0 = [0, 0]^T$, we get $x_1 = [0, 1]^T$ and $x_2 = [0, 1]^T$. So $k_0 = 1$ and $I = \{1\}$ in Theorem 11. The reduced equation is $x_2^3 = 1$ with minimal nonnegative solution $x_2 = 1$, which is positive, and the minimal nonnegative solution of the original equation is then $[0, 1]^T$ by Theorem 11 (we got this solution after just one iteration). We then apply Theorem 15. With $x_0 = [3, 1]^T$, we get $x_1 = [18^{1/3}, 1]^T$. So $k_0 = 0$ and $I = \emptyset$ in Theorem 15. The equation is thus not reduced. We have $x_k = [t_k, 1]^T$, where t_k is determined by the iteration: $t_0 = 3$, $t_{k+1} = (2t_k^2)^{1/3}$. So t_k converges to 2 linearly with rate 2/3, and x_k converges to $[2,1]^T$ at the same rate. The linear convergence of $\{x_k\}$ is also guaranteed by Theorem 14 since condition (d) there is satisfied ($a_{1112} \neq 0$).

There are examples for which iteration (4) converges linearly with a rate very close to 1. We consider another extension of Example 1.1 in [1].

Example 3. We consider equation (1) with $\mathcal{A} \in \mathbb{R}^{[m,2]}$ given by $a_{11\ldots 1} = a_{22\ldots 2} = 1$ and $a_{11\ldots 12} = -2$ and with $b = [0,1]^T$. Then equation (1) has two solutions: $[0,1]^T$ and $[2,1]^T$. Note that $I_0 = \{1\}$. With $x_0 = [3,1]^T$, we get $x_1 = [(2 \cdot 3^{m-2})^{1/(m-1)}, 1]^T$. So $k_0 = 0$ and $I = \emptyset$ in Theorem 15. The equation is thus not reduced. We have $x_k = [t_k, 1]^T$, where t_k is determined by the iteration: $t_0 = 3$, $t_{k+1} = (2t_k^{m-2})^{1/(m-1)}$. So t_k converges to 2 linearly with rate $(m-2)/(m-1)$, and x_k converges to $[2,1]^T$ at the same rate. The linear convergence of $\{x_k\}$ is also guaranteed by Theorem 14 since $a_{11\ldots 12} \neq 0$. However, when m is large, the rate is very close to 1, so x_k converges to $[2,1]^T$ very slowly.

Remark 8. We have the freedom to choose the splitting $M = P - Q$ for iteration (4), when $\mathcal{A}$ is a nonsingular M-tensor. Recall that a smaller Q is going to give faster termwise convergence (see Theorems 10 and 13). For a dense tensor with order $m \geq 4$, computing $\mathcal{N}x^{m-1}$ will require $O(n^m)$ flops (for large n and fixed m) and solving the linear system $My = c$ requires $O(n^3)$ flops. So it is advisable to use $P = M$ and $Q = 0$ in the splitting $M = P - Q$ for a dense tensor with order $m \geq 4$. For a dense tensor with order 3, it is advisable to take P to be the lower triangular part of M or the upper triangular part of M, since solving the linear system $Py = c$ requires $O(n^2)$ flops in this case. If the tensor is very sparse, then the Jacobi iteration is often the most efficient one among the iterative methods we have discussed since it can utilize the sparsity very well and the computational work each iteration is lower than other methods.

4. Numerical experiments

In this section, we perform some numerical experiments to illustrate our theoretical results. In each example of equation (1), the tensor $\mathcal{A}$ is a very sparse Z-tensor and the vector b has many zero elements, so the minimal nonnegative solution (if exists) of equation (1) may have some zero elements. Our first example is obtained from Example 1 by introducing a parameter.

Example 4. We consider $\mathcal{A} \in \mathbb{R}^{4,2k}$, where $k \geq 2$ is an integer. The nonzero elements of $\mathcal{A}$ are given by

$$
\begin{aligned}
&\mathcal{A}(i,i,i,i) = 1,\ i = 1, \ldots, 2k, \\
&\mathcal{A}(2i-1, 2i-1, 2i-1, 2i) = -2,\ i = 1, \ldots, k, \\
&\mathcal{A}(2i, 2i+1, 2i+1, 2i+1) = -\epsilon,\ i = 1, \ldots, k-1, \\
&\mathcal{A}(2k, 2, 2, 2) = -\epsilon,
\end{aligned}
$$

where $\epsilon \geq 0$ is a parameter. We take $b = [0, 1, \ldots, 0, 1]^T \in \mathbb{R}^{2k}$.

Table 1: Number of iterations for finding the extremal solutions using Jacobi iteration for Example 4

k	5	50	500	5000	50000
min, $\epsilon = 0.05$	2	2	2	2	2
min, $\epsilon = 0.005$	2	2	2	2	2
max, $\epsilon = 0.05$	95	135	143	149	155
max, $\epsilon = 0.005$	67	71	74	77	80

The example is reduced to Example 1 when $\epsilon = 0$. For all $\epsilon \leq 0.06$, $\mathcal{A}x^3 > b$ for $x = [3, 1.4, \ldots, 3, 1.4]^T$, so $\mathcal{A}$ is a nonsingular M-tensor. For small ϵ, the equation $\mathcal{A}x^3 = b$ is expected to have 2^k nonnegative solutions that are perturbations of the 2^k solutions given in Example 1. We will use Jacobi iteration to find the two extremal solutions. If we wish (and time permits), we can also try to find $2^k - 2$ non-extremal nonnegative solutions in the following way. Let I be any proper subset of $I_0 = \{1, 3, \ldots, 2k-1\}$. Then $\mathcal{A}$ is reducible with respect to I. We can then find the maximal nonnegative solution of the reduced equation corresponding to the index set $I_c = [2k] \setminus I$, and then insert 0 at each position $i \in I$ to get a nonnegative solution of the original tensor equation. For $k = 3$ and $\epsilon = 0.05$, we actually followed through the above procedure to find 2^3 nonnegative solutions (including the two extremal solutions).

In Table 1, we report the number of iterations of Jacobi iteration for finding the two extremal solutions for $\epsilon = 0.05, 0.005$ and $k = 5, 50, 500, 5000, 50000$. For the minimal solution, we use $x_0 = 0$ in Jacobi iteration. For the maximal solution, we use $x_0 = [3, 1.4, \ldots, 3, 1.4]^T$ in Jacobi iteration. The stopping criterion is $\|\mathcal{A}(x_k)^3 - b\|_2 \leq 10^{-10}$. For this example, the minimal nonnegative solution is $[0, 1, \ldots, 0, 1, 0, (1+\epsilon)^{1/3}]^T$ and can be found by Jacobi iteration in two iterations. For finding the maximal solution, the Jacobi iteration requires more iterations for larger ϵ. The change in the number of iterations is small as k increases, and is largely due to the use of 2-norm (instead of ∞-norm) in the stopping criterion. For the maximal solution, the convergence of Jacobi iteration is linear since condition (d) in Theorem 14 is satisfied.

In our second example, both $\mathcal{A}$ and b are very sparse, but the minimal nonnegative solution of equation (1) is still positive. By computing the extremal nonnegative solutions, we can get a strong indication as to whether there is a unique positive solution for the equation.

Example 5. We consider $\mathcal{A} \in \mathbb{R}^{3,n}$, where $n \geq 4$ is even. The nonzero elements

Table 2: Number of iterations for finding the extremal solutions using Jacobi iteration for Example 5 with $n = 64$

ϵ	0.1	0.15	0.2	0.25	0.252	0.254	0.256	0.258
min, $tol = 10^{-8}$	96	141	265	1908	2524	3719	7039	61018
min, $tol = 10^{-10}$	122	182	346	2518	3330	4908	9283	80297
max, initial guess	1	2	2	12	39	66	129	1177
max, $tol = 10^{-8}$	106	164	311	2307	3192	4733	8957	77565
max, $tol = 10^{-10}$	132	205	392	2916	3998	5921	11202	96844

of $\mathcal{A}$ are given by

$$\begin{aligned}
&\mathcal{A}(i,i,i) = 1,\ i = 1,\ldots,n,\\
&\mathcal{A}(i,i+1,i+1) = -0.25,\ i = 1,\ldots,n-1,\\
&\mathcal{A}(i,i+1,i+2) = -\epsilon,\ i = 1,\ldots,n-2,\\
&\mathcal{A}(i,i-1,i-1) = -0.25,\ i = 2,\ldots,n/2,\\
&\mathcal{A}(i,i-1,i-2) = -0.25,\ i = 3,\ldots,n,
\end{aligned}$$

where $\epsilon > 0$ is a parameter. We take $b = [1, 0, \ldots, 0]^T \in \mathbb{R}^n$.

The tensor $\mathcal{A}$ is irreducible and thus every nonnegative solution of $\mathcal{A}x^2 = b$ must be positive (see [10, Theorem 2.4]). The majorization matrix M associated with $\mathcal{A}$ is block upper triangular and $M^{-1}b > 0$ does not hold. The uniqueness of positive solutions is then unknown (see [13, Theorem 4.5]). For this example, $\mathcal{A} = \mathcal{I} - \mathcal{B}$, where $\mathcal{B} \geq 0$. We take $n = 64$. When $\epsilon = 0.258$, we find $\rho(\mathcal{B}) \approx 0.9998$. When $\epsilon = 0.259$, we find $\rho(\mathcal{B}) \approx 1.0007$. Since $\rho(\mathcal{B})$ increases as ϵ increases, $\mathcal{A}$ is a nonsingular M-tensor for $\epsilon \leq 0.258$, and is not a nonsingular M-tensor for $\epsilon \geq 0.259$.

We now use the Jacobi iteration to find the extremal nonnegative solutions of $\mathcal{A}x^2 = b$ for some values of $\epsilon \leq 0.258$. For the minimal solution, we use Jacobi iteration with $x_0 = 0$ and stop the iteration when $\|\mathcal{A}(x_k)^2 - b\|_2 \leq tol$. For the maximal solution, we first take $\hat{b} = [2, 1, \ldots, 1]^T \in \mathbb{R}^{64}$ and apply the Jacobi iteration with $\hat{x}_0 = 0$ to the equation $\mathcal{A}\hat{x}^2 = \hat{b}$ and stop the iteration as soon as $\mathcal{A}(\hat{x}_k)^2 > b$ is satisfied (the number of iterations for each ϵ is recorded in Table 2 in the row for "max, initial guess"). We then use the Jacobi iteration with $x_0 = \hat{x}_k$ to compute the maximal solution of $\mathcal{A}x^2 = b$ and stop the iteration when $\|\mathcal{A}(x_k)^2 - b\|_2 \leq tol$. In Table 2, we have given the number of iterations for finding the minimal and maximal solutions with $tol = 10^{-8}$ and $tol = 10^{-10}$. For each ϵ value in the table, we also compute $\|x_{\max} - x_{\min}\|_2$ for the computed $x_{\max}$ and $x_{\min}$ with $tol = 10^{-10}$. We find these values to be between 2.08×10^{-8} and 1.04×10^{-7}. This strongly suggests that the tensor equation has a unique positive solution for these ϵ values. The convergence of the Jacobi iteration is linear for this example since condition (b) in Theorem 11 (which is the same as condition (b) in Theorem 14) is satisfied (we just need to check that $\mathcal{L}e$ is an irreducible matrix). Moreover, the Jacobi iteration for the minimal solution and for the maximal solution should have (roughly) the same rate of convergence since the minimal solution and the maximal solution are (roughly) the same. For fixed ϵ, we see from Table 2 that the number of iterations for the maximal solution is significantly larger than that for the minimal solution. This is because of the difference in the early stage of iterations. For $\epsilon = 0.258$ for example, to reduce the residual error from 10^{-8} to 10^{-10}, we need 19279 iterations for both the minimal solution and the maximal solution. To have better efficiency, we can apply Newton's method as a correction method after the residual has been sufficiently reduced by Jacobi iteration. For example, to compute the minimal solution when $\epsilon = 0.258$, we can reduce the residual to under 10^{-2} after 2983 iterations of Jacobi iteration (with $x_0 = 0$), and then reduce the residual to about 1.4×10^{-14} after 5 steps of Newton's method. We also note that for the computed $x_{\max}$ and $x_{\min}$, the last element can be very small. For $\epsilon = 0.1$ for example, we have $(x_{\min})_{64} = 6.7550 \times 10^{-9}$ and $(x_{\max})_{64} = 6.7612 \times 10^{-9}$. But this does not cause any problems with the Jacobi iteration. For $\epsilon = 0.259$, the sequence from the Jacobi iteration (with $x_0 = 0$) increases without bound, so the tensor equation does not have any nonnegative solutions when $\epsilon \geq 0.259$ (see Remark 4 and Corollary 4). The exact turning point ϵ^* is somewhere between 0.258 and 0.259. We see from Table 2 that the Jacobi iteration requires a large number of iterations when ϵ approaches ϵ^* from the left.

We end this section by one more example, for which linear convergence of the iterative methods in Section 3 is not always guaranteed by the convergence results there.

Example 6. We consider $\mathcal{A} \in \mathbb{R}^{3,n}$, where $n = 2p$ and p is even. The nonzero elements of $\mathcal{A}$ are given by

$$\mathcal{A}(i,i,i) = 1,\ i = 1,\dots,n,$$
$$\mathcal{A}(i,p-i+1,p-i+1) = -0.25,\ \mathcal{A}(i,i,i+1) = -\epsilon,\ i = 1,\dots,p,$$
$$\mathcal{A}(i,n-i+1,n-i+1) = -0.25,\ \mathcal{A}(i,i,i-1) = -0.25,\ i = p+1,\dots,n,$$

where $\epsilon > 0$ is a parameter. We take $b = [1,0,\dots,0]^T \in \mathbb{R}^n$.

For this example, $\mathcal{A} = \mathcal{I} - \mathcal{B}$, where $\mathcal{B} \geq 0$. We now take $n = 16$. When $\epsilon = 0.79$, we find $\rho(\mathcal{B}) \approx 0.997$. When $\epsilon = 0.8$, we find $\rho(\mathcal{B}) \approx 1.006$. Therefore, $\mathcal{A}$ is a nonsingular M-tensor for $\epsilon \leq 0.79$, and is not a nonsingular M-tensor for $\epsilon \geq 0.8$. We have $I_0 = [16] \setminus \{1\}$, and may use definition to determine the largest set $I \subseteq I_0$ such that $\mathcal{A}$ is reducible with respect to I. It turns out that $I = I_2 \cup I_3 \cup I_4$, where $I_2 = \{2,7,10,15\}$, $I_3 = \{3,6,11,14\}$ and $I_4 = \{4,5,12,13\}$. This set I can also be found quickly by using Theorem 11. With $x_0 = 0$ in the Jacobi iteration, we find that $k_0 = 3$ in Theorem 11 and $\{i \mid (x_3)_i = 0\} = I_2 \cup I_3 \cup I_4$. Note that $\mathcal{A}$ is also reducible with respect to $I_2, I_3, I_4, I_2 \cup I_3, I_2 \cup I_4, I_3 \cup I_4$.

For $\epsilon = 0.7$ and $\epsilon = 0.79$, we find four nonnegative solutions of the tensor equation by applying the Jacobi iteration to the tensor equation or to suitably

Table 3: Four nonnegative solutions for Example 6 with $n = 16$ and $\epsilon = 0.7$

$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$
1.0620	1.1861	1.2421	1.3496
0	0.2981	0.4233	0.6512
0	0	0.2441	0.6024
0	0	0	0.5901
0	0	0	0.6145
0	0	0.4882	0.6755
0	0.5962	0.6538	0.7731
0.7148	0.7984	0.8361	0.9085
0.4578	0.5113	0.5354	0.5818
0	0.3688	0.4006	0.4660
0	0	0.2992	0.4010
0	0	0	0.3614
0	0	0	0.3437
0	0	0.1220	0.3472
0	0.1490	0.2275	0.3719
0.5310	0.6120	0.6501	0.7229

reduced tensor equations. The four solutions (rounded to four decimal places) are given in Table 3 for $\epsilon = 0.7$ and in Table 4 for $\epsilon = 0.79$. Note that when ϵ is increased from 0.7 to 0.79, all four solutions increase, as predicted by Corollaries 4 and 7 (applied to the original tensor equation or the reduced tensor equations).

We now examine the convergence speed of the iterative methods in Section 3 for finding the extremal solutions of $\mathcal{A}x^2 = b$ for Example 6. The stopping criterion is $\|\mathcal{A}(x_k)^2 - b\|_2 \leq 10^{-10}$. Recall that M is the majorization matrix associated with $\mathcal{A}$ and $M = P - Q$ is a splitting of M. Of particular interest are the following four methods.

- Method 1: P is the diagonal part of M (Jacobi iteration).
- Method 2: P is the lower triangular part of M.
- Method 3: P is the upper triangular part of M.
- Method 4: $P = M$.

For the minimal solution, we use Methods 1–4 with $x_0 = 0$ and stop the iteration when $\|\mathcal{A}(x_k)^2 - b\|_2 \leq 10^{-10}$. The numbers of iterations are reported in Table 5. For the maximal solution, we first take $\hat{b} = [2, 1, \ldots, 1]^T \in \mathbb{R}^{16}$ and apply the Jacobi iteration with $\hat{x}_0 = 0$ to the equation $\mathcal{A}\hat{x}^2 = \hat{b}$ and stop the iteration as soon as $\mathcal{A}(\hat{x}_k)^2 > b$ is satisfied (the number of iterations for each ϵ is recorded in Table 6 in the row for "initial guess"). We then use Methods 1–4 with $x_0 = \hat{x}_k$ to compute the maximal solution of $\mathcal{A}x^2 = b$ and stop the iteration

Table 4: Four nonnegative solutions for Example 6 with $n = 16$ and $\epsilon = 0.79$

$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$
1.0700	1.2532	1.3819	4.9333
0	0.3756	0.6120	5.1982
0	0	0.3719	5.4377
0	0	0	5.4724
0	0	0	5.3017
0	0	0.7438	4.9235
0	0.7512	0.8827	4.3318
0.7611	0.8914	0.9830	3.5092
0.4874	0.5709	0.6295	2.2472
0	0.4537	0.5270	2.4649
0	0	0.4436	2.7891
0	0	0	3.0223
0	0	0	3.1400
0	0	0.1859	3.1395
0	0.1878	0.3301	3.0210
0.5350	0.6505	0.7335	2.8730

Table 5: Number of iterations for finding the minimal solution using Methods 1–4 for Example 6 with $n = 16$

ϵ	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
Method 1	58	69	84	105	138	195	321	842
Method 2	42	50	61	77	102	145	240	631
Method 3	55	66	80	100	132	187	308	810
Method 4	40	48	58	73	96	137	227	599

Table 6: Number of iterations for finding the maximal solution using Methods 1–4 for Example 6 with $n = 16$

ϵ	0.70	0.72	0.74	0.76	0.78	0.79
initial guess	3	3	4	13	34	139
Method 1	326	407	549	877	2068	7924
Method 2	269	335	452	722	1704	6535
Method 3	292	365	493	789	1865	7152
Method 4	235	294	397	635	1501	5757

when $\|\mathcal{A}(x_k)^2 - b\|_2 \leq 10^{-10}$. The numbers of iterations are reported in Table 6. As predicted by Theorems 10 and 13, the convergence of Method 4 is faster than the convergence of Methods 2 and 3, and the convergence of Method 1 is slower than the convergence of Methods 2 and 3. For $\epsilon \geq 0.8$, the tensor equation has no positive solutions. Indeed, if $x > 0$ is a solution, then $(0.001\mathcal{I} + \mathcal{A})x^2 > 0$ and $0.001\mathcal{I} + \mathcal{A} = 1.001\mathcal{I} - \mathcal{B}$ would be a nonsingular M-tensor, contradicting the fact $\rho(\mathcal{B}) > 1.005$. The exact turning point ϵ^* is somewhere between 0.79 and 0.8. We see from Table 6 that, for finding the (positive) maximal solution, Methods 1–4 require a large number of iterations when ϵ approaches ϵ^* from the left. As we have mentioned for Example 5, we can apply Newton's method as a correction method after the residual has been sufficiently reduced, to have better efficiency in those situations. The turning point for the minimal solution is different. The minimal solution has positive elements only in positions $i \in I_1 = \{1, 8, 9, 16\}$ and may be found by applying Methods 1–4 to the reduced tensor equation corresponding to the index set I_1. The reduced tensor has the form $\hat{\mathcal{A}} = \mathcal{I} - \hat{\mathcal{B}}$, where $\hat{\mathcal{B}} \in \mathbb{R}^{3,4}$ is such that $\rho(\hat{\mathcal{B}}) \approx 0.975$ for $\epsilon = 1.4$ and $\rho(\hat{\mathcal{B}}) \approx 1.014$ for $\epsilon = 1.5$. This is why we can still find the minimal nonnegative solution for $\epsilon = 1.4$. The tensor equation has no nonnegative solutions for $\epsilon \geq 1.5$. When ϵ approaches the turning point from the left, Methods 1–4 will require a large number of iterations. We can apply Newton's method as a correction method for better efficiency, but Newton's method has to be applied to the reduced equation (it is not applicable to the original equation due to the presence of zero elements in the approximate solutions). The rate of convergence appears linear in all cases in Tables 5 and 6. Linear convergence is guaranteed for Methods 2 and 4 for finding the minimal solution since $P^{-1}b > 0$ holds for the reduced equation (see Theorem 11). However, none of our sufficient conditions for linear convergence is satisfied for Methods 1 and 3 for finding the minimal solution, and for Methods 1–4 for finding the maximal solution.

For Example 6, the tensor $\mathcal{A}$ is very sparse and Method 1 (Jacobi iteration) can exploit the sparsity very well and may still be the most efficient among the four methods. When n is one million and $\epsilon = 0.7$, we just need a few seconds to get the minimal solution (without using dimension reduction). The minimal solution only has 4 nonzero elements, a very sparse solution indeed. We find that the iterates x_3 and x_4 both have 4 nonzero elements. If we wish, we can perform the remaining computation on a reduced tensor equation with dimension 4 (see Theorem 11).

5. Conclusion

We have presented new proofs for the existence of extremal nonnegative solutions of the M-tensor equation with a nonnegative right-side vector by using some simple fixed-point iterations. We have studied these iterative methods and their generalizations. With a suitable starting point, each of these methods has monotonic convergence (to the maximal nonnegative solution or to the minimal nonnegative solution) and the rate of convergence is (at least) linear under some mild assumptions. These methods are currently the only methods that are guaranteed to compute the maximal nonnegative solution or the minimal nonnegative solution with suitable initial guesses, and they can exploit the sparsity of the coefficient tensor easily. There are examples for which the presented iterations have linear convergence at a rate very close to 1. In those situations, one may use Newton's method as a correction method to improve efficiency. Iterative methods with faster global convergence (without increasing computational work each iteration by too much) are still desirable. Those methods should be ones that can compute the maximal nonnegative solution and/or the minimal nonnegative solution, not just any one of the nonnegative solutions.

Acknowledgments

This work was supported in part by a grant from Natural Sciences and Engineering Research Council of Canada.

Declarations of interest

None.

Data availability

No data was used for the research described in the article.

References

[1] X. Bai, H. He, C. Ling, G. Zhou, A nonnegativity preserving algorithm for multilinear systems with nonsingular M-tensors, Numer. Algorithms **87** (2021) 1301–1320.

[2] A. Berman, R. Plemmons, Nonnegative Matrices in the Mathematical Sciences, SIAM, Philadephia, 1994.

[3] W. Ding, L. Qi, Y. Wei, M-tensors and nonsingular M-tensors, Linear Algebra Appl. **439** (2013) 3264–3278.

[4] W. Ding, Y. Wei, Solving multi-linear systems with M-tensors, J. Sci. Comput. **68** (2016) 683–715.

[5] C.-H. Guo, Nonsymmetric algebraic Riccati equations and Wiener–Hopf factorization for M-matrices, SIAM J. Matrix Anal. Appl. **23** (2001) 225–242.

[6] C.-H. Guo, A. J. Laub, On the iterative solution of a class of nonsymmetric algebraic Riccati equations, SIAM J. Matrix Anal. Appl. **22** (2000) 376–391.

[7] L. Han, A homotopy method for solving multilinear systems with M-tensors, Appl. Math. Lett. **69** (2017) 49–54.

[8] H. He, C. Ling, L. Qi, G. Zhou, A globally and quadratically convergent algorithm for solving multilinear systems with M-tensors, J. Sci. Comput. **76** (2018) 1718–1741.

[9] D.-H. Li, H.-B. Guan, X.-Z. Wang, Finding a nonnegative solution to an M-tensor equation, Pac. J. Optim. **16** (2020) 419–440.

[10] D.-H. Li, H.-B. Guan, J.-F. Xu, Inexact Newton method for M-tensor equations, Pac. J. Optim. **17** (2021) 617–643.

[11] D.-H. Li, J.-F. Xu, H.-B. Guan, Newton's method for M-tensor equations, J. Optim. Theory Appl. **190** (2021) 628–649.

[12] C.-S. Liu, C.-H. Guo, W.-W. Lin, Newton–Noda iteration for finding the Perron pair of a weakly irreducible nonnegative tensor, Numer. Math. **137** (2017) 63–90.

[13] D. Liu, W. Li, S.-W. Vong, The tensor splitting with application to solve multi-linear systems, J. Comput. Appl. Math. **330** (2018) 75–94.

[14] Z. Luo, L. Qi, N. Xiu, The sparsest solutions to Z-tensor complementarity problems, Optim. Lett. **11** (2017) 471–482.

[15] M. Ng, L. Qi, G. Zhou, Finding the largest eigenvalue of a nonnegative tensor, SIAM J. Matrix Anal. Appl. **31** (2009) 1090–1099.

[16] Q. Ni, L. Qi, A quadratically convergent algorithm for finding the largest eigenvalue of a nonnegative homogeneous polynomial map, J. Global Optim. **61** (2015) 627–641.

[17] K. J. Pearson, Essentially positive tensors, Int. J. Algebra **4** (2010) 421–427.

[18] L. Qi, Z. Luo, Tensor Analysis: Spectral Theory and Special Tensors, SIAM, Philadephia, 2017.

[19] Y. Wei, W. Ding, Theory and Computation of Tensors: Multi-Dimensional Arrays, Academic Press, Amsterdam, 2016.

[20] J. Zhang, C. Bu, An iterative method for finding the spectral radius of an irreducible nonnegative tensor, Comput. Appl. Math. **40** (2021), Paper No. 8, 14 pp.