Constructing a High-Order Globally Convergent Iterative Method for Calculating the Matrix Sign Function

Haifa Bin Jebreen

Mathematics Department, College of Science, King Saud University, Riyadh, Saudi Arabia

Correspondence should be addressed to Haifa Bin Jebreen; hjebreen@ksu.edu.sa

Received 9 April 2018; Accepted 29 May 2018; Published 21 June 2018

This work is concerned with the construction of a new matrix iteration in the form of an iterative method which is globally convergent for finding the sign of a square matrix having no eigenvalues on the axis of imaginary. Toward this goal, a new method is built via an application of a new four-step nonlinear equation solver on a particulate matrix equation. It is discussed that the proposed scheme has global convergence with eighth order of convergence. To illustrate the effectiveness of the theoretical results, several computational experiments are worked out.

1. Preliminaries

The sign function for the scalar case is expressed by

\[
\text{sign}(\omega) = \begin{cases} 
1, & \text{Re}(\omega) > 0, \\
-1, & \text{Re}(\omega) < 0,
\end{cases}
\]  

wherein \( \omega \in \mathbb{C} \) is not located on the imaginary axis. Roberts in [1] for the first time extended this definition for matrices, which has several important applications in scientific computing, for example see [2–4] and the references cited therein. For example, the off-diagonal decay of the matrix function of sign is also a well-developed area of study in statistics and statistical physics [5].

To proceed formally, let us consider that \( A \in \mathbb{C}^{n \times n} \) is a square matrix possessing no eigenvalues on the axis of imaginary. We consider

\[
A = TJT^{-1},
\]  

as a form of Jordan canonical written such that

\[
J = \text{diag}(J_1, J_2),
\]  

and the eigenvalues of \( J_1 \in \mathbb{C}^{p \times p} \) are in the open left half-plane, while the eigenvalues of \( J_2 \in \mathbb{C}^{q \times q} \) are in the open right half-plane. It is now possible to write the following [6]:

\[
S = \text{sign}(A) = T \begin{pmatrix} -I_p & 0 \\ 0 & I_q \end{pmatrix} T^{-1},
\]  

wherein \( p + q = n \). Noting that \( \text{sign}(A) \) is definable once \( A \) is nonsingular.

This procedure takes into account a clear application of the form of Jordan canonical and of the matrix \( T \). Herein none of the matrices \( T \) and \( J \) are unique. However, it is possible to investigate that \( \text{sign}(A) \) as provided in (4) does not rely on the special selection of \( T \) or \( J \).

Here, a simpler interpretation for the sign matrix in the case of Hermitian case (namely, all eigenvalues are real) can be given by

\[
S = U \text{diag}(\text{sign}(\lambda_1), \ldots, \text{sign}(\lambda_n)) U^*,
\]  

wherein

\[
U^*AU = \text{diag}(\lambda_1, \ldots, \lambda_n),
\]  

is a diagonalization of the square matrix \( A \).

The significance of calculating and finding \( S \) in (4) is because of the point that the function of sign plays a central role in matrix functions theory, specially for principal matrix...
roots and the polar decomposition; for more one may refer to [7–9].

Bini et al. in [10] proved that the principal $p$-th root of the matrix $A$ could be written as a multiple of the $(2,1)$-block associated with the sign matrix $\text{sign}(C)$, associated with the block companion matrix:

$$
C = \begin{pmatrix}
0 & I \\
0 & I \\
\vdots & \ddots & \ddots & \ddots \\
A & 0 & \cdots & I
\end{pmatrix} \in \mathbb{C}^{p \times p}.
$$

(7)

It is requisite to focus of the most general case, i.e., when all the eigenvalues are complex rather than being narrow over a range of specific matrices, such as in (5).

An important point goes to the fact that although $\text{sign}(A)$ is a square root of the unit matrix, it is not equal to $I$ or $-I$ unless the $A$’s spectrum locates completely in the open right or left half-plane(s), respectively. Thus, the sign function is a nonprimary square root of $I$.

Apart from (4), an efficient way to derive matrix iterations for some matrix functions is to apply the zero-finding iterative methods for solving operator equations which here is a matrix equation. Toward this goal, it is necessary to tackle a nonlinear equation as comes next:

$$
F(X) := X^2 - I = 0,
$$

(8)

wherein $I$ is a unit matrix, so as to propose matrix methods for $S$. The main motivation in this work is to extend the recently published results of the literature [11, 12] in this category by providing a useful novel method for calculating $\text{sign}$ matrix. Furthermore, the proposed procedure can be applied for the calculation of polar decomposition, principal matrix square root, and several other scientific computing problems.

After this brief introduction about the matrix function of $\text{sign}$ in Section 1, the remaining sections of this study are given as comes next. Section 2 shortly surveys the existing matrix iterations and their importance for computing $S$. In Section 3, it is discussed how we construct a new method having global convergence behavior and not belonging to the class of Padé family of iterations for computing $S$. It too manifests that the proposed scheme is convergent with high order of convergence. Computational reports are furnished to illustrate the higher computational precision of the constructed solvers in Section 4. A final remark of the manuscript is given in Section 5 with some directions for future studies.

2. The Literature

In the current research work, iterative methods are the main focus for calculating $S$. As a matter of fact, such matrix iteration methods are Newton-type schemes that are in essence fixed-point schemes by providing a convergent matrix sequence by imposing a sharp initial value.

The connection of matrix iterative expressions with the function of sign is not that straightforward, but in practice, such methods could be constructed by considering a suitable root-finding method to the nonlinear matrix equation (8). Noting that $\text{sign}(A)$ is a solution of this equation (refer to [12] and the references cited therein).

Applying the classic Newton’s method (NM) to (8) yields

$$
X_{k+1} = \frac{1}{2} \left( X_k + X_k^{-1} \right).
$$

(9)

An inversion-free version of (9), called Newton-Schultz iteration [6], is defined by

$$
X_{k+1} = \frac{1}{2} X_k \left( 3I - X_k^2 \right),
$$

(10)

by applying the well-known Schulz inverse-finder in order to remove the computation of the inverse matrix per computing step.

The Newton-Schultz scheme is a second-order convergent, inversion-free method in calculating the sign matrix, but it suffers from the drawback that its convergence unlike the Newton’s method (9) is local.

Analogously, the third-order convergent Halley’s method (HM) [13] for calculating the sign matrix is defined by

$$
X_{k+1} = \left[ I + 3X_k^2 \right] \left[ X_k \left( 3I + X_k^2 \right) \right]^{-1}.
$$

(11)

It is noted that all the above-mentioned schemes are particular cases of the Padé family presented and discussed in [13, 14]. The Padé approximation belongs to a broader category of rational approximations. Coincidentally, the best uniform approximation of the sign function on a pair of symmetric but disjoint intervals can be expressed as a rational function.

Recently a fourth-order iterative method was furnished in [15] as follows:

$$
X_{k+1} = \left[ I + 18X_k^2 + 13X_k^4 \right] \left[ X_k \left( 7I + 22X_k^2 + 3X_k^4 \right) \right]^{-1}.
$$

(12)

3. Construction of a New Matrix Method

Assume the following nonlinear equation:

$$
g(x) = 0,
$$

(13)

wherein $g : D \subseteq \mathbb{C} \rightarrow \mathbb{C}$ is a scalar function. In what follows, let us first present a new scheme in nonlinear equation solving. The idea of increasing the order (see, e.g., [16, 17]) is to consider several substeps, while the newly appearing first derivatives are approximated via a secant-like approximation. Thus, we may write

$$
y_k = x_k - s_k,
$$

$$
z_k = x_k - \left( 1 + \frac{g(y_k)}{g(x_k) - (5/3) g(y_k)} \right) \frac{g(x_k)}{g'(x_k)},
$$
Figure 1: Basins of attractions for (9) in left and (10) in right (shading is done based on the number of cycles to achieve the convergence).

\[ w_k = z_k - \frac{g(z_k)}{g'_{y_k, z_k}} \]
\[ x_{k+1} = w_k - \frac{g(w_k)}{g'_{w_k, z_k}} \]  \hspace{1cm} (14)

whereas the first-order divided difference operator is defined by

\[ g[\ell, h] = \frac{g(h) - g(\ell)}{h - \ell}. \]  \hspace{1cm} (15)

Here the point is that the new method should not aim at having the optimal convergence order (in the sense of Kung-Traub, see, e.g., [16]) since such schemes lose the global convergence behavior when applied to (13). Since the final objective is to propose a method for matrix sign, we should take two things into account, which are having global convergence behavior and the novelty. In fact, the optimality conjecture of Kung-Traub is not useful once we extend iterative methods for calculating the sign of a matrix. Because, the optimality conjecture here ruins the final structure of the matrix method.

One may also ask why method (14) has been selected and what are the other ways for improving it. To answer these, we recall that the best way to improve the performance of (14) is to add one Newton's substep at the end of its fourth step, which is costly since it includes the computation of the first derivative per cycle. In a similar way as in (14), we can add a secant-like substep and increase the convergence order. Generally speaking, a family of iterations can now be constructed in this way. On the other hand, since very higher order methods may not be useful in double precision arithmetic, namely, the practical environment that most researchers work in, we here only provide (14) and discuss its application and extension for matrix sign.

Theorem 1. Assume that \( \alpha \in D \) is a simple root of a function \( g: D \subseteq \mathbb{C} \rightarrow \mathbb{C} \), which is sufficiently differentiable and contains \( x_0 \) as an initial value. Accordingly, the iterative expression (14) reads

\[ e_{k+1} = \frac{1}{g} (c_{j})^7 e_{k}^{8} + O(e_{k}^{9}), \]  \hspace{1cm} (16)

where \( c_{j} = g^{(j)}(\alpha)/j! g'(\alpha) \), and \( e_{k} = x_{k} - \alpha \).

Proof. The steps of proving the convergence order for this iterative method are via Taylor expansion, which is straightforward. \( \square \)

Applying (14) on the matrix equation (8) will yield a novel matrix scheme to calculate (4) in its reciprocal form as follows:

\[ X_{k+1} = X_k \left( 12I + 200X_k^2 + 560X_k^4 + 344X_k^6 + 36X_k^8 \right) \Xi_k^{-1}, \]  \hspace{1cm} (17)

wherein

\[ \Xi_k = \left[ I + 64X_k^2 + 406X_k^4 + 532X_k^6 + 145X_k^8 + 4X_k^{10} \right], \]  \hspace{1cm} (18)

and the initial approximation is

\[ X_0 = A. \]  \hspace{1cm} (19)

Applying (14) on the nonlinear equation \( g(x) = x^2 - 1 \) contributes a global convergence in the complex plane (excluding the values locating on the axis of imaginary). The basins of attraction for (10) (locally convergent) and (9) (globally convergent) are portrayed in Figure 1. This global behavior of the proposed scheme, that is kept for matrix case, has been shown in Figures 2-3.

To draw the basins of attractions, we consider a square \( \Gamma = [-2, 2] \times [-2, 2] \in \mathbb{C} \) and allocate a color to any point.
Figure 2: Basins of attractions for (17) in left and its density plot on the same domain in right (shading is done based on the number of cycles to achieve the convergence).

Figure 3: The density plot of the basins of attractions for (17) on $[-50, 50] \times [-50, 50]$ in left and on $[-100, 100] \times [-100, 100]$ in right (shading is done based on the number of cycles to achieve the convergence).

$x_0 \in \Gamma$ according to the simple zero, at which the new methods (or the existing methods for comparisons) converge. Subsequently, we highlight the point as black once the scheme diverge. Herein, we consider the stopping termination for convergence as $|g(x_k)| \leq 10^{-2}$. Using a different stopping criterion $|x_{k+1} - x_k| \leq 10^{-2}$, the density plot of the basins of attraction only for the new high order of convergence method (17) are brought forward in Figures 2-3 on different domains.

3.1. Convergence Study. Now, it is shown that the proposed schemes are convergent, under standard conditions, namely, when there are no pure imaginary eigenvalues in the absence of rounding errors.

**Theorem 2.** Assume that $A \in \mathbb{C}^{n \times n}$ possess no eigenvalues on the axis of imaginary. Accordingly, the iterations $\{X_k\}_{k=0}^{\infty}$ expressed by (17) is convergent to $S$, using (19).

**Proof.** Assume that $R$ is a rational operator in accordance with (17). Since $X \in \mathbb{C}^{n \times n}$ has a form of Jordan canonical, there exists a matrix $Z$, so that

$$X = ZJZ^{-1}. \tag{20}$$

It is recalled that diagonalizable and nondiagonalizable matrices have a Jordan normal form $A = TJT^{-1}$, whereas $J$ includes the Jordan blocks. So,

$$R(X) =ZR(J)Z^{-1}. \tag{21}$$
An eigenvalue $\lambda$ of $X_k$ is transferred into an eigenvalue of $R(\lambda)$ of $X_{k+1}$ via the iterative expression (17). This relevance and relation among the eigenvalues show that it is required to search how $R(A)$ transforms the complex plane into itself. It is recalled that $R$ has the feature of sign preservation, namely,

$$\text{sign } (R(x)) = \text{sign } (x) \quad x \in \mathbb{C}. \quad (22)$$

Moreover, it should have the global convergence, that is, the sequence defined by

$$x_{k+1} = R(x_k), \quad (23)$$

with $x_0 = x$ converges to $\text{sign}(x)$ while $x$ is not located on the axis of imaginary. At this moment, assume that the square matrix $A$ have a form of canonical Jordan considered just like [6, p. 107]:

$$Z^{-1}AZ = \Lambda = \begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}, \quad (24)$$

wherein $Z$ is a not singular and $C, N$ are Jordan blocks in association with eigenvalues locating in $\mathbb{C}^-$ and $\mathbb{C}^+$, respectively. We show by $\lambda_1, \ldots, \lambda_p$ and $\lambda_{p+1}, \ldots, \lambda_n$ locating on the main diagonals of blocks $C$ and $N$, respectively. Applying (24), it is possible to write

$$\text{sign } (A) = Z \begin{bmatrix} -I_p & 0 \\ 0 & I_{n-p} \end{bmatrix} Z^{-1}. \quad (25)$$

Taking (25) into account, it is easy to deduce

$$\text{sign } (\lambda) = \text{sign } \left( Z^{-1}AZ \right) = Z^{-1} \text{sign } (A) Z \begin{bmatrix} \text{sign } (\lambda_1) & 0 \\ \vdots & \ddots \\ \text{sign } (\lambda_p) & \text{sign } (\lambda_{p+1}) \\ 0 & \cdots & \text{sign } (\lambda_n) \end{bmatrix}. \quad (26)$$

From $D_0 = Z^{-1}AZ$, we define

$$D_k = Z^{-1}X_kZ, \quad (27)$$

in order to obtain a convergent sequence to $\text{sign}(A)$. Thence, from the scheme (17), we simply could obtain

$$D_{k+1} = D_k \left( 12 I_k + 200 D_k^2 + 560 D_k^4 + 344 D_k^6 + 532 D_k^8 + 145 D_k^{10} \right)$$

$$\times \begin{bmatrix} 1 + 64 D_k^2 + 406 D_k^4 + 532 D_k^6 + 145 D_k^8 + 4 D_k^{10} \\ \vdots \\ 1 + 64 D_k^2 + 406 D_k^4 + 532 D_k^6 + 145 D_k^8 + 4 D_k^{10} \end{bmatrix}^{-1}. \quad (28)$$

By rearranging (28) as $n$ uncoupled scalar iterations as comes next:

$$d_{k+1}^i = \frac{12 d_k^i + 200 d_k^i + 560 d_k^i + 344 d_k^i + 532 d_k^i + 145 d_k^i + 4 d_k^{10}}{1 + 64 d_k^i + 406 d_k^i + 532 d_k^i + 145 d_k^i + 4 d_k^{10}} \quad (29)$$

where

$$d_k^i = (D_k)_{ij}, \quad 1 \leq i \leq n. \quad (30)$$

Using (28) and (29), we should investigate the convergence of $\{d_k^i\}$ to $\text{sign}(\lambda_i)$, for all $1 \leq i \leq n$. From (29) and because the eigenvalues of $A$ are not pure imaginary, it is possible to write

$$\text{sign } (\lambda_i) = s_i = \pm 1. \quad (31)$$

Thus, we attain

$$\frac{d_{k+1}^i - s_i}{d_k^i + s_i} = \left( \frac{s_i + d_k^i}{s_i + d_k^i} \right)^8 \left( \frac{s_i - 2 d_k^i}{s_i + 2 d_k^i} \right)^2. \quad (32)$$

Noting that the factor $(s_i - 2 d_k^i)/(s_i + 2 d_k^i)$, is bounded due to choosing an appropriate initial matrix (19). Since

$$\left| \frac{d_0^i - s_i}{d_0^i + s_i} \right| < 1, \quad (33)$$

we attain

$$\lim_{k \to \infty} \left| \frac{d_{k+1}^i - s_i}{d_k^i + s_i} \right| = 0, \quad (34)$$

and, therefore,

$$\lim_{k \to \infty} d_k^i = s_i = \text{sign } (\lambda_i). \quad (35)$$

At this point, it possible to derive that

$$\lim_{k \to \infty} D_k = \text{sign } (A). \quad (36)$$

Noting that if $D_0$ is not diagonal, the relation among the iterates’ eigenvalues must be dealt with for (17). The eigenvalues of $X_k$ are transformed from the iteration $k$ to the iteration $k+1$ via

$$\lambda_{k+1}^i = \left( 12 \lambda_k^i + 200 \lambda_k^i + 560 \lambda_k^i + 344 \lambda_k^i + 532 \lambda_k^i + 145 \lambda_k^i + 4 \lambda_k^{10} \right)$$

$$\times \begin{bmatrix} 1 + 64 \lambda_k^i + 406 \lambda_k^i + 532 \lambda_k^i + 145 \lambda_k^i + 4 \lambda_k^{10} \\ \vdots \\ 1 + 64 \lambda_k^i + 406 \lambda_k^i + 532 \lambda_k^i + 145 \lambda_k^i + 4 \lambda_k^{10} \end{bmatrix}^{-1}. \quad 1 \leq i \leq n. \quad (37)$$

The relation (37) manifests that the eigenvalues generally are convergent to $s_i = \pm 1$, namely,

$$\lim_{k \to \infty} \left| \frac{\lambda_{k+1}^i - s_i}{\lambda_{k+1}^i + s_i} \right| = 0. \quad (38)$$
Ultimately, it would be easy to write that
\[
\lim_{k \to \infty} X_k = Z \left( \lim_{k \to \infty} D_k \right) Z^{-1} = Z \text{sign} (A) \ Z^{-1}
\]
\[
= \text{sign} (A).
\]
This ends the convergence proof for (17) to calculate $S$.

Now by considering (18) and the facts that $X_k$ are rational functions of $A$, so similar to $A$, commute with $S$, and $S^2 = I$, $S^{-1} = S$, $S[j] = I$, and $S[2j+1] = S$, $j \geq 1$. It can be shown that the new scheme reads the following error inequality:
\[
\|X_{k+1} - S\| \leq \|Z^{-1} \| \|I - 2X_k\|^2 \|X_k - S\|^6.
\]
Inequality (40) shows the eighth order of convergence.

A scaling technique to accelerate the initial phase of convergence is normally requisite since the convergence rate cannot be seen in the initial iterates. Such an idea was discussed fully in [18] for Newton's method. A robust procedure to improve the initial convergence speed is to scale the iterations before each iteration; i.e., $X_k$ should be moved to $\mu_k X_k$.

If the scaling parameter (for the Newton's method) is defined by [18],
\[
\zeta_k = \begin{cases} 
\sqrt{\frac{\|X_k\|}{\|X_0\|}}, & \text{(norm scaling)}, \\
\sqrt{\frac{\rho(X^{-1}_k)}{\rho(X_k)}}, & \text{(spectral scaling)}, \\
\sqrt{\frac{\det(X_k)^{-1/\alpha}}{\det(X_0)^{-1/\alpha}}}, & \text{(determinantal scaling)},
\end{cases}
\] (41)

then the accelerated forms of the proposed matrix iteration for $S$ is defined as follows:
\[
X_0 = A,
\]
\[
\zeta_0 = \text{the scaling parameter computed by (41)},
\]
\[
X_{k+1} = \zeta_k X_k \left( 12I + 200e^2 X_k^2 + 560e^4 X_k^4 + 344e^6 X_k^6 + 36\zeta_k X_k^{8} \right) \times \left[ I + 64\zeta_k X_k^2 + 406\zeta_k^2 X_k^4 + 532\zeta_k^6 X_k^6 + 145\zeta_k^{10} X_k^{10} \right]^{-1}.
\] (42)

Noting that here $I = \sqrt{-1}$.

The numerical reports are provided in Tables 1-2 for various sizes of the input matrices based on the required number of steps and the elapsed CPU times. The results uphold the analytical parts and discussions of Sections 2-3. They manifest that there is a clear improvement in the iterations’ numbers and the total elapsed CPU time by applying (17). As a matter of fact, the mean of number of iterations and the CPU times listed in the last rows of each

\[
\lim_{k \to \infty} \zeta_k = 1,
\] (43)
\[
\lim_{k \to \infty} X_k = S.
\] (44)

### 4. Experiments

Herein, several experiments are discussed for the calculation of the sign matrix. The direct application of the new formulas for finding $S$ is given below, though the application for computing the polar decomposition, finding the Yang-Baxter matrix equation can be given similarly. The simulations are run on an office laptop with Windows 7 Ultimate equipped Intel(R) Core(TM) i5-2430M CPU 3.15 2.40GHz processor and 16.00 GB of RAM on a 64-bit operating system. In addition, the simulations are done in Mathematica 11.0 [19].

Various schemes are compared in respect to the iteration numbers and the elapsed CPU time. Globally convergent schemes are only included for comparison. The compared matrix methods are NM, HM, ANM, and PM1 (i.e., (17)). We do not include comparisons with methods having local convergence behavior such as the Newton-Schulz method (10) or (computationally expensive) methods from different categories such as the ones based on the computation of the Cauchy integral
\[
\text{sign} (A) = S = \frac{2}{\pi} \int_0^\infty \left( t^2 I + A^2 \right)^{-1} dt.
\] (45)

The stopping criterion for our simulations is defined by
\[
\|X_k^2 - I\|_2 \leq 10^{-4}.
\] (46)

The reason of choosing (46) lies in the fact that, at each iterate, the obtained approximation should satisfy the main matrix equation. Thus, this criterion is much more trustful than other Cauchy-like terminations when calculating the sign of a matrix.

**Experiment 1.** Here, we calculate the sign matrix of 20 generated randomly complex matrices (with uniform distributions) via the following piece of codes in the Mathematica environment) as comes next
\begin{verbatim}
SeedRandom[123]; number = 20;
Table[A[1] = RandomComplex[{-5 - 5 I, 5 + 5 I}, {50 1, 50 1}];, {1, number}];
\end{verbatim}
Table 1: Comparison of iterations’ numbers for Experiment 1.

<table>
<thead>
<tr>
<th>Matrix No.</th>
<th>NM</th>
<th>ANM</th>
<th>HM</th>
<th>PM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{50 \times 50}$</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>$A_{100 \times 100}$</td>
<td>15</td>
<td>14</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>$A_{150 \times 150}$</td>
<td>18</td>
<td>17</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>$A_{200 \times 200}$</td>
<td>15</td>
<td>14</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>$A_{250 \times 250}$</td>
<td>18</td>
<td>18</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>$A_{300 \times 300}$</td>
<td>23</td>
<td>21</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>$A_{350 \times 350}$</td>
<td>19</td>
<td>17</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>$A_{400 \times 400}$</td>
<td>17</td>
<td>16</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>$A_{450 \times 450}$</td>
<td>16</td>
<td>16</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>$A_{500 \times 500}$</td>
<td>20</td>
<td>20</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>$A_{550 \times 550}$</td>
<td>17</td>
<td>18</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>$A_{600 \times 600}$</td>
<td>20</td>
<td>19</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>$A_{650 \times 650}$</td>
<td>21</td>
<td>21</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>$A_{700 \times 700}$</td>
<td>17</td>
<td>18</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>$A_{750 \times 750}$</td>
<td>20</td>
<td>21</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>$A_{800 \times 800}$</td>
<td>20</td>
<td>22</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>$A_{850 \times 850}$</td>
<td>20</td>
<td>21</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>$A_{900 \times 900}$</td>
<td>20</td>
<td>21</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>$A_{950 \times 950}$</td>
<td>21</td>
<td>20</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>$A_{1000 \times 1000}$</td>
<td>21</td>
<td>24</td>
<td>14</td>
<td>8</td>
</tr>
</tbody>
</table>

Mean 18.45 18.45 11.95 6.35

The results for Experiment 2 are given in Figure 4 showing a stable and consistent behavior of the proposed scheme for finding matrix sign function.

The numerical reports and evidences in Section 4 improve the mean of the CPU time, clearly. This was the main target of this paper in order to propose an efficient method.

5. Discussion

In various fields of numerical linear algebra and scientific computing, the theory and computation of matrix functions are very much useful. In the modern numerical linear algebra, it underlies an effective way introducing one to resolve the topical problems of the control theory. The function of a matrix can be defined in several ways, of which the following three are generally the most useful: Jordan canonical form, polynomial interpolation, and finally Cauchy integral.

In this research work, we have focused on iterative methods for this purpose. Hence, a high-order nonlinear equation solver has been employed for constructing a novel scheme in calculating the sign matrix, which does not have pure imaginary eigenvalues.

It was shown that the convergence is global via attraction basins in the complex plane and the rate of convergence is eight. Finally, some numerical experiments in double precision arithmetic were performed to manifest the superiority and applicability of (17). Outlines for future works can be forced to extend the discussed matrix iterations for calculating polar decompositions in future studies based on the application of the new schemes.

Data Availability

All the data used for comparison of different methods in this article have been generated using random generators, via the programming package Mathematica. The data can be generated in this way. Moreover, interested readers may contact the corresponding authors if they need any of such programming codes for further studies.
Table 2: Comparison of the elapsed (CPU) times for Experiment 1.

<table>
<thead>
<tr>
<th>Matrix No.</th>
<th>NM</th>
<th>ANM</th>
<th>HM</th>
<th>PM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{50\times50}$</td>
<td>0.148498</td>
<td>0.0230943</td>
<td>0.00861964</td>
<td>0.0102797</td>
</tr>
<tr>
<td>$A_{100\times100}$</td>
<td>0.0943384</td>
<td>0.109786</td>
<td>0.0727258</td>
<td>0.0392688</td>
</tr>
<tr>
<td>$A_{150\times150}$</td>
<td>0.232303</td>
<td>0.506882</td>
<td>0.189903</td>
<td>0.21293</td>
</tr>
<tr>
<td>$A_{200\times200}$</td>
<td>0.330597</td>
<td>0.709762</td>
<td>0.280812</td>
<td>0.21293</td>
</tr>
<tr>
<td>$A_{250\times250}$</td>
<td>0.657373</td>
<td>1.44232</td>
<td>0.555998</td>
<td>0.458002</td>
</tr>
<tr>
<td>$A_{300\times300}$</td>
<td>1.29408</td>
<td>2.56028</td>
<td>1.13594</td>
<td>0.80164</td>
</tr>
<tr>
<td>$A_{350\times350}$</td>
<td>4.8739</td>
<td>9.09064</td>
<td>4.03174</td>
<td>3.20028</td>
</tr>
<tr>
<td>$A_{400\times400}$</td>
<td>4.8925</td>
<td>10.7225</td>
<td>4.83132</td>
<td>3.62766</td>
</tr>
<tr>
<td>$A_{450\times450}$</td>
<td>7.07534</td>
<td>14.5142</td>
<td>7.22801</td>
<td>4.66437</td>
</tr>
<tr>
<td>$A_{500\times500}$</td>
<td>9.72633</td>
<td>19.8393</td>
<td>8.51814</td>
<td>6.87333</td>
</tr>
<tr>
<td>$A_{550\times550}$</td>
<td>9.34249</td>
<td>21.6787</td>
<td>8.03064</td>
<td>7.05715</td>
</tr>
<tr>
<td>$A_{600\times600}$</td>
<td>13.9398</td>
<td>31.1319</td>
<td>11.5503</td>
<td>8.65203</td>
</tr>
<tr>
<td>$A_{650\times650}$</td>
<td>16.7687</td>
<td>39.0702</td>
<td>14.9441</td>
<td>10.4224</td>
</tr>
<tr>
<td>$A_{700\times700}$</td>
<td>19.9251</td>
<td>45.4667</td>
<td>18.0246</td>
<td>18.9427</td>
</tr>
<tr>
<td>$A_{750\times750}$</td>
<td>24.1888</td>
<td>52.8144</td>
<td>20.6969</td>
<td>19.2772</td>
</tr>
<tr>
<td>$A_{800\times800}$</td>
<td>28.187</td>
<td>58.5243</td>
<td>24.4635</td>
<td>17.3038</td>
</tr>
<tr>
<td>$A_{850\times850}$</td>
<td>34.3487</td>
<td>82.3412</td>
<td>29.5465</td>
<td>26.1742</td>
</tr>
<tr>
<td>Mean</td>
<td>9.09109</td>
<td>20.1257</td>
<td>7.98652</td>
<td>6.67063</td>
</tr>
</tbody>
</table>

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research project was supported by a grant from the “Research Center of the Female Scientific and Medical Colleges”, Deanship of Scientific Research, King Saud University.

References


