KRYLOV SUBSPACE APPROXIMATIONS TO THE TOEPLITZ MATRIX EXPONENTIAL *

SPIKE T. LEE †

Abstract. The shift-invert Arnoldi method, which belongs to the family of Krylov subspace methods, is employed to generate an orthonormal basis from the Krylov subspace corresponding to a nonsymmetric Toeplitz matrix and an initial vector. The vectors and recurrence coefficients produced by this method are exploited to approximate the Toeplitz matrix exponential. Toeplitz matrix inversion formula and rapid Toeplitz matrix-vector multiplications are used to lower the computational costs. For error analysis, a sufficient condition is established to guarantee that the error bound is independent of the norm of the matrix. Numerical results are given to demonstrate the efficiency of the method.

Key words. Toeplitz matrix, matrix exponential, Krylov subspace, shift-invert Arnoldi method, numerical range

AMS subject classifications. 65L05; 65N22; 65F10; 65F15

1. Introduction. An $n \times n$ *Toeplitz* matrix A_n is defined as $[A_n]_{j,k} = a_{j-k}$, which means A_n has constant diagonals. We consider the approximation to the *Toeplitz matrix exponential* (TME):

$$w(t) = \exp(-tA_n)v, \tag{1.1}$$

where A_n is a real nonsymmetric Toeplitz matrix, v is a given vector, and $t > 0$ is a scaling factor.

Toeplitz matrices arise from topics like signal and image processing, queueing networks, and numerical solutions of partial differential equations and integral equations, see [2] and the references therein. Methods for solving Toeplitz systems have been under thorough study over the last twenty years. Apart from Toeplitz system solvers, the TME plays a key role in various application fields. In computational finance, Toeplitz matrices can be seen from the option pricing framework in jump-diffusion models, where a partial integro-differential equation (PIDE) needs to be solved. Tangman et al. [15] reduced the problem to the approximation of a real nonsymmetric TME. In integral equations, the TME also takes part in the numerical solution of Volterra-Wiener-Hopf equations [1].

However, Toeplitz matrices are generally dense. Therefore the classic methods in [12] for approximating the exponential of Toeplitz matrix will suffer from $\mathcal{O}(n^3)$ complexity. On the other hand, the Krylov subspace methods have recently become an efficient means to approximate the matrix exponential multiplied by a vector [3, 4, 5, 6, 9, 10, 13, 14], especially when the matrix is very large and sparse. The computational cost can be brought down to $\mathcal{O}(n)$ in some cases. The primary objective of these methods is to construct an orthonormal basis from a Krylov subspace with regard to a certain matrix. This is achieved by the Lanczos process for symmetric matrices or the Arnoldi process for nonsymmetric matrices, while both processes require only matrix-vector multiplications. Once the basis is constructed, preferably at fewer costs, all that is left to do is to approximate a comparatively smaller matrix exponential. In particular, Moret and Novati [13] improved the Arnoldi method with a shift-invert technique, which allowed them to speed up the Arnoldi process. They also presented an error estimation in terms of the numerical range of a matrix. In [4], van den Eshof and Hochbruck also applied a similar idea to revise the Lanczos process for symmetric matrices, though from a different point of view. The brilliant performance of such kind of modified Krylov subspace methods arouses our interest, and what is more, we recall that matrix-vector products are included during the process. It is well known that Toeplitz matrices possess great structures and properties, and their matrix-vector multiplications can be computed by the fast Fourier transform (FFT) with $\mathcal{O}(n \log n)$ complexity [2]. For this reason we expect that the operation cost of TME should be less than $\mathcal{O}(n^3)$.

In this paper, we propose an algorithm to approximate the TME (1.1). To our knowledge, the approximation to TME has never been studied before. Our scheme resembles the one in [13], i.e., to

*The research was partially supported by the research grant 033/2009/A from FDCT of Macao, UL020/08-Y2/MAT/JXQ01/FST and RG063/08-09S/SHW/FST from University of Macau.

† Department of Mathematics, University of Macau, Macao, China (ma76522@umac.mo).

adjust the Arnoldi process for better productivity. Meanwhile, the transformed formulation requires the inverse of the Toeplitz matrix. By making use of the Toeplitz structure and the famous Gohberg-Semencul formula (GSF) in [7], we can reduce the computational cost to $\mathcal{O}(n \log n)$ in total. As in [13], we will establish a sufficient condition for an error bound which is independent of $\|tA_n\|_2$, but in Toeplitz fashion instead. As an application, a TME which stems from a PIDE is considered. Numerical results will illustrate the efficiency and robustness of our method. The rest of the paper is arranged as follows. In Section 2 we introduce the background of Toeplitz matrices. In Section 3 we illustrate the shift-invert Arnoldi method for the approximation to matrix exponential multiplying a vector. Implementation and error estimation of the shift-invert Arnoldi method for Toeplitz matrices are presented in Section 4. In Section 5 we report the numerical results. At last we give the concluding remarks in Section 6.

2. Background of Toeplitz matrices. As a special case of Toeplitz matrix, an $n \times n$ matrix C_n is called *circulant* if it is a Toeplitz matrix $[C_n]_{j,k} = c_{j-k}$ with $c_l = c_{l-n}$ for $l = 1, 2, \ldots, n-1$. Moreover, a circulant matrix can be diagonalized by the Fourier matrix F_n, i.e.,

$$C_n = F_n^* \Lambda_n F_n, \tag{2.1}$$

where the entries of F_n are given by

$$[F_n]_{j,k} = \frac{1}{\sqrt{n}} e^{2\pi i jk/n}, \qquad i \equiv \sqrt{-1}, \qquad 0 \le j, k \le n-1,$$

and Λ_n is a diagonal matrix holding the eigenvalues of C_n.

From (2.1), we can determine Λ_n in $\mathcal{O}(n \log n)$ operations by taking only one n-length FFT of the first column of C_n [2]. Furthermore, we can consider the computation of a circulant matrix multiplying a vector. Suppose u is the given vector. The multiplication $C_n u$ or $C_n^{-1} u$ is then computed by a couple of FFTs in $\mathcal{O}(n \log n)$ operations provided that Λ_n is already obtained. If the Toeplitz matrix-vector product $A_n u$ is wanted, we can first embed A_n into a $2n \times 2n$ circulant matrix, i.e.,

$$\begin{bmatrix} A_n & \times \\ \times & A_n \end{bmatrix} \begin{bmatrix} u \\ 0 \end{bmatrix} = \begin{bmatrix} A_n u \\ \dagger \end{bmatrix}. \tag{2.2}$$

Now that we are back to the circulant case, the multiplication is carried out as discussed before, with $\mathcal{O}(n \log n)$ complexity [2].

2.1. Generating functions. It is common to assume that the diagonals $\{a_k\}_{k=-n+1}^{n-1}$ of a Toeplitz matrix A_n are the Fourier coefficients of a function f:

$$a_k = a_k(f) \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-ik\theta} d\theta, \quad k = -n+1, \ldots, n-1.$$

Then the function f is known as the *generating function* of A_n.

In general, f is a complex-valued function. We symbolize the real part and imaginary part of f by $Re(f)$ and $Im(f)$ respectively. Let $\mathcal{T}_n[f]$ denote a Toeplitz matrix generated by f. One can easily check that a Toeplitz matrix $\mathcal{T}_n[f]$ is a real matrix when $Re(f)$ is an even function and $Im(f)$ is an odd function. We remark that $\mathcal{T}_n[f]$ is reduced to a nonsymmetric Toeplitz matrix in $\mathbb{R}^{n \times n}$ when $Im(f) \neq 0$.

Throughout this paper we consider a Toeplitz matrix in $\mathbb{R}^{n \times n}$. It is assumed that its generating function $f \in \mathcal{C}_{2\pi}$, where $\mathcal{C}_{2\pi}$ contains all the 2π-periodic continuous complex-valued functions. Furthermore, we define

$$\|f\|_\infty = \max_{\theta \in [-\pi, \pi]} |f(\theta)|,$$

and let f satisfy the assumptions below:

$$Re(f) \ge 0 \quad \text{and} \quad \|Im(f)/Re(f)\|_\infty = \mathcal{O}(1). \tag{2.3}$$

The two assumptions in (2.3) will be used later for error estimation.

2.2. Gohberg-Semencul formula. The Toeplitz matrix inversion is also studied thoroughly besides Toeplitz matrix-vector multiplication [8]. In [7], Gohberg and Semencul discovered the GSF for the inverse of a Toeplitz matrix A_n. The formula shows that the inverse A_n^{-1} can be *explicitly* represented by its first column $x = [x_1, x_2, \ldots, x_n]^\intercal$ and last column $y = [y_1, y_2, \ldots, y_n]^\intercal$ provided that $x_1 \neq 0$. The GSF is given by:

$$A_n^{-1} = \frac{1}{x_1}\left(X_n Y_n^\intercal - \hat{Y}_n \hat{X}_n^\intercal\right), \tag{2.4}$$

where X_n, Y_n, $\hat{X}_n$ and $\hat{Y}_n$ are all lower triangular Toeplitz matrices given by

$$X_n = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ x_2 & x_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n-1} & \cdots & x_1 \end{bmatrix}, \quad Y_n = \begin{bmatrix} y_n & 0 & \cdots & 0 \\ y_{n-1} & y_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ y_1 & y_2 & \cdots & y_n \end{bmatrix},$$

$$\hat{X}_n = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ x_n & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ x_2 & \cdots & x_n & 0 \end{bmatrix} \quad \text{and} \quad \hat{Y}_n = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ y_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ y_{n-1} & \cdots & y_1 & 0 \end{bmatrix}.$$

Note that x and y can also be regarded as the solutions of two linear systems:

$$A_n x = e_1 \quad \text{and} \quad A_n y = e_n, \tag{2.5}$$

where e_1 and e_n are the first and last column of an identity matrix. By the GSF (2.4), the inverse of A_n is decided through the solvability of two linear systems in (2.5) which both hold the Toeplitz matrix A_n as the coefficient matrix, and also the condition $x_1 \neq 0$.

Particularly if we want the solutions of many Toeplitz systems which share the same coefficient matrix A_n, we only need to solve two Toeplitz systems (2.5) to obtain the first and last column of A_n^{-1}. Then the GSF (2.4) yields an explicit representation of A_n^{-1} in terms of four triangular Toeplitz matrices, and all the desired solutions are derived via Toeplitz matrix-vector multiplications (2.2) in $\mathcal{O}(n \log n)$ operations instead of solving many Toeplitz systems.

The GSF (2.4) gives an exact representation of the inverse of a Toeplitz matrix, but instead we have to solve two Toeplitz systems in (2.5). In this paper, we prefer the iterative methods with complexity $\mathcal{O}(n \log n)$ over the direct methods with complexity $\mathcal{O}(n \log^2 n)$ [2]. For example, one can choose the conjugate gradient normal equation method with T. Chan's circulant preconditioner for solving nonsymmetric Toeplitz systems like (2.5). It is well known that T. Chan's circulant preconditioner suits a wide class of Toeplitz matrices, see [2] for more discussions. Alternatively, the GMRES method with T. Chan's circulant preconditioner is another choice for solving (2.5). In many applications, the GMRES method may converge very fast.

3. Krylov subspace methods for matrix exponential. Krylov subspace methods for computing $w(t) = \exp(-tA_n)v$ are widely investigated over the years [3, 4, 5, 6, 9, 10, 13, 14]. The main concept of such methods is to approximately project the exponential of a large matrix onto an m-th dimension Krylov subspace

$$\mathcal{K}_m \equiv \text{span}\{v, A_n v, \ldots, A_n^{m-1} v\}.$$

Meanwhile, the renowned Arnoldi process is used to generate an orthonormal basis of $\mathcal{K}_m$, see [14] for details. However, it is shown in [9] that the number of steps m of the Krylov subspace approximation is close to $\mathcal{O}(\|tA_n\|_2)$. That means standard methods like Arnoldi method or Lanczos method could be unsatisfactory if $\|tA_n\|_2$ is large. To untangle this knot, one can exploit a potential advantage of Krylov subspace methods, i.e., they incline to locate well-separated eigenvalues faster [4]. For instance, Moret and Novati [13] put this advantage into practice by filling in a shift-invert technique. Such maneuver can also be found in numerical methods for eigenvalue problems.

3.1. Shift-invert Arnoldi method. Let I be the identity matrix. The shift-invert technique is to apply the Arnoldi process to a shifted and inverted matrix $(I+\gamma A_n)^{-1}$, which stresses the required eigenvalues, with a shift parameter $\gamma > 0$.

Algorithm 1: Arnoldi process with shift-invert technique

1. Initialize: Compute $v_1 = v/\|v\|_2$
2. Iterate: Do $j = 1, \ldots, m$
   (a) Compute $u := (I+\gamma A_n)^{-1} v_j$
   (b) Do $k = 1, \ldots, j$
      i. Compute $h_{k,j} := (u, v_k)$
      ii. Compute $u := u - h_{k,j} v_k$
   (c) Compute $h_{j+1,j} := \|u\|_2$ and $v_{j+1} := u/h_{j+1,j}$

By using $v_1 = v/\|v\|_2$ as an initial vector, we reach the following formulation:

$$(I+\gamma A_n)^{-1} V_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^{\intercal}, \tag{3.1}$$

where $V_m = [v_1, \ldots, v_m]$ is the resulting $n \times m$ matrix containing the orthonormal basis, H_m is an $m \times m$ upper Hessenberg matrix, and e_j denotes the j-th column of the identity matrix. More specifically, the matrix $(I+\gamma A_n)^{-1}$ is projected onto a Krylov subspace. Then this formulation creates an approximation for $\exp(-tA_n)v$ [13]:

$$\exp(-tA_n)v \approx \beta V_m \exp(-\tau(H_m^{-1} - I))e_1 \equiv \beta V_m g(H_m) e_1, \tag{3.2}$$

where $\tau = t/\gamma$, $\beta = \|v\|_2$, and $g(z) = \exp(-\tau(z^{-1}-1))$.

The approximation (3.2) shows that the large matrix exponential $\exp(-tA_n)$ is replaced by a small matrix exponential of size $m \times m$. Let $w_m(t)$ denote the approximation in (3.2):

$$w_m(t) = \beta V_m g(H_m) e_1. \tag{3.3}$$

Then this algorithm is called the *shift-invert Arnoldi method*. Here we remark that this treatment has also been presented independently by [4] for symmetric matrices.

3.2. Error estimation of shift-invert Arnoldi approximation. Moret and Novati [13] estimated the error between the approximation $w_m(t)$ and the vector $w(t) = \exp(-tA_n)v$ from the viewpoint of sectorial operator and numerical range. For convenience, we simply assume $\beta = \|v\|_2 = 1$ in this part. Following [13], we first introduce several relevant concepts.

The numerical range of a matrix A_n is defined as a subset in the complex plane $\mathbb{C}$:

$$W(A_n) \equiv \{u^* A_n u,\ u \in \mathbb{C}^n,\ u^* u = 1\}.$$

Then we let $\Sigma_{\alpha,\vartheta}$ be the following set:

$$\Sigma_{\alpha,\vartheta} = \{z \in \mathbb{C} :\ |\arg(z-\alpha)| < \vartheta,\ \alpha \geq 0,\ 0 < \vartheta < \frac{\pi}{2}\},$$

i.e., $\Sigma_{\alpha,\vartheta}$ is an unbounded sector in the right-half plane with semi-angle $\vartheta < \pi/2$ and vertex lying on the non-negative real axis. Note that $\Sigma_{\alpha,\vartheta}$ is symmetric about the real-axis. The matrix A_n is called a *sectorial operator* [13] if A_n satisfies:

$$W(A_n) \subseteq \Sigma_{\alpha,\vartheta}.$$

In [13], Proposition 2.1 and Proposition 3.2 provide a sufficient condition for error estimate of shift-invert Arnoldi approximation in terms of sectorial operator. For convenience, we summarize them in the following theorem.

THEOREM 3.1. ([13], Proposition 2.1 and Proposition 3.2) *Let A_n be a sectorial operator. Then the shift-invert Arnoldi approximation $w_m(t)$ in (3.3) to the matrix exponential $w(t) = \exp(-tA_n)v$ satisfies*

$$\|w(t) - w_m(t)\| \leq \frac{\Phi_m}{\pi \sin(\vartheta^* - \vartheta)},$$

where $\vartheta < \vartheta^* < \pi/2$, the term $\Phi_m \to 0$ as $m \to \infty$ and its convergence is independent of $\|tA_n\|_2$.

From Theorem 3.1, we first derive that

$$w_m(t) \to w(t), \quad m \to \infty.$$

Theorem 3.1 also hints that $\|tA_n\|_2$ is not connected with the error bound, which has been a bottleneck for standard Arnoldi method. In addition, we can also see that a ϑ not close to $\pi/2$ is more favored. It is because a ϑ close to $\pi/2$ leads to a small $\sin(\vartheta^* - \vartheta)$, and hence turns the factor $(\pi \sin(\vartheta^* - \vartheta))^{-1}$ towards infinity. Particularly when $\vartheta = 0$, it is reduced to the symmetric case, and related error estimations can be found in [4].

In conclusion, if A_n is a sectorial operator, then Theorem 3.1 shows that the shift-invert Arnoldi approximation $w_m(t)$ converges to the matrix exponential $w(t)$, and the error bound does not depend on $\|tA_n\|_2$, see [13] for more details.

4. Implementation and error estimation. In this section, we go into further details of approximating the real TME by shift-invert Arnoldi method. Recall the assumptions that the real Toeplitz matrix $A_n = \mathcal{T}_n[f]$ is generated by $f \in \mathcal{C}_{2\pi}$, and f satisfies the assumptions in (2.3):

$$Re(f) \geq 0 \quad \text{and} \quad \|Im(f)/Re(f)\|_\infty = \mathcal{O}(1).$$

We remark that $Re(f) \geq 0$ is not a necessary condition since the other possible cases can be easily handled by shifting the matrix and multiplying a certain factor [4, 13]. We first clarify how the shift-invert Arnoldi method practically works, and then investigate the error estimation by using generating functions.

4.1. Implementation of shift-invert Arnoldi method. In the standard Arnoldi algorithm, each matrix-vector product $A_n v_j$ for $j = 1, \ldots, m$ is evaluated at each iteration step. If A_n is a Toeplitz matrix, then these multiplications can be carried out by (2.2) in $\mathcal{O}(n \log n)$ operations. Once we have included the shift-invert technique, the required matrix-vector multiplication becomes $(I + \gamma A_n)^{-1} v_j$ at each iteration step. Suppose the inverse of $I + \gamma A_n$ can be found beforehand. Then all $(I + \gamma A_n)^{-1} v_j$ are obtained by matrix-vector products instead of solving systems. Since A_n is a real nonsymmetric Toeplitz matrix, the shifted matrix $I + \gamma A_n$ is also a real nonsymmetric Toeplitz matrix for a fixed γ. Recall that the GSF (2.4) provides an explicit representation of the inverse of a Toeplitz matrix, therefore this celebrated formula would come in handy in our case.

We first gather the shift-invert Arnoldi method for real TME as the algorithm below:

Algorithm 2: shift-invert Arnoldi method for real TME

1.	Solve $(I + \gamma A_n)x = e_1$ and $(I + \gamma A_n)y = e_n$ by fast Toeplitz solvers
2.	Perform the Arnoldi process in which each multiplication $(I + \gamma A_n)^{-1} v_j$ is calculated through (2.4) by FFTs
3.	Evaluate the approximation $w_m(t) = \beta V_m g(H_m) e_1$

In order to apply the GSF (2.4), we first need to verify that $x_1 \neq 0$. Since $I + \gamma A_n$ is real and nonsingular, x should be a real vector not equal to zero. We first have the following relations:

$$(I + \gamma A_n)x = e_1 \quad \text{and} \quad x^\intercal (I + \gamma A_n)x = x_1.$$

Note that $\gamma > 0$ and $Re(f) \geq 0$, it follows that:

$$x_1 = x^\intercal x + \gamma x^\intercal \mathcal{T}_n(Re(f))x > 0,$$

i.e., x_1 is not equal to zero, and hence the GSF (2.4) is feasible.

To seek the inverse of the Toeplitz matrix $I + \gamma A_n$, we have to start with finding the first and last column x and y of $(I + \gamma A_n)^{-1}$ by solving the following two systems by fast Toeplitz solvers:

$$(I + \gamma A_n)x = e_1 \quad \text{and} \quad (I + \gamma A_n)y = e_n, \tag{4.1}$$

where the coefficient matrix $I + \gamma A_n$ is real, nonsymmetric, and Toeplitz. See Section 2.2 for details.

After collecting the two columns x, y and making sure $x_1 \neq 0$, we can compute all the matrix-vector products $(I + \gamma A_n)^{-1} v_j$ exactly through FFTs in $\mathcal{O}(n \log n)$ operations [2]. After performing the shift-invert Arnoldi process, we derive the matrix formulation (3.1). Finally, the resulting small matrix exponential in (3.3) can be evaluated by the scaling and squaring method or other classic methods [12] provided that m is small enough.

Note that step 1 and step 2 carry most of the workloads in the shift-invert Arnoldi algorithm for general matrices. For a Toeplitz matrix, we manage to reduce them all to $\mathcal{O}(n \log n)$ operations by making use of the Toeplitz properties. Thus the computational cost of the shift-invert Arnoldi method for approximating the real TME is of $\mathcal{O}(n \log n)$.

4.2. Error estimation by generating functions. In [13], Moret and Novati diagnosed the error bound of the shift-invert Arnoldi approximation. The premise is that A_n is a sectorial operator. Therefore, the next thing we do is to sort out such characteristics of Toeplitz matrices by using their generating functions.

We first introduce the *convex hull* of a given set U, which is the intersection of all the convex sets containing U, and denoted by $\mathrm{conv}(U)$. Suppose $A_n = \mathcal{T}_n[f]$. Let

$$\Omega(f) \equiv \{f(\theta),\ \forall \theta \in [-\pi, \pi]\}$$

be the range of the generating function f. There exists a relation between the numerical range and the generating function of a Toeplitz matrix according to Theorem 5.1 in [16]. In this paper, we only need a special case of Theorem 5.1 in [16]. For brevity, it is summarized as the following theorem:

THEOREM 4.1. ([16], Theorem 5.1) *Let $W(A_n)$ be the numerical range of $A_n = \mathcal{T}_n[f]$, where $f \in \mathcal{C}_{2\pi}$. Then $W(A_n)$ is a subset of the closure of* $\mathrm{conv}(\Omega(f))$*, i.e.,*

$$W(A_n) \subseteq \overline{\mathrm{conv}(\Omega(f))}.$$

To show that a Toeplitz matrix A_n is a sectorial operator, we need the following lemma.

LEMMA 4.2. *Let $f \in \mathcal{C}_{2\pi}$ and $\vartheta = \arctan \|Im(f)/Re(f)\|_\infty$. If f satisfies the two assumptions in* (2.3)*, then we have $\vartheta < \pi/2$ and*

$$\overline{\mathrm{conv}(\Omega(f))} \subseteq \overline{\Sigma_{0,\vartheta}}.$$

Proof. It is known from the assumption that $\|Im(f)/Re(f)\|_\infty \leq M < \infty$. Thus

$$\vartheta = \arctan \|Im(f)/Re(f)\|_\infty \leq \arctan M < \pi/2.$$

Since $Re(f) \geq 0$, we have for any $z = f(\theta) \in \Omega(f)$ that

$$|\arg z| = \arctan |Im(f(\theta))/Re(f(\theta))| \leq \arctan \|Im(f)/Re(f)\|_\infty = \vartheta.$$

It follows that $z \in \overline{\Sigma_{0,\vartheta}}$, which implies

$$\Omega(f) \subseteq \overline{\Sigma_{0,\vartheta}}.$$

Apparently $\overline{\Sigma_{0,\vartheta}}$ is a closed convex set, the proof is completed. □

THEOREM 4.3. *Suppose* $A_n = \mathcal{T}_n[f] \in \mathbb{R}^{n\times n}$ *with* $f \in \mathcal{C}_{2\pi}$. *If* f *satisfies the two assumptions in* (2.3)*, then* A_n *is a sectorial operator.*

Proof. By combining Lemma 4.2 and Theorem 4.1, we simply have $\vartheta < \pi/2$ and

$$W(A_n) \subseteq \overline{\Sigma_{0,\vartheta}},$$

i.e., A_n is a sectorial operator. The proof is completed. □

Theorem 4.3 gives a sufficient condition of whether a real Toeplitz matrix A_n is a sectorial operator, in terms of its generating function f. Recall that Theorem 3.1 guarantees the absence of $\|tA_n\|_2$ in the error bound when A_n is a sectorial operator. In conclusion, if the generating function f of A_n satisfies the condition of Theorem 4.3, then the error bound of the shift-invert Arnoldi approximation does not depend on $\|tA_n\|_2$. In the next section, we will verify this conclusion by numerical experiments.

5. Numerical results. In the following numerical tests, we consider approximating the real nonsymmetric TME (1.1), namely

$$w(t) = \exp(-tA_n)v,$$

by shift-invert Arnoldi method and standard Arnoldi method [14]. All experiments are conducted in MATLAB. We regard the MATLAB command `expm` as the exact value for $w(t)$. For all tables, "n" denotes the matrix size, "*tol*" stands for the tolerance of $\|w(t) - w_m(t)\|_2/\|w(t)\|_2 < tol$, where $w_m(t)$ is the numerical approximation to $w(t)$. The column "shft-inv" displays the iteration numbers of shift-invert Arnoldi method, while "stdrd" shows the ones of standard Arnoldi method. For any "-" showing up in the column, it means the number of iterations exceeds 250.

In the implementation of shift-invert Arnoldi method, the choice of γ is an intricate issue. In the symmetric matrix case, there are detailed studies on the optimal choice of γ [4], though an exact value $\gamma = t/10$ is used throughout the numerical experiments therein. However, for the general case, how to pick an appropriate γ remains a puzzle. In [13], which studies the nonsymmetric matrix case, the parameter γ is chosen as t divided by a number varying in the range $[1, 10]$. In fact, numerical experiences show that γ is not sensitive to the behavior of iteration numbers, hence it is similarly selected as $\gamma = t/10$ in our case.

Three examples are given to demonstrate the shift-invert Arnoldi method and standard Arnoldi method. The first two examples mainly explain how the assumption (2.3) makes a difference, where the vector v is chosen to be the vector of all ones. The third example is an application in computational finance, in which a real nonsymmetric TME is involved.

Example 1. We consider a Toeplitz matrix A_n which is generated by the function

$$f(\theta) = \theta^2 + i \cdot \theta^3, \qquad \theta \in [-\pi, \pi].$$

Note that $Re(f) = \theta^2 \geq 0$ is an even function and $Im(f) = \theta^3$ is an odd function. According to Section 2.1, A_n is a real Toeplitz matrix. It is obvious that the generating function satisfies $\|Im(f)/Re(f)\|_\infty = \mathcal{O}(1)$ and leads to a semi-angle

$$\vartheta = \arctan \|\theta^3/\theta^2\|_\infty = \arctan \pi < \pi/2.$$

By Theorem 4.3 and Theorem 3.1, the error bound of the shift-invert Arnoldi method should be independent of $\|tA_n\|_2$. Note that $\|A_n\|_2$ does not depend on the matrix size n in this example, hence we set $n = 512$ and try out different values of t. Numerical results in Table 5.1 show that the iteration numbers of shift-invert Arnoldi method are indeed independent of $\|tA_n\|_2$, or t in this case. Oppositely, the standard Arnoldi method needs more iterations as t increases.

Example 2. We consider a Toeplitz matrix A_n which is generated by the function

$$f(\theta) = \theta^2 + i \cdot \mathrm{sgn}(\theta), \qquad \theta \in [-\pi, \pi],$$

where $\mathrm{sgn}(\theta)$ is the sign function defined as

$$\mathrm{sgn}(\theta) = \begin{cases} 1, & 0 < \theta \leq \pi, \\ 0, & \theta = 0, \\ -1, & -\pi \leq \theta < 0. \end{cases}$$

Note that $Re(f) = \theta^2 \geq 0$ is an even function and $Im(f) = \mathrm{sgn}(\theta)$ is an odd function. According to Section 2.1, A_n is a real Toeplitz matrix. It is easy to see that the quotient $|\mathrm{sgn}(\theta)/\theta^2|$ is unbounded when $\theta \to 0$. Therefore f does not satisfy the condition (2.3).

As in Example 1, $\|A_n\|_2$ does not rely on n, hence the matrix size is fixed at $n = 512$ and different values of t should be put to the test. In Table 5.1, we see that the shift-invert Arnoldi method is inferior to the previous example and the number of iterations gradually increases in accordance with $\|tA_n\|_2$, or simply t in this example. It is due to the incapability of meeting the error bound condition $\|Im(f)/Re(f)\|_\infty = \mathcal{O}(1)$ in Theorem 4.3. For the standard Arnoldi method, the iteration numbers still fail to stay steady, just like their counterparts in Example 1.

TABLE 5.1
The numbers of iterations of shift-invert Arnoldi method and standard Arnoldi method in Example 1 and Example 2.

t	Example 1				Example 2			
	$tol = 10^{-4}$		$tol = 10^{-7}$		$tol = 10^{-4}$		$tol = 10^{-7}$	
	shft-inv	stdrd	shft-inv	stdrd	shift-inv	stdrd	shift-inv	stdrd
1	11	31	31	41	7	10	11	15
10	10	147	22	183	18	43	28	54
100	9	-	18	-	59	148	84	193
1000	9	-	16	-	-	-	-	-

Example 3. We consider pricing options for a single underlying asset in Merton's jump-diffusion model [11] as an application of the shift-invert Arnoldi method. In Merton's model, jumps are normally distributed with mean μ and variation σ. The option value $\omega(\xi, t)$ with logarithmic price ξ and backward time t satisfies a forward PIDE on $(-\infty, +\infty) \times [0, T]$:

$$\omega_t = \frac{\nu^2}{2}\omega_{\xi\xi} + (r - \lambda\kappa - \frac{\nu^2}{2})\omega_\xi - (r + \lambda)\omega + \lambda \int_{-\infty}^{\infty} \omega(\xi + \eta, t)\phi(\eta)d\eta, \tag{5.1}$$

where T is the maturity time, ν is the stock return volatility, r is the risk-free interest rate, λ is the arrival intensity of a Poisson process, $\kappa = e^{(\mu+\sigma^2/2)} - 1$ is the expectation of the impulse function and ϕ is the Gaussian distribution given by

$$\phi(\eta) = \frac{e^{-(\eta-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}. \tag{5.2}$$

For a European call option, the initial condition is

$$\omega(\xi, 0) = \max(Ke^\xi - K, 0), \tag{5.3}$$

where K is the strike price [11]. We first truncate the infinite ξ-domain $(-\infty, \infty)$ to $[\xi_{\min}, \xi_{\max}]$, and then divide $[\xi_{\min}, \xi_{\max}]$ into $n+1$ subintervals with a uniform mesh size Δ_ξ. By approximating the differential part of (5.1) by central difference discretization, we obtain an $n \times n$ tridiagonal Toeplitz matrix

$$\mathcal{D}_n = \mathrm{tridiag}\left[\frac{\nu^2}{2\Delta_\xi^2} - \frac{2r - 2\lambda\kappa - \nu^2}{4\Delta_\xi},\ -\frac{\nu^2}{\Delta_\xi^2} - r - \lambda,\ \frac{\nu^2}{2\Delta_\xi^2} + \frac{2r - 2\lambda\kappa - \nu^2}{4\Delta_\xi}\right].$$

For the integral term in (5.1), the localized part can be expressed in discrete form by using the rectangle rule. The corresponding operator is an $n \times n$ Toeplitz matrix

$$[\mathcal{I}_n]_{j,k} = \Delta_\xi \cdot \phi((k - j)\Delta_\xi).$$

Let $A_n = \mathcal{D}_n + \lambda\mathcal{I}_n$ be the real nonsymmetric Toeplitz matrix. Then A_n is the coefficient matrix of the semi-discretized system with regard to t [15]. The option price at $t = T$ requires evaluating the exponential term $\exp(TA_n)\omega_0$, where ω_0 is the discretized form of the initial value in (5.3), see [15] for details. By simple calculations, we can show that the assumption (2.3) is also satisfied in this example under the condition $r > 0$.

The input parameters are $\xi_{\min} = -2$, $\xi_{\max} = 2$, $K = 100$, $\nu = 0.25$, $r = 0.05$, $\lambda = 0.1$, $\mu = -0.9$ and $\sigma = 0.45$. The shift parameter is selected as $\gamma = T/10$ just like before. Note that $\|A_n\|_2$ increases with n as we refine grid nodes in the spatial direction. Therefore we use various matrix size n to tell the effectiveness of shift-invert Arnoldi method. In Table 5.2, numerical results show that the shift-invert Arnoldi method outperforms the standard one and the error bound is independent of $\|A_n\|_2$.

TABLE 5.2
The numbers of iterations of shift-invert Arnoldi method and standard Arnoldi method in Example 3.

n	$T = 0.5$				$T = 1$			
	$tol = 10^{-4}$		$tol = 10^{-7}$		$tol = 10^{-4}$		$tol = 10^{-7}$	
	shft-inv	stdrd	shft-inv	stdrd	shft-inv	stdrd	shft-inv	stdrd
256	9	44	17	62	10	65	17	88
512	10	88	17	122	10	128	18	175
1024	10	174	17	242	10	-	18	-
2048	10	-	17	-	10	-	18	-

Apart from showing the iteration behavior of the two methods, we now continue to cover another numerical aspect of approximating the matrix exponential. In the implementation of shift-invert Arnoldi method, it is common to first find the inverse $(I + \gamma A_n)^{-1}$ before going into the iterative process. For instance, an LU decomposition would be a natural choice for a general matrix [13]. Then every term $(I + \gamma A_n)^{-1}v_j$ is computed in each iteration by solving triangular systems. However, the factorization of a dense matrix costs $\mathcal{O}(n^3)$ operations, and also those triangular systems require $\mathcal{O}(n^2)$ operations to solve. For the TME case, the GSF (2.4) takes the upper hand in finding the inverse of a Toeplitz matrix, and it only needs to be done for once and for all. The two Toeplitz systems in (4.1) can be solved by the GMRES method with T. Chan's preconditioner. In the iterative process, $(I + \gamma A_n)^{-1}v_j$ is computed by FFTs with $\mathcal{O}(n \log n)$ complexity.

TABLE 5.3
CPU times (in seconds) of standard Arnoldi method, shift-invert Arnoldi method with GSF or LU decomposition in Example 3 with $T = 1$.

n	$tol = 10^{-4}$			$tol = 10^{-7}$		
	stdrd	shft-inv		stdrd	shft-inv	
		LU	GSF		LU	GSF
256	0.0238	0.0191	0.0126	0.0380	0.0265	0.0182
512	0.0970	0.0804	0.0187	0.1718	0.0985	0.0263
1024	0.6446	0.4258	0.0331	3.0102	0.4932	0.0466
2048	14.4769	1.9983	0.0821	33.8326	2.2678	0.1016

Table 5.3 contains the numerical results of Example 3 with $T = 1$. This time we report the CPU times (in seconds) of standard Arnoldi method, and shift-invert Arnoldi method with GSF or LU decomposition to reach the final accuracy of 10^{-4} and 10^{-7}. It is easy to see that the shift-invert Arnoldi method with GSF is less time-consuming. The standard Arnoldi method is plagued by the heavy iteration numbers and happens to be the worst among them, even worse than the costly shift-invert Arnoldi method with LU decomposition. The difference in CPU times is more obvious when the matrix size n grows larger.

6. Concluding remarks. In this paper, we have employed the shift-invert Arnoldi method to compute the real nonsymmetric TME. We show that under the two assumptions in (2.3), the real Toeplitz

matrix A_n is a sectorial operator, and hence the error bound of the shift-invert Arnoldi approximation is independent of $\|tA_n\|_2$. Moreover, we have reduced the computational costs to $\mathcal{O}(n \log n)$ by exploiting the Toeplitz structure. Several numerical examples, including an application in computational finance, illustrate that the shift-invert Arnoldi method needs far fewer iterations and is unaffected by the change of $\|tA_n\|_2$ when (2.3) is satisfied.

Finally we remark that if A_n is not an exact Toeplitz matrix, e.g., it is a block Toeplitz Toeplitz block matrix in the two-dimensional case, there will not be any efficient inversion formula just like the GSF for standard Toeplitz matrices. Accordingly in the shift-invert Arnoldi process, it is inevitable to solve a Toeplitz-like system at each iteration step. In future work, iterative methods would be studied for solving such Toeplitz-like systems, instead of using direct representation from the matrix inversion formula.

Acknowledgements. The author would like to thank Raymond H. Chan and Tao Wu for the inspiration of this topic, Igor Moret for his useful comments, and Eugene Tyrtyshnikov for his helpful suggestions. The author is also grateful to Hong-Kui Pang and Hai-Wei Sun for numerous fruitful discussions.

REFERENCES

[1] M. Abdou and A. Badr, *On a method for solving an integral equation in the displacement contact problem*, Appl. Math. Comput., Vol. 127 (2002), pp. 65–78.
[2] R. Chan and M. Ng, *Conjugate gradient methods for Toeplitz systems*, SIAM Rev., Vol. 38 (1996), pp. 427–482.
[3] F. Diele, I. Moret and S. Ragni, *Error estimates for polynomial Krylov approximations to matrix functions*, SIAM. J. Matrix Anal. Appl., Vol. 30 (2008), pp. 1546–1565.
[4] J. van den Eshof and M. Hochbruck, *Preconditioning Lanczos approximations to the matrix exponential*, SIAM J. Sci. Comput., Vol. 27 (2006), pp. 1438–1457.
[5] A. Frommer and V. Simoncini, *Stopping criteria for rational matrix functions of Hermitian and symmetric matrices*, SIAM J. Sci. Comput., Vol. 30 (2008), pp. 1387–1412.
[6] E. Gallopoulos and Y. Saad, *Efficient solution of parabolic equations by Krylov approximation methods*, SIAM J. Sci. Statist. Comput., Vol. 13 (1992), pp. 1236–1264.
[7] I. Gohberg and A. Semencul, *On the inversion of finite Toeplitz matrices and their continuous analogs*, Mat. Issled., Vol. 2 (1972), pp. 201–233.
[8] G. Heinig and K. Rost, *Algebraic Methods for Toeplitz-like Matrices and Operators*, Birkhäuser Verlag, Basel, Boston, Stuttgart, 1984.
[9] M. Hochbruck and C. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., Vol. 34 (1997), pp. 1911–1925.
[10] L. Lopez and V. Simoncini, *Analysis of projection methods for rational function approximation to the matrix exponential*, SIAM J. Numer. Anal., Vol. 44 (2006). pp. 613–635.
[11] R. Merton, *Option pricing when underlying stock returns are discontinuous*, J. Financ. Eco., Vol. 3 (1976), pp. 125–144.
[12] C. Moler and C. Van Loan, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Rev., Vol. 45 (2003) pp. 3–49.
[13] I. Moret and P. Novati, *RD-rational approximations of the matrix exponential*, BIT, Vol. 44 (2004), pp. 595–615.
[14] Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., Vol. 29 (1992), pp. 209–228.
[15] D. Tangman, A. Gopaul and M. Bhuruth, *Exponential time integration and Chebychev discretisation schemes for fast pricing of options*, Appl. Numer. Math., Vol. 58 (2008), pp. 1309–1319.
[16] P. Tilli, *Singular values and eigenvalues of non-Hermitian block Toeplitz matrices*, Linear Algebra Appl., Vol. 272 (1998), pp. 59–89.