NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2004; **11**:75–89 (DOI: 10.1002/nla.334)

Estimation of $\mathbf{u}^{\mathrm{T}} f(A)\mathbf{v}$ for large-scale unsymmetric matrices

Hongbin Guo*,† and Rosemary A. Renaut†

Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287 1804, U.S.A.

SUMMARY

Fast algorithms, based on the unsymmetric look-ahead Lanczos and the Arnoldi process, are developed for the estimation of the functional $\Phi(f)=\mathbf{u}^{\mathrm{T}} f(A)\mathbf{v}$ for fixed $\mathbf{u},\mathbf{v}$ and A, where $A\in\mathscr{R}^{n\times n}$ is a large-scale unsymmetric matrix. Numerical results are presented which validate the comparable accuracy of both approaches. Although the Arnoldi process reaches convergence more quickly in some cases, it has greater memory requirements, and may not be suitable for especially large applications. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: matrix function; look-ahead Lanczos; Arnoldi; large-scale unsymmetric matrix

1. INTRODUCTION

We consider algorithms for the approximation of the functional

$$\Phi(f)=\mathbf{u}^{\mathrm{T}} f(A)\mathbf{v} \tag{1}$$

where matrix $A\in\mathscr{R}^{n\times n}$ is a large-scale unsymmetric matrix, and $\mathbf{u},\mathbf{v}\in\mathscr{R}^{n}$ are such that $\|\mathbf{u}\|=\|\mathbf{v}\|=1$. Here $\|\cdot\|$ denotes the 2-norm, and $f(\lambda)$ is a sufficiently smooth function such that the definition of $f(A)$ is meaningful, [1, 2]. Fast algorithms using the Lanczos procedure, and their applications, for the estimation of (1) for symmetric positive define (SPD) matrices A were presented in References [3, 4], and [5–8], respectively. Here we extend the approach for A large and unsymmetric. Krylov subspace methods for estimating matrix exponentials were analysed in Reference [9].

First we review the existing algorithm for the computation of (1) with $\mathbf{u}=\mathbf{v}$ and SPD A. Assume that the SPD matrix A has the eigen decomposition $A=Q\Lambda Q^{\mathrm{T}}$, where Q is an orthogonal matrix, and Λ is a diagonal matrix with entries $\Lambda_{ii}=\lambda_i$ ordered such that $a\leqslant\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n\leqslant b$. Then, rewriting (1) using this decomposition, and introducing $\tilde{\mathbf{v}}=Q^{\mathrm{T}}\mathbf{v}$, we obtain

$$\mathbf{v}^{\mathrm{T}} f(A)\mathbf{v}=\mathbf{v}^{\mathrm{T}} Q f(\Lambda) Q^{\mathrm{T}}\mathbf{v}=\tilde{\mathbf{v}}^{\mathrm{T}} f(\Lambda)\tilde{\mathbf{v}}=\sum_{i=1}^{n} f(\lambda_i)\tilde{v}_i^2$$

---

*Correspondence to: H. Guo, Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287 1804, U.S.A.

†E-mail: hb_guo@asu.edu and renaut@asu.edu

Received 11 January 2002
Revised 31 October 2002

This leads to the reexpression of (1) as the Riemann–Stieltjes integral

$$I(f) = \mathbf{v}^{\mathrm{T}} f(A)\mathbf{v} = \int_a^b f(\lambda)\,\mathrm{d}\mu(\lambda) \tag{2}$$

where

$$\mu(\lambda) = \begin{cases} 0, & a \leqslant \lambda < \lambda_1 \\ \sum_{j=1}^{i} \tilde{v}_j^2, & \lambda_i \leqslant \lambda < \lambda_{i+1} \\ \sum_{j=1}^{n} \tilde{v}_j^2, & \lambda_n \leqslant \lambda \leqslant b \end{cases}$$

Integral (2) may be estimated using the Gauss quadrature rule

$$I[f] = \sum_{j=1}^{m} \omega_j f(\theta_j) + R[f] \tag{3}$$

where weights ω_j and nodes θ_j need to be determined, and $R[f]$ is the quadrature error, [3]. Because this rule can be made exact for polynomials of degree up to at most $2m-1$, the weights and nodes can be obtained through the recursive construction of the orthogonal polynomial series $\{p_j(\lambda)\}$, $j = 0, 1, \ldots, m$. Then, if the recursive coefficients are considered as the elements of a tridiagonal matrix T_m, the nodes θ_j are the eigenvalues of T_m and the weights ω_j are the squares of the last elements of the corresponding unit eigenvectors. Moreover, this construction of the Gaussian rule (3) is equivalent to the Lanczos process with initial vector $\mathbf{v}$, [10, 7]. Specifically, both constructions generate the tridiagonal matrix T_m, and the Lanczos vectors $\mathbf{v}_i$ and the polynomials $p_i(\lambda)$ are related through the relation

$$\mathbf{v}_i = p_{i-1}(A)\mathbf{v}_1, \quad i = 1, \ldots, m+1$$

Significantly, the Lanczos procedure can be used to obtain an efficient estimation of $\mathbf{v}^{\mathrm{T}} f(A)\mathbf{v}$ for large-scale SPD matrix A.

In this paper, we extend and examine this idea to develop efficient algorithms for the estimation of (1) for the general case. Using the Krylov subspace method leads to two different approaches; the unsymmetric Lanczos method and the Arnoldi process. These approaches are described and evaluated by implementation in Sections 2 and 3, respectively. Conclusions are presented in Section 4.

2. UNSYMMETRIC LANCZOS

2.1. Introduction to the look-ahead Lanczos procedure

For ease of presentation of the theoretical development, we consider the Lanczos recurrence with 'monic' vectors (see [11])

$$AV = VH + \mathbf{v}_{m+1}\mathbf{e}_m^{\mathrm{T}} \tag{4}$$

$$A^{\mathrm{T}}U = UH + \mathbf{u}_{m+1}\mathbf{e}_m^{\mathrm{T}} \tag{5}$$

where

$$H = \begin{pmatrix} \alpha_1 & \beta_1 & & & \\ 1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & \alpha_{m-1} & \beta_{m-1} \\ & & & 1 & \alpha_m \end{pmatrix}$$

$$U^{\mathrm{T}} V = D = \operatorname{diag}(\delta_1, \delta_2, \ldots, \delta_m)$$

The main problem with the use of the Lanczos procedure for unsymmetric matrices is the occurrence of breakdown at some stage, $\mathbf{u}_{m+1}^{\mathrm{T}}\mathbf{v}_{m+1} = 0$, [12]. Let $K_i(A, \mathbf{v}) := \operatorname{span}\{\mathbf{v}, A\mathbf{v}, \ldots, A^{i-1}\mathbf{v}\}$ denote the ith Krylov subspace and let

$$L_r = L_r(A, \mathbf{v}_1) := \dim(K_n(A, \mathbf{v}_1))$$

$$L_l = L_l(A^{\mathrm{T}}, \mathbf{u}_1) := \dim(K_n(A^{\mathrm{T}}, \mathbf{u}_1))$$

denote the grades of $\mathbf{v}_1$ and $\mathbf{u}_1$ with respect to A and A^{T}, respectively, [13]. Denote by L the first stage at which breakdown occurs, $\mathbf{u}_{L+1}^{\mathrm{T}}\mathbf{v}_{L+1} = 0$. Breakdown occurs in one of two ways:

1. *Regular termination*:

   $$\mathbf{u}_{L+1}^{\mathrm{T}}\mathbf{v}_{L+1} = 0 \quad \text{and} \quad \mathbf{u}_{L+1} = 0 \quad \text{or} \quad \mathbf{v}_{L+1} = 0$$

   In this case $L_l = L$ or $L_r = L$ and the iteration has terminated with an A-invariant or an A^{T}-invariant subspace, respectively.
2. *Serious breakdown*: breakdown for which neither $\mathbf{u}_{L+1}$ nor $\mathbf{v}_{L+1}$ are zero. In this case $L < L_* := \min\{L_r, L_l\}$ and we have found neither an A-invariant nor an A^{T}-invariant subspace.

Serious breakdown can be further classified as curable or incurable. Look-ahead Lanczos, introduced by Freund *et al.* [11, 14], is utilized to overcome the breakdown for the curable case. Thus, before extending the application of the Lanczos algorithm to the calculation of (1), we first review the look-ahead Lanczos procedure.

For an SPD matrix A, we can define the inner product

$$(\phi, \psi) = \mathbf{v}_1^{\mathrm{T}} \phi(A) \psi(A) \mathbf{v}_1$$

where ϕ, ψ are polynomials, $\phi, \psi \in \mathscr{P}_k$, and $\mathscr{P}_k$ denotes the set of polynomials of degree k. In the unsymmetric case, the Lanczos vectors satisfy $\mathbf{v}_k = \phi_{k-1}(A)\mathbf{v}_1$ and $\mathbf{u}_k = \phi_{k-1}(A^{\mathrm{T}})\mathbf{u}_1$, which leads to the following definition for the formal inner product [10].

Definition 2.1 (*Formal inner product (FIP)*)
We define the FIP by

$$\langle \phi, \psi \rangle = \mathbf{u}_1^{\mathrm{T}} \phi(A) \psi(A) \mathbf{v}_1$$

This inner product does not guarantee non-negativity, and hence is not an inner product in the standard sense. Based on this FIP, we introduce the family of formal orthogonal polynomials (FOP):

Definition 2.2 (*FOP*)
A polynomial ψ_{k-1} is an FOP of degree $k-1$ if $\langle \psi_{k-1}, \phi \rangle = 0$, $\forall \phi \in \mathscr{P}_{k-2}$.

From the above definition, $\psi_{k-1}(\lambda) = \gamma_0 + \gamma_1 \lambda + \cdots + \gamma_{k-1} \lambda^{k-1}, \gamma_{k-1} \neq 0$, is an FOP of degree $k-1$ if and only if its coefficients are a non-trivial solution of the following linear system:

$$\begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_{k-2} \\ \mu_1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \mu_{2k-5} \\ \mu_{k-2} & \cdots & \mu_{2k-5} & \mu_{2k-4} \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{k-2} \end{pmatrix} = -\gamma_{k-1} \begin{pmatrix} \mu_{k-1} \\ \mu_k \\ \vdots \\ \mu_{2k-3} \end{pmatrix} \tag{6}$$

where $\mu_j = \mathbf{u}_1^{\mathrm{T}} A^j \mathbf{v}_1$.

Definition 2.3 (*Regular FOP*)
An FOP ψ_{k-1} is called **regular** if the solution of (6) is unique up to a scalar, otherwise ψ_{k-1} is called a **singular** FOP.

When (6) is inconsistent, there is no FOP ψ_{k-1} of degree $k-1$. We refer to this case, which occurs due to the indefiniteness of the formal inner product, as **deficient**.

Proposition 2.4
Let ψ_{k-1} be a *regular* FOP of degree $k-1$. Then, a *regular* FOP of degree k exists if and only if $\langle \psi_{k-1}, \psi_{k-1} \rangle \neq 0$, i.e. $\mathbf{u}_k^{\mathrm{T}} \mathbf{v}_k \neq 0$.

There is a maximal subset of indices

$$\{k_1, k_2, \ldots, k_J\} \subseteq \{1, 2, \ldots, L_*\}, \quad k_1 = 1 < k_2 < \cdots, k_J \leqslant L_*$$

such that $\forall j = 1, \ldots, J$, there exists a monic regular FOP $\psi_{k_j - 1} \in \mathscr{P}_{k_j - 1}$. The regular FOPs $\psi_{k_{j-1}-1}$, $\psi_{k_j - 1}$ and $\psi_{k_{j+1}-1}$ are connected through a three-term recurrence [15].

Definition 2.5 (*Regular vector*)
Corresponding to the regular FOPs we define the regular vector series $\{\mathbf{v}_{k_j}\}$ and $\{\mathbf{u}_{k_j}\}$,

$$\mathbf{v}_{k_j} = \psi_{k_j - 1}(A) \mathbf{v}_1, \quad \mathbf{u}_{k_j} = \psi_{k_j - 1}(A^{\mathrm{T}}) \mathbf{u}_1 \tag{7}$$

From Definition 2.3, the regular vectors satisfy

$$\begin{aligned} \mathbf{u}_{k_j}^{\mathrm{T}}\mathbf{v} &= 0 \quad \forall\ \mathbf{v} \in K_{k_j-1}(A, \mathbf{v}_1) \\ \mathbf{v}_{k_j}^{\mathrm{T}}\mathbf{u} &= 0 \quad \forall\ \mathbf{u} \in K_{k_j-1}(A^{\mathrm{T}}, \mathbf{u}_1) \end{aligned} \tag{8}$$

Thus, if $k_J = L_*$, either or both of the regular vector series $\{\mathbf{v}_{k_j}\}_{j=1}^{J}$, $\{\mathbf{u}_{k_j}\}_{j=1}^{J}$ can be extended to span an A-invariant or an A^{T}-invariant subspace by the look-ahead Lanczos process. On the other hand, if $k_J < L_*$, an invariant subspace cannot be constructed and the breakdown is incurable.

The vectors which are used to extend the series

$$\begin{gathered} \mathbf{v}_k \in K_k(A, \mathbf{v}_1) \backslash K_{k-1}(A, \mathbf{v}_1), \quad \mathbf{u}_k \in K_k(A^{\mathrm{T}}, \mathbf{u}_1) \backslash K_{k-1}(A^{\mathrm{T}}, \mathbf{u}_1) \\ k = k_{j-1} + 1, \ldots, k_j - 1, \quad j = 1, 2, \ldots, J \end{gathered}$$

are called inner vectors, and the extended series provide complete bases for the Krylov subspaces $K_k(A, \mathbf{v}_1)$ and $K_k(A^{\mathrm{T}}, \mathbf{u}_1)$. These bases are divided into J blocks, $V_j, U_j,\ j = 1, \ldots, J$,

$$V_j = \{\mathbf{v}_i \mid i = k_j, k_j + 1, \ldots, k_{j+1} - 1\}, \quad U_j = \{\mathbf{u}_i \mid i = k_j, k_j + 1, \ldots, k_{j+1} - 1\}$$

which are block biorthogonal

$$(U^{(m)})^{\mathrm{T}} V^{(m)} = D^{(m)}, \quad U^{(m)} = [U_1, \ldots, U_l], \quad V^{(m)} = [V_1, \ldots, V_l]$$

Here m is the total number of generated left (right) Lanczos vectors and $D^{(m)}$ is a non-singular block diagonal matrix.

The Lanczos process with look-ahead replaces (4) and (5) by

$$AV^{(m)} = V^{(m)} H^{(m)} + \mathbf{v}_{m+1} e_m^{\mathrm{T}} \tag{9}$$

$$A^{\mathrm{T}} U^{(m)} = U^{(m)} H^{(m)} + \mathbf{u}_{m+1} e_m^{\mathrm{T}} \tag{10}$$

where

$$\begin{gathered} (U^{(m)})^{\mathrm{T}} V^{(m)} = D^{(m)} = \operatorname{diag}(\delta_1, \ldots, \delta_l) \\ \delta_i = U_i^{\mathrm{T}} V_i \quad i = 1, 2, \ldots, l(m) \end{gathered}$$

and $H^{(m)}$ is a block tridiagonal upper Hessenberg matrix

$$H^{(m)} = \begin{pmatrix} \alpha_1 & \beta_1 & & \\ \gamma_1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_{l-1} \\ & & \gamma_{l-1} & \alpha_l \end{pmatrix}$$

Here

$$\alpha_i = \begin{pmatrix} * & \cdots & \cdots & * \\ 1 & \ddots & & \vdots \\ & \ddots & \ddots & \\ & & 1 & * \end{pmatrix}$$

$\gamma_i = \mathbf{e}_1 \mathbf{e}_{p_i}^{\mathrm{T}}$, the p_i are the number of columns of block i, and the β_i's are, in general, full matrices.

Let k_j denote the index of the regular vector of the jth block, and $l = l(k)$ the number of the block to which $\mathbf{v}_k$ and $\mathbf{u}_k$ belong. If $m = k_{l+1} - 1$, block $l(m)$ is called **complete**.

For numerical stability, $\mathbf{u}_k$ and $\mathbf{v}_k$ are normalized, $\hat{\mathbf{u}}_k = \mathbf{u}_k / \|\mathbf{u}_k\|$, $\hat{\mathbf{v}}_k = \mathbf{v}_k / \|\mathbf{v}_k\|$.

2.2. *Look-ahead Lanczos for estimating* $\mathbf{u}^{\mathrm{T}} f(A)\mathbf{v}$

Here, where it causes no confusion, we omit the superscript (m). We can use $VH^{-1}\mathbf{e}_1$ to approximate $A^{-1}\mathbf{v}_1$ providing $\|\mathbf{v}_{m+1}\| \cdot |e_m^{\mathrm{T}} H^{-1} e_1|$ is small, as is common for solving linear systems of equations, see Reference [16, Chapter 7.2]. Similarly,

$$(zI - A)^{-1}\mathbf{v}_1 \approx V(zI - H)^{-1}\mathbf{e}_1$$

when z is not an eigenvalue of A and H_m.

Let Γ be a closed contour which encircles both $\mathscr{F}(A) = \{\mathbf{x}^H A\mathbf{x}\colon \mathbf{x} \in C^n, \|\mathbf{x}\| = 1\}$ and $\mathscr{F}(H)$, and such that $f(z)$ is analytic inside Γ.

Using the definition for the matrix function given in Reference [2], we have

$$f(A)\mathbf{v}_1 = \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma} f(z)(zI - A)^{-1}\mathbf{v}_1 \, \mathrm{d}z$$

and

$$\frac{1}{2\pi \mathrm{i}} \oint_{\Gamma} f(z) V(zI - H)^{-1}\mathbf{e}_1 \, \mathrm{d}z = Vf(H)\mathbf{e}_1$$

The previous two equations suggest that we can approximate $\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1$ by $\mathbf{u}_1^{\mathrm{T}} Vf(H)\mathbf{e}_1$,

$$\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1 = \mathbf{u}_1^{\mathrm{T}} Vf(H)\mathbf{e}_1 + \tilde{R}[f] \tag{11}$$

provided that $R(f)$ satisfies reasonable bounds relative to $\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1$.

In Reference [9], which deals with Krylov subspace methods for matrix exponentials, general upper bounds of $\|f(A)\mathbf{v}_1 - Vf(H)\mathbf{e}_1\|$ were estimated. These results were refined for Arnoldi approximations of $\exp(\tau A)\mathbf{v}_1$ for certain Hermitian matrices. Here, we use an alternative approach, which uses specific properties of the look-ahead Lanczos, and which is based on demonstrating that the estimate of $\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1$ by Gaussian quadrature on the complex plane is just $\mathbf{e}_1^{\mathrm{T}} Df(H)\mathbf{e}_1$.

In order to obtain our main result we need the following four intermediary Lemmas 2.6–2.9.

Lemma 2.6
Let $\xi_i, i=1,2,\dots,m-1$, be the lower off-diagonal entries of the upper Hessenberg matrix H, then

$$\mathbf{e}_m^{\mathrm{T}} H^j \mathbf{e}_1 = \begin{cases} 0, & 0 \leqslant j < m-1 \\ \prod_{i=1}^{m-1} \xi_i, & j = m-1 \end{cases}$$

For future reference we use ρ to denote the value $\mathbf{e}_m^{\mathrm{T}} H^{m-1} \mathbf{e}_1$.

Lemma 2.7

$$A^j V \mathbf{e}_1 = \begin{cases} VH^j \mathbf{e}_1, & 0 \leqslant j \leqslant m-1 \\ VH^m \mathbf{e}_1 + \rho \mathbf{v}_{m+1}, & j = m \end{cases}$$

$$(A^{\mathrm{T}})^j U \mathbf{e}_1 = \begin{cases} UH^j \mathbf{e}_1, & 0 \leqslant j \leqslant m-1 \\ UH^m \mathbf{e}_1 + \rho \mathbf{u}_{m+1}, & j = m \end{cases}$$

Proof
These relationships follow immediately from Lemma 2.6 and

$$A^j V = VH^j + \sum_{i=0}^{j-1} A^i \mathbf{v}_{m+1} \mathbf{e}_m^{\mathrm{T}} H^{j-i-1}$$

$$(A^{\mathrm{T}})^j U = UH^j + \sum_{i=0}^{j-1} (A^{\mathrm{T}})^i \mathbf{u}_{m+1} \mathbf{e}_m^{\mathrm{T}} H^{j-i-1} \qquad \square$$

Lemma 2.8
If the block $l(m)$ is complete, then, for $0 \leqslant j \leqslant m$, $DH^j \mathbf{e}_1 = (H^{\mathrm{T}})^j D \mathbf{e}_1$.

Proof
Because the block $l(m)$ is complete, by (8) $U^{\mathrm{T}} \mathbf{v}_{m+1} = 0$. But, by Lemma 2.7,

$$U^{\mathrm{T}}(A^j V \mathbf{e}_1) = DH^j \mathbf{e}_1, \quad 0 \leqslant j \leqslant m$$

and, using (8), $\mathbf{u}_{m+1}^{\mathrm{T}} \mathbf{v} = 0$, $\forall\ \mathbf{v} \in K_m(A, \mathbf{v}_1)$. Thus, $\mathbf{u}_{m+1}^{\mathrm{T}} A^{j-i-1} \mathbf{v}_1 = 0$, $0 \leqslant j \leqslant m$, $0 \leqslant i \leqslant j-1$. Furthermore,

$$(U^{\mathrm{T}} A^j)\mathbf{v}_1 = \left((H^{\mathrm{T}})^j U^{\mathrm{T}} + \sum_{i=0}^{j-1} (H^{\mathrm{T}})^i \mathbf{e}_m \mathbf{u}_{m+1}^{\mathrm{T}} A^{j-i-1} \right) \mathbf{v}_1 = (H^{\mathrm{T}})^j D \mathbf{e}_1$$

and the result follows. $\square$

Lemma 2.9
D is a symmetric matrix.

Theorem 2.10
Let $A \in R^{n\times n}$ be unsymmetric, and block $l(m)$ of the look-ahead Lanczos procedure (9), (10) be complete. Then:

1. $\mathbf{u}_1^{\mathrm{T}}\phi(A)\mathbf{v}_1 = \mathbf{e}_1^{\mathrm{T}} D\phi(H)\mathbf{e}_1$, for any $\phi \in \mathscr{P}_{2m-1}$;
2. $\mathbf{u}_1^{\mathrm{T}} A^{2m}\mathbf{v}_1 = \mathbf{e}_1^{\mathrm{T}} DH^{2m}\mathbf{e}_1 + \rho^2\mathbf{u}_{m+1}^{\mathrm{T}}\mathbf{v}_{m+1}$.

Proof
We present the proof for the first result, the second follows similarly. We first show the result for $\phi(\lambda)=\lambda^i$, $i=0,1,\dots,2m-1$. Suppose $i\leqslant 2m-2$. For any j_1, j_2 which satisfy $0\leqslant j_1, j_2 \leqslant m-1$,

$$\begin{aligned}\mathbf{u}_1^{\mathrm{T}} A^{j_1+j_2}\mathbf{v}_1 &= \mathbf{e}_1^{\mathrm{T}} U^{\mathrm{T}} A^{j_1} A^{j_2} V\mathbf{e}_1\\ &= \mathbf{e}_1^{\mathrm{T}} (H^{\mathrm{T}})^{j_1} U^{\mathrm{T}} V H^{j_2}\mathbf{e}_1\\ &= \mathbf{e}_1^{\mathrm{T}} (H^{\mathrm{T}})^{j_1} D H^{j_2}\mathbf{e}_1\end{aligned}$$

by Lemma 2.7. Thus, by Lemma 2.8, the result holds for $i\leqslant 2m-2$.

For the case $i=2m-1$ we immediately obtain

$$\mathbf{u}_1^{\mathrm{T}} A^m A^{m-1}\mathbf{v}_1 = (\mathbf{e}_1^{\mathrm{T}}(H^{\mathrm{T}})^m U^{\mathrm{T}} + \rho\mathbf{u}_{m+1}^{\mathrm{T}}) V H^{m-1}\mathbf{e}_1$$

by Lemma 2.7, and by (8) $\mathbf{u}_{m+1}^{\mathrm{T}} V = 0$. Then, using the symmetry of D and Lemma 2.8,

$$\begin{aligned}\mathbf{e}_1^{\mathrm{T}}(H^{\mathrm{T}})^m U^{\mathrm{T}} V H^{m-1}\mathbf{e}_1 &= \mathbf{e}_1^{\mathrm{T}}(H^{\mathrm{T}})^m D H^{m-1}\mathbf{e}_1\\ &= \mathbf{e}_1^{\mathrm{T}}(H^{\mathrm{T}})^m (H^{\mathrm{T}})^{m-1} D\mathbf{e}_1\\ &= \mathbf{e}_1^{\mathrm{T}} D H^{2m-1}\mathbf{e}_1\end{aligned}$$

□

Theorem 2.10 shows that, when $l(m)$ is complete, application of Gaussian quadrature on the complex plane for $\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1$ is given by $\mathbf{e}_1^{\mathrm{T}} D f(H)\mathbf{e}_1$. Thus, when using the look-ahead Lanczos process to estimate (1), we need to make sure that we complete the last block $l(m)$. Moreover, this theorem generalizes the following result given in Reference [17].

Theorem 2.11

$$\mathbf{u}_1^{\mathrm{T}} A^j \mathbf{v}_1 = \begin{cases} \delta_1 \mathbf{e}_1^{\mathrm{T}} H^j \mathbf{e}_1, & j\leqslant 2m-1\\ \delta_1 \mathbf{e}_1^{\mathrm{T}} H^{2m}\mathbf{e}_1 + \delta_{m+1}\prod_{i=1}^m \gamma_i^2, & j=2m\end{cases}$$

where γ_i are the lower off-diagonal entries of H.

Remark 2.1
In summary, the functional Φ, (1), can be estimated by Gaussian quadrature,

$$\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1 = \mathbf{e}_1^{\mathrm{T}} D f(H)\mathbf{e}_1 + \tilde{R}[f]$$

2.3. Algorithm

We use the results of the previous section to obtain the numerical algorithm for the calculation of (1).

Algorithm 1

```
For $l = 1, 2, \ldots$ until a specified error tolerance is satisfied.
    1. Run the look-ahead Lanczos process with initial vectors $\mathbf{u}_1$ and $\mathbf{v}_1$ until
       block $l$ is complete.
    2. Compute $\mathbf{e}_1^{\mathrm{T}} D f(H)\mathbf{e}_1$.
End for
```

Remark 2.2
To stably compute $f(H)$ we may use the real Schur decomposition of H. If $H = WTW^{\mathrm{T}}$, then $f(H) = Wf(T)W^{\mathrm{T}}$. Thus, $\mathbf{e}_1^{\mathrm{T}} D f(H)\mathbf{e}_1 = \mathbf{e}_1^{\mathrm{T}} DWf(T)W^{\mathrm{T}}\mathbf{e}_1$. To calculate $f(T)$ for a block upper triangular matrix, T, we use the perfect recursive algorithm described in Reference [18].

2.4. Error analysis

Theorem 2.12
Suppose that the Taylor series representation of $f(z)$ is $f(z) = \sum_{k=0}^{\infty} c_k z^k$ on *an open disk containing* $\lambda(H)$, and that the last block $l(m)$ of the look-ahead Lanczos procedure is complete. Then, $|\tilde{R}(f)|$ is bounded,

$$\begin{aligned}|\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1 - \mathbf{e}_1^{\mathrm{T}} D f(H)\mathbf{e}_1| \leqslant{} & |c_{2m}\rho^2 \mathbf{u}_{m+1}^{\mathrm{T}}\mathbf{v}_{m+1}| \\ & + \frac{n}{(2m+1)!}\max_{0\leqslant s\leqslant 1} \|A^{2m+1} f^{(2m+1)}(As)\|_2 \\ & + \frac{m}{(2m+1)!}\max_{0\leqslant s\leqslant 1} \|H^{2m+1} f^{(2m+1)}(Hs)\|_2 \|De_1\|\end{aligned}$$

Proof
The result follows from Theorems 2.10 and 11.2.4 of Reference [1]. □

Error analysis results for general functions can be obtained similarly to the results presented in Reference [7]. These theoretical results are, however, of limited utility for the development of practical stopping criteria. Thus, because of the importance of $\mathbf{u}_1^{\mathrm{T}} A^{-1}\mathbf{v}_1$ in applications, we emphasize the error analysis for $f(\lambda) = 1/\lambda$, and, where needed, assume that H^{-1} exists.

Theorem 2.13
If the last block $l(m)$ of the look-ahead Lanczos procedure is complete, then

$$\mathbf{u}_1^{\mathrm{T}} A^{-1}\mathbf{v}_1 - \mathbf{e}_1^{\mathrm{T}} D H^{-1}\mathbf{e}_1 = \mathbf{u}_{m+1}^{\mathrm{T}} A^{-1}\mathbf{v}_{m+1}(\mathbf{e}_m^{\mathrm{T}} H^{-1}\mathbf{e}_1)^2$$

Proof
We multiply (9) and (10) on the right by H^{-1}, and then on the left by A^{-1} and $(A^{\mathrm{T}})^{-1}$, respectively.

$$VH^{-1} = A^{-1}V + A^{-1}\mathbf{v}_{m+1}\mathbf{e}_m^{\mathrm{T}}H^{-1}$$

$$UH^{-1} = (A^{\mathrm{T}})^{-1}U + (A^{\mathrm{T}})^{-1}\mathbf{u}_{m+1}\mathbf{e}_m^{\mathrm{T}}H^{-1}$$

Multiplying on the right by $\mathbf{e}_1$, and on the left by $\mathbf{u}_1^{\mathrm{T}}$ and $\mathbf{v}_{m+1}^{\mathrm{T}}$, respectively, yields

$$\mathbf{e}_1^{\mathrm{T}}DH^{-1}\mathbf{e}_1 = \mathbf{u}_1^{\mathrm{T}}A^{-1}\mathbf{v}_1 + \mathbf{u}_1^{\mathrm{T}}A^{-1}\mathbf{v}_{m+1}\mathbf{e}_m^{\mathrm{T}}H^{-1}\mathbf{e}_1$$

$$0 = \mathbf{v}_{m+1}^{\mathrm{T}}(A^{\mathrm{T}})^{-1}\mathbf{u}_1 + \mathbf{v}_{m+1}^{\mathrm{T}}(A^{\mathrm{T}})^{-1}\mathbf{u}_{m+1}\mathbf{e}_m^{\mathrm{T}}H^{-1}\mathbf{e}_1$$

The result follows by solving the latter equation for $\mathbf{u}_1^{\mathrm{T}}A^{-1}\mathbf{v}_{m+1}$ and substituting in the former. □

Numerical tests show that $\mathbf{u}_{m+1}^{\mathrm{T}}A^{-1}\mathbf{v}_{m+1}$ does not change too much when m increases and $\mathbf{e}_m^{\mathrm{T}}H^{-1}\mathbf{e}_1$ dominates the convergence. Thus, we only need to give a coarse estimate of $\mathbf{u}_{m+1}^{\mathrm{T}}A^{-1}\mathbf{v}_{m+1}$. For the computation of $\mathbf{e}_m^{\mathrm{T}}H^{-1}\mathbf{e}_1$, the LU decomposition of $H^{(m)} = L_mU_m$ can be obtained recursively. The speed of convergence is estimated through the following result.

Theorem 2.14
Given the QL decomposition $H^{\mathrm{T}} = QL$, we have

$$|\mathbf{e}_m^{\mathrm{T}}H^{-1}\mathbf{e}_1| \leqslant \varepsilon/|l_{11}|$$

where $\varepsilon = \min_{p(0)=1, p\in\tilde{\mathscr{P}}_{m-1}} \|\mathbf{e}_m^{\mathrm{T}}p(H)\|^2$, and $\tilde{\mathscr{P}}_m$ denotes the set of polynomials of degree less than or equal m.

Proof
For any $g \in \tilde{\mathscr{P}}_{m-2}$, by Lemma (2.6),

$$|\mathbf{e}_m^{\mathrm{T}}H^{-1}\mathbf{e}_1| = |\mathbf{e}_m^{\mathrm{T}}[H^{-1} + g(H)]\mathbf{e}_1|$$

$$= |\mathbf{e}_m^{\mathrm{T}}[I + g(H)H]H^{-1}\mathbf{e}_1|$$

Thus,

$$|\mathbf{e}_m^{\mathrm{T}}H^{-1}\mathbf{e}_1| \leqslant \|H^{-1}\mathbf{e}_1\|_2 \min_{p(0)=1, p\in\tilde{\mathscr{P}}_{m-1}} \|\mathbf{e}_m^{\mathrm{T}}p(H)\|_2$$

$$= \|H^{-1}\mathbf{e}_1\|_2\varepsilon$$

Since $HQ = L^{\mathrm{T}}$, $HQ\mathbf{e}_1 = l_{11}\mathbf{e}_1$. Thus, $\|H^{-1}\mathbf{e}_1\|_2 = \|Q\mathbf{e}_1\|_2/|l_{11}| = 1/|l_{11}|$, and the result follows. □

Table I. Estimation of (1) for unsymmetric matrix A.

Matrices	κ	Function	Steps m	Relative error
grcar	5.39	exp(x)	6	4.84e-11
grcar	5.39	exp(x)	41	6.71e-14
parter	104.8	1/x	43	1.49e-03

2.5. Numerical results

We present results for two examples, the unsymmetric matrices *grcar* and *parter* of degree 200 from Higham's test toolbox, [19]. The initial vectors $\mathbf{u}_1$ and $\mathbf{v}_1$ are the vectors $(1,2,\dots,n)^{\mathrm{T}}$ and $(n,\dots,1)^{\mathrm{T}}$, with normalization, $\|\mathbf{u}_1\| = \|\mathbf{v}_1\| = 1$. Throughout we also normalize $\mathbf{u}_i$ and $\mathbf{v}_i$. The results are presented in Table I. Here κ is the 1-norm condition number $\|A\|_1 \|A^{-1}\|_1$. For matrix *parter*, the 40th block has degree 4, and the 39th block of matrix *grcar*, has degree 3.

3. THE ARNOLDI PROCEDURE

The Arnoldi process may also be used to construct an algorithm for the estimation of (1). Given the initial vector $\mathbf{v}_1$, the Arnoldi process generates the orthogonal basis $V_m = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$ for the Krylov subspace $K_m(A, \mathbf{v}_1)$. In matrix form

$$AV_m = V_m H_m + h_{m+1,m} \mathbf{v}_{m+1} \mathbf{e}_m^{\mathrm{T}} \tag{12}$$

where H_m is an upper Hessenberg matrix.

Proposition 3.1
If $0 \leqslant j \leqslant m-1$, then $A^j \mathbf{v}_1 = V_m H_m^j \mathbf{e}_1$.

Proof
The result follows from $\mathbf{e}_m^{\mathrm{T}} H_m^j \mathbf{e}_1 = 0$ when $j \leqslant m-2$. □

The same arguments employed in Section 2.2 for the look-ahead Lanczos can also be used to show that $\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1$ can be approximated by $\mathbf{u}_1^{\mathrm{T}} V_m f(H_m)\mathbf{e}_1$. In the Arnoldi case, because $\mathscr{F}(H_m) \subseteq \mathscr{F}(A)$, it is sufficient for the closed contour Γ to encircle $\mathscr{F}(A)$. We note that Proposition 3.1 does not hold for $j > m-1$. Thus, for approximation of $f(A)\mathbf{v}_1$ the Arnoldi method has algebraic accuracy $m-1$. Because the initial vector $\mathbf{u}_1^{\mathrm{T}}$ does not impact the Arnoldi procedure, the angle between $\mathbf{u}_1$ and the residual, $\mathbf{r} = f(A)\mathbf{v}_1 - V_m f(H_m)\mathbf{e}_1$, is the second factor which affects the speed of approximation $\mathbf{u}_1^{\mathrm{T}} f(A)\mathbf{v}_1 \approx \mathbf{u}_1^{\mathrm{T}} V_m f(H_m)\mathbf{e}_1$.

The Arnoldi method presents two advantages compared to the look-ahead Lanczos method; it will not breakdown and the algorithm is easily implemented. The Lanczos process requires two matrix–vector operations per step, as compared with one for the Arnoldi process. But the algebraic accuracy for m steps of Lanczos is comparable to that for $2m$ steps of Arnoldi. On the other hand, the Arnoldi process requires long recurrences, with linear increase in memory. Consequently, storage may limit the accuracy achievable by Arnoldi for large-scale A. Moreover, for this application the Arnoldi process cannot be restarted because the initial vector $\mathbf{v}_1$ is fixed.

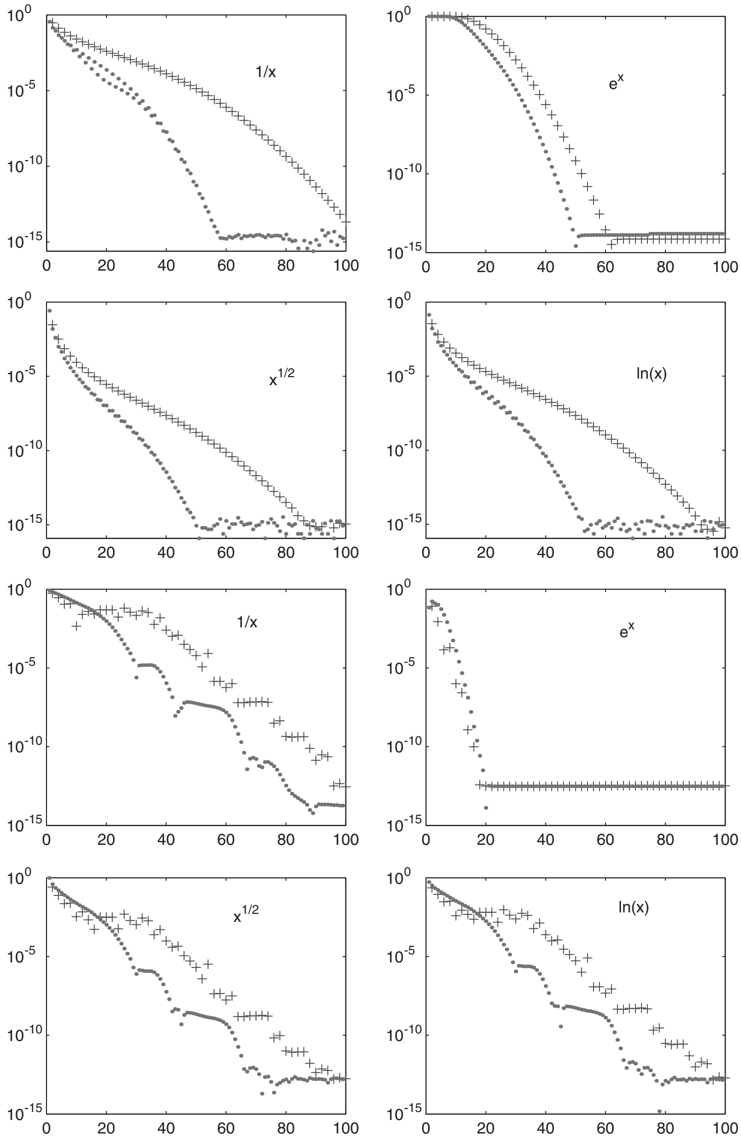

Figure 1. Relative error against number of matrix vector operations, for given functions f and with matrices A_1, and A_2, above and below, resp. In each case '$+$' denotes the Lanczos result and '$\cdot$' the Arnoldi result.

To compare the performance of these two methods we evaluate the functional $\mathbf{u}^{\mathrm{T}} f(A)\mathbf{v}$ for four choices of f and three different matrices. The vectors $\mathbf{u}_1$ and $\mathbf{v}_1$ are initialized as in Section 2.5. The cases tested are as follows:

- Matrix A_1 is the bidiagonal matrix with diagonal $[1, 2, \ldots, 100]$ and all upper off-diagonal elements identically 1. This matrix has spectrum condition number 116.75.

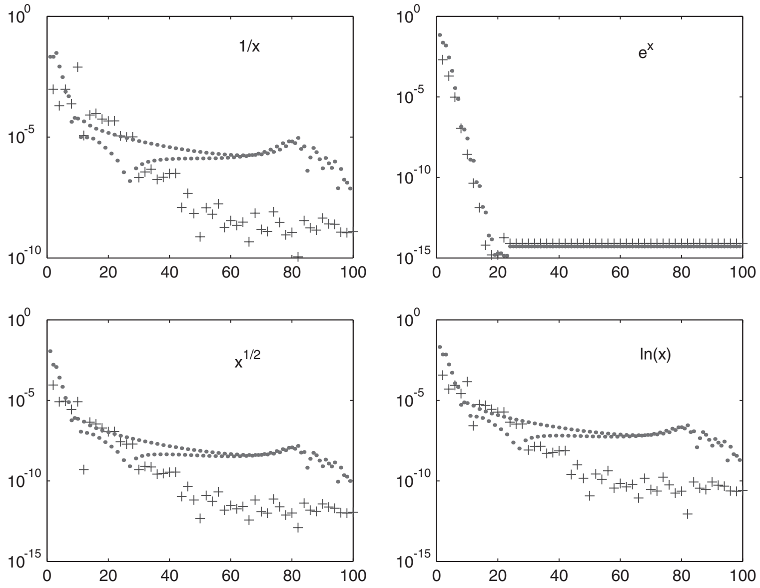

Figure 2. Relative error against number of matrix vector operations, for given functions f and with matrices A_3. In each case '$+$' denotes the Lanczos result and '$\cdot$' the Arnoldi result.

- Matrix A_2 is obtained from the second-order discretization of the convection–diffusion equation

$$\begin{aligned} u_{xx} + u_{yy} + u_x &= f, \quad (x, y) \quad \text{in } [0,1]^2 \\ u &= g, \quad (x, y) \quad \text{on } \partial[0,1]^2 \end{aligned}$$

  with $h = 1/21$. This matrix has spectrum condition number 257.6.
- Matrix $A_3 = gallery('grcar', 200)$ is generated from MATLAB GALLERY, Higham test matrices [19]. While it has spectrum condition number 3.6178, its eigenvalues are sensitive. Originally, it was designed to test the eigenvalue problem.

The functions chosen are $f(x) = x^{-1}, e^x, \sqrt{x}$ and $\ln x$, respectively. The results are illustrated in Figures 1 and 2. In each case, the relative errors are plotted against the number of matrix–vector operations, rather than the number of steps.

The presented results for both matrices A_1 and A_2 verify the analysis of algebraic accuracy, in particular for the function $f(x) = e^x$. From the results for matrices A_1 and A_2 we see that the Arnoldi process outperforms the Lanczos approach, while for matrix A_3 the Lanczos method converges more quickly. Clearly, from Figure 2, the behaviour of the Arnoldi is impacted by the sensitivity of the eigenvalues for A_3 while stagnation occurs with Arnoldi, this is not the case for the Lanczos method. To examine the better performance of Arnoldi, compared to Lanczos, for matrices A_1 and A_2, we illustrate in Figure 3 the inner product $\mathbf{u}_1^{\mathrm{T}}\mathbf{r}/\|\mathbf{r}\|_2$. This confirms that $\mathbf{u}_1^{\mathrm{T}}\mathbf{r}/\|\mathbf{r}\|_2$ goes to zero quickly in the early stages, providing fast convergence for Arnoldi.

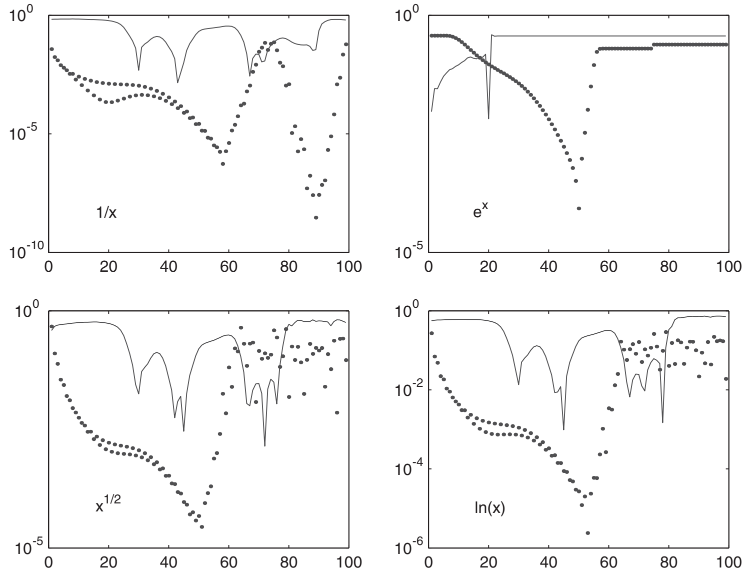

Figure 3. Arnoldi method: $\mathbf{u}_1^{\mathrm{T}}\mathbf{r}/\|\mathbf{r}\|_2$ against the number of steps, for matrix A_1 and A_2. In each case '-' denotes the inner products for matrix A_2 and '·' the matrix result A_1.

4. CONCLUSIONS

Algorithms for the estimation of $\mathbf{u}^{\mathrm{T}}f(A)\mathbf{v}$ for large-scale unsymmetric matrices have been developed and numerically validated. Their ability to accurately estimate $\mathbf{u}^{\mathrm{T}}f(A)\mathbf{v}$ depends on the function f and the matrix A. Future work will focus on development of the error analysis for other functions.

ACKNOWLEDGEMENTS

We thank the anonymous referee for drawing our attention to Reference [9], which describes relevant related work for estimating $f(A)\mathbf{v}$. This work has been supported in part by the Arizona Alzheimer's Disease Research Center which is funded by the Arizona Department of Health Services.

REFERENCES

1. Golub GH, van Loan C. *Matrix Computations* (3rd edn). John Hopkins Press: Baltimore, MD, 1996.
2. Rinehart RF. The equivalence of definitions of a matrix function. *American Mathematical Monthly* 1995; **62**:395–414.
3. Golub GH. Some modified matrix eigenvalue problems. *SIAM Review* 1973; **15**:318–334.
4. Bai Z, Fahey M, Golub G. Some large-scale matrix computation problems. *Journal of Computational and Applied Mathematics* 1996; **74**:71–89.
5. Bai Z, Fahey M, Golub G, Menon M, Richter E. Computing partial eigenvalue sum in electronic structure calculations. *Technical Report*, SCCM, Stanford University, SCCM-98-03, 1998.
6. Nguyen N, Milanfar P, Golub G. Blind restoration and superresolution using generalized cross-validation with Gauss quadrature rules. *Technical Report*, SCCM, Stanford University, SCCM-00-02, 2000.
7. Guo H. Computing trace of function of matrix. *Numerical Mathematics* (A Journal of Chinese Universities) 2000; **22**(2):204–215.

8. Guo H. IRR: an algorithm for computing the smallest singular value of large-scale matrices. *International Journal of Computer Mathematics* 2001; **77**(2):89–104.
9. Hochbruck M, Lubich C. On Krylov subspace approximations to the matrix exponential operator. *SIAM Journal on Numerical Analysis* 1997; **34**:1911–1925.
10. Lanczos C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards* 1950; **45**:225–280.
11. Freund RW, Gutknecht M, Nachtigal N. An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices. *SIAM Journal of Science and Statistical Computing* 1993; **14**:137–158.
12. Gutknecht MH. A completed theory of the unsymmetric Lanczos process and related algorithm, Part I. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:594–639.
13. Wilkinson JH. *Algebraic Eigenvalue Problems.* Clarendon Press: Oxford, 1965.
14. Gutknecht MH. A completed theory of the unsymmetric Lanczos process and related algorithm, Part II. *SIAM Journal on Matrix Analysis and Applications* 1994; **15**:15–58.
15. Draux A. *Polynômes Orthogonaux formels-Applications*. Lecture Notes in Mathematics, vol. 974. Springer: Berlin, 1983.
16. Saad Y. *Iterative Methods for Sparse Linear Systems.* PWS Press, 1996.
17. Bai Z, Ye Q. Error estimation of the Padé approximation of transfer function via the Lanczos process. *Electronic Transactions on Numerical Analysis* 1998; **7**:1–17.
18. Parlett BN. Computation of functions of triangular matrices. *Memorandum*, No. ERLM481, Electronics Research Laboratory, College of Engineering, University of California, Berkeley, 1974.
19. Higham NJ. *The test matrix toolbox for Matlab.* No. 237, University of Manchester, England, 1993.