THE RADAU-LANCZOS METHOD FOR FUNCTIONS OF MATRICES*

ANDREAS FROMMER†, KATHRYN LUND-NGUYEN‡, MARCEL SCHWEITZER†, AND DANIEL B. SZYLD‡

Abstract. We present a new iterative method for computing $f(A)\boldsymbol{b}$, derived from a relationship between the standard Lanczos method and a Gauss-Radau quadrature rule. We show that this method, called the Radau-Lanczos method, converges when A is Hermitian positive definite and f is a Stieltjes function. We also show that the restarted version of this method converges and present numerical results showing this method performing better than the standard Lanczos method in terms of attainable error norm and iteration count.

1. Introduction. A problem of increasing importance in scientific computations is the evaluation of $f(A)\boldsymbol{b}$, where f is a scalar function, $A \in \mathbb{C}^{n\times n}$, and $\boldsymbol{b} \in \mathbb{C}^n$. Frequently A is large and sparse, making the direct computation of $f(A)$ infeasible, but not $f(A)\boldsymbol{b}$.

The main contribution of this paper is a new method, which we call the Radau-Lanczos method, for computing $f(A)\boldsymbol{b}$. We show it converges when A is Hermitian positive definite (HPD) and f is a Stieltjes function. Furthermore, this method improves upon the standard restarted Lanczos method for functions of HPD matrices [1, 2, 3, 4, 11].

An outline of the paper is as follows. We begin by establishing properties of the standard method and of the Lanczos relation. In section 2, we describe the Radau-Lanczos method for linear systems, including a variational characterization that yields error bounds similar to those for the conjugate gradients (CG) method. In section 3, we apply the Radau-Lanczos method to Stieltjes functions of HPD matrices and show that the restarted version converges by providing convergence bounds. Finally, in section 4, we illustrate how our method compares to the standard Lanczos method with numerical experiments.

1.1. The standard Lanczos method. We begin by considering the unrestarted standard Lanczos method for $A \in \mathbb{C}^{n\times n}$ HPD and $f(z) = z^{-1}$, also known as CG. Consider the linear system

$$A\boldsymbol{x} = \boldsymbol{b}. \tag{1.1}$$

Let $\boldsymbol{x}_*$ be the exact solution to (1.1); $\boldsymbol{x}_0$ the starting approximation; $\boldsymbol{x}_m$ the iterates; $\boldsymbol{e}_m = \boldsymbol{x}_* - \boldsymbol{x}_m$ the errors; and $\boldsymbol{r}_m = A\boldsymbol{e}_m = \boldsymbol{b} - A\boldsymbol{x}_m$ the residuals. We also let $\mathcal{K}_m(A, \boldsymbol{r}_0)$ denote the m-th Krylov subspace and Π_m the space of all polynomials of degree at most m. Then $\mathcal{K}_m(A, \boldsymbol{r}_0) = \{p(A)\boldsymbol{r}_0 : p \in \Pi_{m-1}\}$. We refer to the following as the *Lanczos relation*:

$$AV_m = V_m T_m + t_{m+1,m}\boldsymbol{v}_{m+1}\widehat{\boldsymbol{e}}_m^H, \tag{1.2}$$

where the columns of $V_m \in \mathbb{C}^{n\times m}$ form an orthonormal basis of $\mathcal{K}_m(A, \boldsymbol{r}_0)$; $T_m = V_m^H A V_m \in \mathbb{C}^{m\times m}$ is the restriction and projection of A onto $\mathcal{K}_m(A, \boldsymbol{r}_0)$; and $\widehat{\boldsymbol{e}}_m$ is the m-th standard unit vector with appropriate dimension. Since A is HPD, T_m is tridiagonal and real.

Following the more general framework of the Lanczos method, we write $\boldsymbol{x}_m$ as

$$\boldsymbol{x}_m = \boldsymbol{x}_0 + V_m T_m^{-1} V_m^H \boldsymbol{r}_0 = \boldsymbol{x}_0 + V_m q_{m-1}(T_m) V_m^H \boldsymbol{r}_0,$$

where $q_{m-1} \in \Pi_{m-1}$ is the Hermite interpolating polynomial of $f(z) = z^{-1}$ at the eigenvalues of T_m. Indeed, $T_m^{-1} = q_{m-1}(T_m)$; see, e.g., [10, Ch. 1].

We also know that $\boldsymbol{x}_m = \boldsymbol{x}_0 + p(A)\boldsymbol{r}_0$, for some polynomial $p \in \Pi_{m-1}$. The fact that $\boldsymbol{x}_m$ has a unique representation in $\boldsymbol{x}_0 + \mathcal{K}_m(A, \boldsymbol{r}_0)$, plus the following lemma [13, Lemma 3.1], ensures that p and q_{m-1} are in fact the same.

*This version dated January 14, 2016.

†Fakultät für Mathematik und Naturwissenschaften, Bergische Universität Wuppertal, 42097 Wuppertal, Germany ({frommer,schweitzer}@math.uni-wuppertal.de).

‡Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, Pennsylvania 19122-6094, USA ({lund-nguyen,szyld}@math.temple.edu). Supported in part by the U.S. National Science Foundation grant DMS-1418882

LEMMA 1.1 (Lanczos polynomial relation). *For all* $q \in \Pi_{m-1}$,

$$V_m q(T_m) V_m^H \boldsymbol{r}_0 = q(A)\boldsymbol{r}_0. \tag{1.3}$$

Since A is HPD, we can introduce the following inner product $(\cdot,\cdot)$ on Π_m:

$$(p,q) = \langle q(A)\boldsymbol{r}_0, \bar{p}(A)\bar{\boldsymbol{r}}_0\rangle = (p(A)\boldsymbol{r}_0)^H q(A)\boldsymbol{r}_0.$$

Given the expansion $\boldsymbol{r}_0 = \sum_{i=1}^n \beta_i \boldsymbol{u}_i$ in terms of the unit eigenvectors $\boldsymbol{u}_i$ of A with corresponding eigenvalues λ_i, we can express this inner product as

$$(p,q) = \sum_{i=1}^n |\beta_i|^2 q(\lambda_i)\bar{p}(\lambda_i) =: \int_{\lambda_{\min}}^{\lambda_{\max}} q(z)\bar{p}(z)\ \mathrm{d}\alpha(z),$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalue of A, respectively; and the measure $\mathrm{d}\alpha$ is defined by the function $\alpha(z) = \sum_{i=1}^n |\beta_i|^2 H(z-\lambda_i)$, where H is the Heaviside function.

If p is a polynomial with $p(0) \neq 0$, then we denote $\widetilde{p} = \frac{1}{p(0)}p$ as its normalized variant, so that $\widetilde{p}(0) = 1$. We denote by p_m the sequence of orthogonal polynomials with respect to $(\cdot,\cdot)$. It is known that the zeros of p_m are the eigenvalues of T_m, as well as the nodes of the m-point Gauss quadrature rule with respect to $\mathrm{d}\alpha$ on $[\lambda_{\min}, \lambda_{\max}]$; see [7, 8]. These orthogonal polynomials are unique up to a scaling factor, and we call the corresponding normalized $\widetilde{p}_m$ the CG polynomials because of the following well-known result; see, e.g., [12, Ch. 8].

THEOREM 1.2. *The CG iterates* $\boldsymbol{x}_m$ *satisfy*
(i) $\boldsymbol{e}_m = \widetilde{p}_m(A)\boldsymbol{e}_0$, $\boldsymbol{r}_m = \widetilde{p}_m(A)\boldsymbol{r}_0$.
(ii) $\|\boldsymbol{e}_m\|_A = \min\{\|\boldsymbol{x}_* - \boldsymbol{x}\|_A : \boldsymbol{x} \in \boldsymbol{x}_0 + \mathcal{K}_m(A, \boldsymbol{r}_0)\}$.

1.2. Rank-one modifications. In [3], the authors consider a particular rank-one modification of the standard method to broaden the class of matrices for which convergence properties can be derived. Their modified method is based on the Arnoldi relation rather than the Lanczos relation (1.2), as it is defined and shown to converge for positive real matrices. We only show how the modification works for A HPD (which is also positive real), to serve as motivation for our new method.

Define $\widetilde{T}_m := T_m + \big(t_{m+1,m} T_m^{-1}\widehat{\boldsymbol{e}}_m\big)\widehat{\boldsymbol{e}}_m^H$. Also define a new iteration

$$\widetilde{\boldsymbol{x}}_m := \boldsymbol{x}_0 + V_m \widetilde{T}_m^{-1} V_m^H \boldsymbol{r}_0.$$

By the following lemma [15, Lemma 3], we can conclude, just as with CG, that

$$\widetilde{\boldsymbol{x}}_m = \boldsymbol{x}_0 + h(A)\boldsymbol{r}_0,$$

where $h \in \Pi_{m-1}$ is the Hermite interpolating polynomial of $f(z) = z^{-1}$ at the eigenvalues of $\widetilde{T}_m$, and $h(\widetilde{T}_m) = \widetilde{T}_m^{-1}$. The eigenvalues of $\widetilde{T}_m$ are termed the *harmonic Ritz values* of A. Then $\widetilde{x}_m \in \boldsymbol{x}_0 + \mathcal{K}_m(A, \boldsymbol{r}_0)$ as well, and as is shown in [3], $\widetilde{x}_m$ is in fact the GMRES approximation to $A\boldsymbol{x} = \boldsymbol{b}$.

The following lemma also shows that there are further rank-one modifications for which (1.3) holds.

LEMMA 1.3. *Let* $\boldsymbol{u} \in \mathbb{C}^m$ *be a nonzero vector. Denote* $\widehat{T}_m := T_m + \boldsymbol{u}\widehat{\boldsymbol{e}}_m^H$. *Then for any* $q \in \Pi_{m-1}$,

$$V_m q\big(\widehat{T}_m\big) V_m^H \boldsymbol{r}_0 = q(A)\boldsymbol{r}_0. \tag{1.4}$$

It is worth mentioning that the only such modifications for which (1.4) can be preserved must be rank-one with nonzero entries only in the last column, as stated in the following lemma. The proof this lemma can be found in the forthcoming, longer version of this paper [5].

LEMMA 1.4. *Let* $M \in \mathbb{C}^{m\times m}$, *and denote* $\widehat{T}_m := T_m + M$. *If for all* $q \in \Pi_{m-1}$, $V_m q\big(\widehat{T}_m\big) V_m^H \boldsymbol{r}_0 = q(A)\boldsymbol{r}_0$, *then there exists* $\boldsymbol{u} \in \mathbb{C}^m$ *such that* $M = \boldsymbol{u}\widehat{\boldsymbol{e}}_m^H$.

We are therefore motivated to look for such rank-one modifications that may lead to improved convergence properties.

2. The Radau-Lanczos method for linear systems. In section 1, we saw that the standard Lanczos method for a HPD matrix is related to the CG polynomials, as well as an m-point Gauss quadrature rule with respect to the measure $\mathrm{d}\alpha$ and with nodes at the eigenvalues of T_m. In this section, we show how a particular $m+1$-point Gauss-Radau quadrature rule for a modified measure is related to a rank-one update of the tridiagonal matrix T_{m+1}.

In an m-point Gauss quadrature rule, the quadrature nodes are determined so that the rule is exact for polynomials up to degree $2m-1$. A Gauss-Radau quadrature rule is a Gauss rule in which one node is fixed. We fix $\theta_0 > \lambda_{\max}$, and consider the $m+1$-point Gauss-Radau rule on the interval $[\lambda_{\min}, \lambda_{\max}]$ for a new measure $\mathrm{d}\alpha_{\mathrm{R}}$ defined as $\mathrm{d}\alpha_{\mathrm{R}}(t) = (\theta_0 - t)\,\mathrm{d}\alpha(t)$. Following the work of [7, 8], we then seek a matrix related to T_{m+1} whose eigenvalues are the nodes of this rule.

First we write T_m explicitly as

$$T_m = \begin{bmatrix} \omega_1 & \gamma_1 & & & \\ \gamma_1 & \omega_2 & \gamma_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_{m-2} & \omega_{m-1} & \gamma_{m-1} \\ & & & \gamma_{m-1} & \omega_m \end{bmatrix}.$$

Then we solve for $\boldsymbol{d} \in \mathbb{C}^m$ satisfying $(T_m - \theta_0 I)\boldsymbol{d} = \gamma_m^2 \widehat{\boldsymbol{e}}_m$ and define

$$T_{m+1}^{\mathrm{R}} := \begin{bmatrix} T_m & \gamma_m \widehat{\boldsymbol{e}}_m \\ \gamma_m \widehat{\boldsymbol{e}}_m^H & d_m \end{bmatrix},$$

where d_m is the m-th component of $\boldsymbol{d}$. Note that T_{m+1}^{R} can also be expressed as a rank-one modification of T_{m+1}, satisfying Lemma 1.3: $T_{m+1}^{\mathrm{R}} = T_{m+1} + (d_m - \omega_{m+1})\widehat{\boldsymbol{e}}_{m+1}\widehat{\boldsymbol{e}}_{m+1}^H$. Furthermore, the eigenvalues of T_{m+1}^{R} are in fact the nodes of our $m+1$-point Gauss-Radau rule, with one node fixed at θ_0. We denote the eigenvalues of T_{m+1}^{R} different from θ_0 by θ_i^{R}, $i = 1, \ldots, m$.

As with CG, there is a connection to a particular set of orthogonal polynomials, given the appropriate inner product. According to Gautschi [6], this inner product is

$$(p, q)_{\mathrm{R}} = \sum_{i=1}^{n} |\beta_i|^2 (\theta_0 - \lambda_i) q(\lambda_i)\bar{p}(\lambda_i) =: \int_{\lambda_{\min}}^{\lambda_{\max}} q(z)\bar{p}(z)\,\mathrm{d}\alpha_{\mathrm{R}}(z), \tag{2.1}$$

which we refer to as the *Radau inner product*, with $\mathrm{d}\alpha_{\mathrm{R}}$ as the *Radau measure* defined earlier. We let p_m^{R} denote the polynomials orthogonal with respect to this inner product, whose roots are θ_i^{R}, $i = 1, \ldots, m$. Note that θ_0 is not a root of p_m^{R} for any m.

We finally define the $m+1$-st Radau-Lanczos approximation as

$$\boldsymbol{x}_{m+1}^{\mathrm{R}} := \boldsymbol{x}_0 + V_{m+1}\big(T_{m+1}^{\mathrm{R}}\big)^{-1} V_{m+1}^H \boldsymbol{r}_0.$$

As in section 1, we now explore polynomial relations amongst the approximations $\boldsymbol{x}_{m+1}^{\mathrm{R}}$, the errors $\boldsymbol{e}_{m+1}^{\mathrm{R}}$, and the residuals $\boldsymbol{r}_{m+1}^{\mathrm{R}}$. Let q_m^{R} denote the Hermite interpolating polynomial of degree m through $f(z) = z^{-1}$ at the eigenvalues of T_{m+1}^{R}. By Lemma 1.3,

$$\boldsymbol{x}_{m+1}^{\mathrm{R}} = \boldsymbol{x}_0 + V_{m+1} q_m^{\mathrm{R}}\big(T_{m+1}^{\mathrm{R}}\big) V_{m+1}^H \boldsymbol{r}_0 = \boldsymbol{x}_0 + q_m^{\mathrm{R}}(A)\boldsymbol{r}_0.$$

Then

$$\boldsymbol{e}_{m+1}^{\mathrm{R}} = \boldsymbol{e}_0 - q_m^{\mathrm{R}}(A)\boldsymbol{r}_0 = \boldsymbol{e}_0 - A q_m^{\mathrm{R}}(A)\boldsymbol{e}_0 = \pi_{m+1}^{\mathrm{R}}(A)\boldsymbol{e}_0,$$

where $\pi_{m+1}^{\mathrm{R}}(z) = 1 - z q_m^{\mathrm{R}}(z)$ and $\pi_{m+1}^{\mathrm{R}} \in \Pi_{m+1}$. Consequently, $\boldsymbol{r}_{m+1}^{\mathrm{R}} = \pi_{m+1}^{\mathrm{R}}(A)\boldsymbol{r}_0$. Note that

$$\pi_{m+1}^{\mathrm{R}}(z) = \frac{1}{\theta_0}(\theta_0 - z)\widetilde{p}_m^{\mathrm{R}}(z), \quad \text{where} \quad \widetilde{p}_m^{\mathrm{R}}(z) = \frac{1}{p_m^{\mathrm{R}}(0)} p_m^{\mathrm{R}}(z), \tag{2.2}$$

since the roots of π_{m+1}^{R} are the eigenvalues of T_{m+1}^{R}, and $\pi_{m+1}^{\mathrm{R}}(0) = 1$.

2.1. Variational characterization. We now derive a variational characterization of the Radau-Lanczos method. We begin with a useful orthogonality property.

LEMMA 2.1. *For any* $q \in \Pi_{m-1}$, $\langle \boldsymbol{e}_0 - q_m^{\mathrm{R}}(A)\boldsymbol{r}_0, q(A)\boldsymbol{r}_0\rangle_A = 0$.

Proof. The proof follows from the definition (2.1) and the polynomial equivalence (2.2):

$$\begin{aligned} 0 = (\bar{q}, {\theta_0}^{-1}\widetilde{p}_m^{\mathrm{R}})_{\mathrm{R}} &= \sum_{i=1}^{n} {\theta_0}^{-1}\widetilde{p}_m^{\mathrm{R}}(\lambda_i)q(\lambda_i)(\theta_0 - \lambda_i)\,|\beta_i|^2 \\ &= \langle (\theta_0 I - A){\theta_0}^{-1}\widetilde{p}_m^{\mathrm{R}}(A)\boldsymbol{r}_0, q(A)\boldsymbol{r}_0\rangle_2 = \langle \pi_{m+1}^{\mathrm{R}}(A)\boldsymbol{r}_0, q(A)\boldsymbol{r}_0\rangle_2 \\ &= \langle A(\boldsymbol{e}_0 - q_m^{\mathrm{R}}(A)\boldsymbol{r}_0), q(A)\boldsymbol{r}_0\rangle_2 = \langle \boldsymbol{e}_0 - q_m^{\mathrm{R}}(A)\boldsymbol{r}_0, q(A)\boldsymbol{r}_0\rangle_A. \quad \square \end{aligned}$$

By definition, q_m^{R} interpolates z^{-1} at the eigenvalues of T_{m+1}^{R}, and in particular, at θ_0. In fact, for every m, q_m^{R} interpolates z^{-1} at θ_0. Then for some $s_{m-1}^{\mathrm{R}} \in \Pi_{m-1}$

$$q_m^{\mathrm{R}}(z) = (\theta_0 - z)s_{m-1}^{\mathrm{R}}(z) + \frac{1}{\theta_0}, \tag{2.3}$$

for all m. We therefore regard q_m^{R}, a polynomial of degree m, as being completely determined by the $m-1$-degree polynomial s_{m-1}. It is precisely this fact that leads to the following variational characterization of the Radau-Lanczos method.

THEOREM 2.2. *The approximation* $\boldsymbol{x}_{m+1}^{\mathrm{R}}$ *is such that the error* $\boldsymbol{e}_{m+1}^{\mathrm{R}} = \boldsymbol{x}_* - \boldsymbol{x}_{m+1}^{\mathrm{R}}$ *satisfies*

$$\left\|\boldsymbol{e}_{m+1}^{\mathrm{R}}\right\|_{A(\theta_0 I - A)^{-1}} = \min_{\boldsymbol{y}\in(\theta_0 I - A)\mathcal{K}_m(A,\boldsymbol{r}_0)} \left\|\boldsymbol{e}_0 - \frac{1}{\theta_0}\boldsymbol{r}_0 - \boldsymbol{y}\right\|_{A(\theta_0 I - A)^{-1}}.$$

Proof. By Lemma 2.1 and (2.3), we have for all $q \in \Pi_{m-1}$ that

$$\begin{aligned} 0 &= \langle \boldsymbol{e}_0 - q_m^{\mathrm{R}}(A)\boldsymbol{r}_0, q(A)\boldsymbol{r}_0\rangle_A = \langle \boldsymbol{e}_0 - \frac{1}{\theta_0}\boldsymbol{r}_0 - (\theta_0 I - A)s_{m-1}^{\mathrm{R}}(A)\boldsymbol{r}_0, q(A)\boldsymbol{r}_0\rangle_A \\ &= \langle \boldsymbol{e}_0 - \frac{1}{\theta_0}\boldsymbol{r}_0 - (\theta_0 I - A)s_{m-1}^{\mathrm{R}}(A)\boldsymbol{r}_0, (\theta_0 I - A)q(A)\boldsymbol{r}_0\rangle_{A(\theta_0 I - A)^{-1}}. \end{aligned}$$

Since $(\theta_0 I - A)s_{m-1}^{\mathrm{R}}(A)\boldsymbol{r}_0 \in (\theta_0 I - A)\mathcal{K}_m(A, \boldsymbol{r}_0)$, and since $(\theta_0 I - A)q(A)\boldsymbol{r}_0$ describes all the elements of $(\theta_0 I - A)\mathcal{K}_m(A, \boldsymbol{r}_0)$, the desired minimization property holds. $\square$

COROLLARY 2.3. *The error* $\boldsymbol{e}_{m+1}^{\mathrm{R}}$ *further satisfies*

$$\left\|\boldsymbol{e}_{m+1}^{\mathrm{R}}\right\|_{A(\theta_0 I - A)^{-1}} = \frac{1}{\theta_0} \min_{\substack{p(0)=1 \\ p\in\Pi_m}} \left\|(\theta_0 I - A)p(A)\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}}.$$

Proof. By Theorem 2.2, we have

$$\begin{aligned} \left\|\boldsymbol{e}_{m+1}^{\mathrm{R}}\right\|_{A(\theta_0 I - A)^{-1}} &= \left\|\boldsymbol{e}_0 - \frac{1}{\theta_0}\boldsymbol{r}_0 - (\theta_0 I - A)s_{m-1}^{\mathrm{R}}(A)\boldsymbol{r}_0\right\|_{A(\theta_0 I - A)^{-1}} \\ &\le \left\|\boldsymbol{e}_0 - \frac{1}{\theta_0}\boldsymbol{r}_0 - (\theta_0 I - A)q(A)\boldsymbol{r}_0\right\|_{A(\theta_0 I - A)^{-1}}, \end{aligned}$$

for all $q \in \Pi_{m-1}$. Note that

$$\begin{aligned} \boldsymbol{e}_0 - \frac{1}{\theta_0}\boldsymbol{r}_0 - (\theta_0 I - A)q(A)\boldsymbol{r}_0 &= \left(I - \frac{1}{\theta_0}A - (\theta_0 I - A)q(A)A\right)\boldsymbol{e}_0 \\ &= \frac{1}{\theta_0}(\theta_0 I - A)\left(I - \theta_0 A q(A)\right)\boldsymbol{e}_0 \\ &= \frac{1}{\theta_0}(\theta_0 I - A)p(A)\boldsymbol{e}_0, \end{aligned}$$

where $p(z)$ is any polynomial in Π_m with $p(0) = 1$. Then, in fact,

$$\left\|\boldsymbol{e}_{m+1}^{\mathrm{R}}\right\|_{A(\theta_0 I - A)^{-1}} \leq \frac{1}{\theta_0} \left\|(\theta_0 I - A)p(A)\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}},$$

for all $p \in \Pi_m$ with $p(0) = 1$. In particular, from (2.2) we have that

$$\boldsymbol{e}_{m+1}^{\mathrm{R}} = \frac{1}{\theta_0}(\theta_0 I - A)\widetilde{p}_m^{\mathrm{R}}(A)\boldsymbol{e}_0,$$

so the minimum is attained. ∎

2.2. Convergence for linear systems. We can further show that the Radau-Lanczos method converges and provide a convergence bound related to that of CG. Define the following quantities:

$$\kappa := \frac{\lambda_{\max}}{\lambda_{\min}}, \quad c := \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}, \quad \text{and} \quad \xi_m := \frac{1}{\cosh(m \ln c)}. \tag{2.4}$$

If $\kappa = 1$, then we set $\xi_m = 0$.

THEOREM 2.4. *The error for Radau-Lanczos approximation $\boldsymbol{x}_{m+1}^{\mathrm{R}}$ can be bounded as*

$$\left\|\boldsymbol{e}_{m+1}^{\mathrm{R}}\right\|_{A(\theta_0 I - A)^{-1}} \leq \left(1 - \frac{\lambda_{\min}}{\theta_0}\right) \xi_m \left\|\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}} \leq 2\left(1 - \frac{\lambda_{\min}}{\theta_0}\right) c^m \left\|\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}}.$$

Proof. Since A is HPD, $A^{\frac{1}{2}}(\theta_0 I - A)^{-\frac{1}{2}}$ commutes with $(\theta_0 I - A)p(A)$, and we have

$$\left\|(\theta_0 I - A)p(A)\right\|_{A(\theta_0 I - A)^{-1}} = \left\|A^{\frac{1}{2}}(\theta_0 I - A)^{-\frac{1}{2}}(\theta_0 I - A)p(A)A^{-\frac{1}{2}}(\theta_0 I - A)^{\frac{1}{2}}\right\|_2 = \left\|(\theta_0 I - A)p(A)\right\|_2.$$

Then along with Corollary 2.3, we have the following:

$$\begin{aligned}
\left\|\boldsymbol{e}_{m+1}^{\mathrm{R}}\right\|_{A(\theta_0 I - A)^{-1}} &= \frac{1}{\theta_0} \min_{\substack{p(0)=1 \\ p \in \Pi_m}} \left\|(\theta_0 I - A)p(A)\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}} \\
&\leq \frac{1}{\theta_0} \min_{\substack{p(0)=1 \\ p \in \Pi_m}} \left\|(\theta_0 I - A)p(A)\right\|_{A(\theta_0 I - A)^{-1}} \left\|\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}} \\
&= \frac{1}{\theta_0} \min_{\substack{p(0)=1 \\ p \in \Pi_m}} \left\|(\theta_0 I - A)p(A)\right\|_2 \left\|\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}} \\
&\leq \frac{1}{\theta_0} \min_{\substack{p(0)=1 \\ p \in \Pi_m}} \max_{[\lambda_{\min}, \lambda_{\max}]} (\theta_0 - \lambda) \left|p(\lambda)\right| \left\|\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}} \\
&\leq \frac{\theta_0 - \lambda_{\min}}{\theta_0} \min_{\substack{p(0)=1 \\ p \in \Pi_m}} \max_{[\lambda_{\min}, \lambda_{\max}]} \left|p(\lambda)\right| \left\|\boldsymbol{e}_0\right\|_{A(\theta_0 I - A)^{-1}}.
\end{aligned} \tag{2.5}$$

By a classical result on Chebyshev polynomials (see, e.g., [14, Section 6.11]) and the relation $\cosh^{-1}(z) = \ln\left(z + \sqrt{(z+1)(z-1)}\right)$, we have that

$$\min_{\substack{p(0)=1 \\ p \in \Pi_m}} \max_{[\lambda_{\min}, \lambda_{\max}]} \left|p(\lambda)\right| \leq \xi_m \leq 2c^m. \tag{2.6}$$

Combining (2.5) and (2.6), we arrive at the desired conclusion. ∎

Convergence follows since ξ_m goes to 0 as m increases.

3. The Radau-Lanczos method for Stieltjes functions of matrices. Let f be a Stieltjes function of the form

$$f(z) = \int_0^\infty \frac{1}{t+z}\ \mathrm{d}\mu(t), \tag{3.1}$$

where μ is monotonically increasing and nonnegative on $[0,\infty)$ with the property $\int_0^\infty \frac{1}{t+1}\ \mathrm{d}\mu(t) < \infty$. Define the Radau-Lanczos approximation to $f(A)\boldsymbol{b}$ as

$$\boldsymbol{f}_{m+1}^{\mathrm{R}} := V_{m+1} f(T_{m+1}^{\mathrm{R}}) V_{m+1}^H \boldsymbol{b},$$

where the columns of V_{m+1} form the Lanczos basis of $\mathcal{K}_{m+1}(A,\boldsymbol{b})$. Note that $f(T_{m+1}^{\mathrm{R}}) = q_*(T_{m+1}^{\mathrm{R}})$, where $q_* \in \Pi_m$ is the Hermite interpolating polynomial of f at the eigenvalues of T_{m+1}^{R}. By Lemma 1.3, $\boldsymbol{f}_{m+1}^{\mathrm{R}} = q_*(A)\boldsymbol{b} \in \mathcal{K}_{m+1}(A,\boldsymbol{b})$.

To show that $\boldsymbol{f}_{m+1}^{\mathrm{R}}$ converges to $f(A)\boldsymbol{b}$, we will take advantage of the integral form (3.1) of f in the expressions of $f(A)\boldsymbol{b}$ and $\boldsymbol{f}_{m+1}^{\mathrm{R}}$,

$$f(A)\boldsymbol{b} = \int_0^\infty (A+tI)^{-1}\boldsymbol{b}\ \mathrm{d}\mu(t) \quad \text{and} \quad \boldsymbol{f}_{m+1}^{\mathrm{R}} = \int_0^\infty V_{m+1}(T_{m+1}^{\mathrm{R}}+tI)^{-1}V_{m+1}^H\boldsymbol{b}\ \mathrm{d}\mu(t). \tag{3.2}$$

We therefore need shifted versions of the results in section 2.

LEMMA 3.1. *Let $\widehat{T}_m$ and $\boldsymbol{r}_0$ be as in Lemma 1.3. Then for all $t \in \mathbb{C}$ and for all $q \in \Pi_{m-1}$,*

$$V_m q(\widehat{T}_m + tI) V_m^H \boldsymbol{r}_0 = q(A+tI)\boldsymbol{r}_0. \tag{3.3}$$

Proof. It is easily verified that the Lanczos relation (1.2) is shift invariant, i.e., $AV_m = V_m T_m + t_{m+1,m}\boldsymbol{v}_{m+1}\widehat{\boldsymbol{e}}_m^H$ implies $(A+tI)V_m = V_m(T_m+tI) + t_{m+1,m}\boldsymbol{v}_{m+1}\widehat{\boldsymbol{e}}_m^H$, so that the columns of V_m are also a basis of $\mathcal{K}(A+tI,\boldsymbol{r}_0)$. Applying Lemma 1.3 to $A+tI$ and $\widehat{T}_m+tI$ gives the desired result. ∎

We also define the following shifted quantities for $t \geq 0$:

$$\begin{aligned}
\boldsymbol{x}_*(t) &:= (A+tI)^{-1}\boldsymbol{b} \\
\boldsymbol{x}_{m+1}^{\mathrm{R}}(t) &:= V_{m+1}(T_{m+1}^{\mathrm{R}}+tI)^{-1}V_{m+1}^H\boldsymbol{b} \\
\boldsymbol{e}_{m+1}^{\mathrm{R}}(t) &:= \boldsymbol{x}_*(t) - \boldsymbol{x}_{m+1}^{\mathrm{R}}(t) \\
\boldsymbol{r}_{m+1}^{\mathrm{R}}(t) &:= (A+tI)\boldsymbol{e}_{m+1}^{\mathrm{R}}(t).
\end{aligned} \tag{3.4}$$

Note that $\boldsymbol{x}_{m+1}^{\mathrm{R}}(t)$ is *not* the $m+1$-st Radau-Lanczos approximation to the shifted system $A+tI$. However, the shifted residuals are collinear with $\boldsymbol{r}_{m+1}^{\mathrm{R}}(0)$.

LEMMA 3.2. *Assume that the initial shifted residuals are all collinear with $\boldsymbol{r}_0^{\mathrm{R}}(0)$, i.e., $\boldsymbol{r}_0^{\mathrm{R}}(t) = \rho_0(t)\boldsymbol{r}_0^{\mathrm{R}}(0)$ with $\rho_0(t) \in \mathbb{C}$. Then*

(i) the shifted residuals $\boldsymbol{r}_{m+1}^{\mathrm{R}}(t)$ are collinear with $\boldsymbol{r}_{m+1}^{\mathrm{R}}(0)$ in the following sense:

$$\boldsymbol{r}_{m+1}^{\mathrm{R}}(t) = \rho_{m+1}(t)\boldsymbol{r}_{m+1}^{\mathrm{R}}(0),$$

where $\rho_{m+1}(t) := \frac{\rho_0(t)}{\pi_{m+1}^{\mathrm{R}}(-t)}$; and

(ii) $\rho_{m+1}(t) \leq \rho_0(t)$.

Proof. To show (i) we first apply Lemma 3.1 and relations from section 2 to $A+tI$ and $T_{m+1}^{\mathrm{R}}+tI$ to obtain that

$$\boldsymbol{r}_{m+1}^{\mathrm{R}}(t) = \pi_{m+1,t}^{\mathrm{R}}(A+tI)\boldsymbol{r}_0^{\mathrm{R}}(t),$$

where $\pi_{m+1,t}^{\mathrm{R}}(z) = 1 - zq_{m,t}^{\mathrm{R}}(z)$, and $q_{m,t}^{\mathrm{R}} \in \Pi_m$ interpolates z^{-1} at the eigenvalues of $T_{m+1}^{\mathrm{R}}+tI$, which are θ_0+t and $\theta_i^{\mathrm{R}}+t$, $i=1,\ldots,m$. We can write $\pi_{m+1,t}^{\mathrm{R}}$ explicitly as

$$\pi_{m+1,t}^{\mathrm{R}}(z) = \frac{(\theta_0+t-z)\prod_{i=1}^m(\theta_i^{\mathrm{R}}+t-z)}{(\theta_0+t)\prod_{i=1}^m(\theta_i^{\mathrm{R}}+t)}.$$

Then

$$\pi^{\mathrm{R}}_{m+1,0}(-t) = \frac{(\theta_0+t)\prod_{i=1}^{m}(\theta_i^{\mathrm{R}}+t)}{\theta_0\prod_{i=1}^{m}\theta_i^{\mathrm{R}}} \text{ and } \pi^{\mathrm{R}}_{m+1,0}(z-t) = \frac{(\theta_0+t-z)\prod_{i=1}^{m}(\theta_i^{\mathrm{R}}+t-z)}{\theta_0\prod_{i=1}^{m}\theta_i^{\mathrm{R}}},$$

implying

$$\pi^{\mathrm{R}}_{m+1,t}(z) = \frac{\pi^{\mathrm{R}}_{m+1,0}(z-t)}{\pi^{\mathrm{R}}_{m+1,0}(-t)} = \frac{\pi^{\mathrm{R}}_{m+1}(z-t)}{\pi^{\mathrm{R}}_{m+1}(-t)}.$$

Therefore,

$$\boldsymbol{r}^{\mathrm{R}}_{m+1}(t) = \pi^{\mathrm{R}}_{m+1,t}(A+tI)\boldsymbol{r}^{\mathrm{R}}_0(t) = \frac{1}{\pi^{\mathrm{R}}_{m+1}(-t)}\pi^{\mathrm{R}}_{m+1}(A)\boldsymbol{r}^{\mathrm{R}}_0(t) = \frac{\rho_0(t)}{\pi^{\mathrm{R}}_{m+1}(-t)}\boldsymbol{r}^{\mathrm{R}}_{m+1}(0),$$

where the last equality holds by the collinearity assumption $\boldsymbol{r}^{\mathrm{R}}_0(t) = \rho_0(t)\boldsymbol{r}^{\mathrm{R}}_0(0)$ and by the equality $\pi^{\mathrm{R}}_{m+1}(A)\boldsymbol{r}^{\mathrm{R}}_0(0) = \boldsymbol{r}^{\mathrm{R}}_{m+1}(0)$.

As for part (ii), since $t \geq 0$ we can easily deduce the following bound:

$$\rho_m(t) = \frac{\rho_0(t)}{\pi^{\mathrm{R}}_{m+1}(-t)} = \frac{\theta_0\prod_{i=1}^{m}\theta_i^{\mathrm{R}}}{(\theta_0+t)\prod_{i=1}^{m}(\theta_i^{\mathrm{R}}+t)} \leq \rho_0(t). \quad \square$$

At this point, we have all the necessary tools to show that the Radau-Lanczos method converges for Stieltjes functions of HPD matrices. Note that the norm for this convergence bound is the $A^{-1}(\theta_0 I - A)^{-1}$-norm, which is different from the norm used for the bounds in section 2.

THEOREM 3.3. *The following error bound holds for the Radau-Lanczos method:*

$$\left\|f(A)\boldsymbol{b} - \boldsymbol{f}^{\mathrm{R}}_{m+1}\right\|_{A^{-1}(\theta_0 I-A)^{-1}} \leq C\left(1 - \frac{\lambda_{\min}}{\theta_0}\right)\xi_m \leq 2C\left(1 - \frac{\lambda_{\min}}{\theta_0}\right)c^m,$$

where c and ξ_m are as in (2.4), and $C = \frac{1}{\sqrt{\lambda_{\min}(\theta_0-\lambda_{\max})}}\|\boldsymbol{b}\|_2 f(\lambda_{\min})$.

Proof. We begin by looking at the integral expression for the error (3.2):

$$f(A)\boldsymbol{b} - \boldsymbol{f}^{\mathrm{R}}_{m+1} = \int_0^\infty \left[\boldsymbol{x}_*(t) - \boldsymbol{x}^{\mathrm{R}}_{m+1}(t)\right]\ \mathrm{d}\mu(t) = \int_0^\infty \boldsymbol{e}^{\mathrm{R}}_{m+1}(t)\ \mathrm{d}\mu(t).$$

Applying Lemma 3.2 to the shifted residuals $\boldsymbol{r}^{\mathrm{R}}_{m+1}(t)$, we obtain the following:

$$\begin{aligned}
\left\|f(A)\boldsymbol{b} - \boldsymbol{f}^{\mathrm{R}}_{m+1}\right\|_{A^{-1}(\theta_0 I-A)^{-1}} &\leq \int_0^\infty \left\|\boldsymbol{e}^{\mathrm{R}}_{m+1}(t)\right\|_{A^{-1}(\theta_0 I-A)^{-1}}\ \mathrm{d}\mu(t) \\
&= \int_0^\infty \left\|(A+tI)^{-1}\boldsymbol{r}^{\mathrm{R}}_{m+1}(t)\right\|_{A^{-1}(\theta_0 I-A)^{-1}}\ \mathrm{d}\mu(t) \\
&= \int_0^\infty \left\|(A+tI)^{-1}\rho_{m+1}(t)\boldsymbol{r}^{\mathrm{R}}_{m+1}(0)\right\|_{A^{-1}(\theta_0 I-A)^{-1}}\ \mathrm{d}\mu(t) \\
&\leq \int_0^\infty \left\|(A+tI)^{-1}\rho_0(t)\boldsymbol{r}^{\mathrm{R}}_{m+1}(0)\right\|_{A^{-1}(\theta_0 I-A)^{-1}}\ \mathrm{d}\mu(t) \\
&= \int_0^\infty |\rho_0(t)|\left\|(A+tI)^{-1}\boldsymbol{r}^{\mathrm{R}}_{m+1}(0)\right\|_{A^{-1}(\theta_0 I-A)^{-1}}\ \mathrm{d}\mu(t). \qquad (3.5)
\end{aligned}$$

Since A, its inverse, and the shifted matrices $A+tI$, $t>0$, are all HPD, we have the following relation:

$$\begin{aligned}
\left\|(A+tI)^{-1}\boldsymbol{r}^{\mathrm{R}}_{m+1}(0)\right\|^2_{A^{-1}(\theta_0 I-A)^{-1}} &= \langle (A+tI)^{-1}A\boldsymbol{e}^{\mathrm{R}}_{m+1}(0), A^{-1}(\theta_0 I-A)^{-1}(A+tI)^{-1}A\boldsymbol{e}^{\mathrm{R}}_{m+1}(0)\rangle \\
&\leq \left(\frac{1}{\lambda_{\min}+t}\right)^2 \left\|\boldsymbol{e}^{\mathrm{R}}_{m+1}(0)\right\|^2_{A(\theta_0 I-A)^{-1}}. \qquad (3.6)
\end{aligned}$$

Then by applying (3.6) to the integrand of (3.5) and by Theorem 2.4, we have

$$\begin{aligned}\left\|f(A)\boldsymbol{b}-\boldsymbol{f}_{m+1}^{\mathrm{R}}\right\|_{A^{-1}(\theta_0 I-A)^{-1}} &\leq \int_0^\infty \frac{|\rho_0(t)|}{\lambda_{\min}+t}\left\|\boldsymbol{e}_{m+1}^{\mathrm{R}}(0)\right\|_{A(\theta_0 I-A)^{-1}}\ \mathrm{d}\mu(t)\\ &\leq \left(1-\frac{\lambda_{\min}}{\theta_0}\right)\xi_m\left\|\boldsymbol{e}_0^{\mathrm{R}}(0)\right\|_{A(\theta_0 I-A)^{-1}}\int_0^\infty \frac{|\rho_0(t)|}{\lambda_{\min}+t}\ \mathrm{d}\mu(t), \qquad (3.7)\end{aligned}$$

Furthermore,

$$\begin{aligned}\left\|\boldsymbol{e}_0^{\mathrm{R}}(0)\right\|_{A(\theta_0 I-A)^{-1}}^2 &= \langle \boldsymbol{e}_0^{\mathrm{R}}(0), A(\theta_0 I-A)^{-1}\boldsymbol{e}_0^{\mathrm{R}}(0)\rangle\\ &= \langle \boldsymbol{r}_0^{\mathrm{R}}(0), A^{-1}(\theta_0 I-A)^{-1}\boldsymbol{r}_0^{\mathrm{R}}(0)\rangle\\ &\leq \frac{1}{\lambda_{\min}(\theta_0-\lambda_{\max})}\left\|\boldsymbol{r}_0^{\mathrm{R}}(0)\right\|_2^2. \qquad (3.8)\end{aligned}$$

Finally, by the definitions in (3.4), we implicitly have that $\boldsymbol{x}_0^{\mathrm{R}}(t)=\boldsymbol{x}_0^{\mathrm{R}}(0)=0$, thus implying that $\boldsymbol{r}_0^{\mathrm{R}}(t)=\boldsymbol{r}_0^{\mathrm{R}}(0)=\boldsymbol{b}$; therefore $\rho_0(t)=1$. Combining this with (3.7) and (3.8), we then obtain the desired bound. ∎

3.1. The restarted Radau-Lanczos method. In many practical situations where one wants to approximate $f(A)\boldsymbol{b}$, the available storage limits the number of Lanczos iterations that can be performed, as one needs to store the entire basis V_m in order to form $\boldsymbol{f}_m^{\mathrm{R}}$. Therefore, *restarts* are of vital importance in this setting; see, e.g., [1, 2, 3, 4, 11]. The idea is as follows: after a (small) number m of Lanczos steps, one forms a first approximation $\boldsymbol{f}_m^{(1)}$ for $f(A)\boldsymbol{b}$. If this approximation is not accurate enough, one uses m further Lanczos steps to obtain an approximation $\boldsymbol{a}_m^{(1)}$ for the error $f(A)\boldsymbol{b}-\boldsymbol{f}_m^{(1)}$, which is then used as an additive correction to form $\boldsymbol{f}_m^{(2)}=\boldsymbol{f}_m^{(1)}+\boldsymbol{a}_m^{(1)}$. Continuing like so, we obtain the sequences $\boldsymbol{f}_m^{(k)}$ and $\boldsymbol{a}_m^{(k)}$, where k denotes the index of the restarted cycle, and m the length of the cycle.

We further note that, for every cycle, one must use a suitable representation of the error as the action of a new matrix function $e^{(k)}(A)$ on a vector [2, 4]. Ultimately, it is crucial to find a representation for $e^{(k)}(z)$ that can be evaluated in a numerically stable manner. Following [3, 4] we represent $e^{(k)}(z)$ as a Stieltjes function with a new measure $\mathrm{d}\mu^{(k)}$ for the Radau-Lanczos method, allowing us to prove the following theorem, which shows that a restarted variant of the Radau-Lanczos method is also convergent. Details will be given in [5].

THEOREM 3.4. *Let k be the number of restart cycles, and $m+1$ the length of each cycle. Let $\boldsymbol{f}_{m+1}^{(k)}$ denote the restarted Radau-Lanczos approximation after k cycles. (We drop the superscript* R *for ease of notation.) Then*

$$\left\|f(A)\boldsymbol{b}-\boldsymbol{f}_{m+1}^{(k)}\right\|_{A^{-1}(\theta_0 I-A)^{-1}} \leq C\left(1-\frac{\lambda_{\min}}{\theta_0}\right)^k \xi_m^k,$$

where $C=\frac{1}{\sqrt{\lambda_{\min}(\theta_0-\lambda_{\max})}}\left\|\boldsymbol{b}\right\|_2 f(\lambda_{\min})$.

Proof. We only give a sketch. As with $\boldsymbol{f}_{m+1}^{(k)}$, let the superscript (k) denote all the corresponding restarted quantities. Following the proof of Theorem 3.3, we again note that

$$\left\|f(A)\boldsymbol{b}-\boldsymbol{f}_{m+1}^{(k)}\right\|_{A^{-1}(\theta_0 I-A)^{-1}} = \int_0^\infty \left\|(A+tI)^{-1}\boldsymbol{r}_{m+1}^{(k)}(t)\right\|_{A^{-1}(\theta_0 I-A)^{-1}}\ \mathrm{d}\mu(t).$$

Assuming the restarted initial residuals are collinear, i.e., $\boldsymbol{r}_0^{(k)}(t)=\rho_0^{(k)}(t)\boldsymbol{r}_0^{(k)}(0)$, then by Lemma 3.2(i), we have that $\boldsymbol{r}_{m+1}^{(k)}(t)=\rho_{m+1}^{(k)}(t)\boldsymbol{r}_{m+1}^{(k)}(0)$.

Furthermore,

$$\begin{aligned}\left\|(A+tI)^{-1}\boldsymbol{r}_{m+1}^{(k)}(0)\right\|_{A^{-1}(\theta_0 I-A)^{-1}} &\leq \frac{1}{\lambda_{\min}+t}\left\|\boldsymbol{e}_{m+1}^{(k)}(0)\right\|_{A(\theta_0 I-A)^{-1}}\\ &\leq \frac{1}{\lambda_{\min}+t}\left(1-\frac{\lambda_{\min}}{\theta_0}\right)\xi_m\left\|\boldsymbol{e}_0^{(k)}(0)\right\|_{A(\theta_0 I-A)^{-1}},\end{aligned}$$

where the last inequality holds by Theorem 2.4 for a particular restart cycle k.

Since $\boldsymbol{e}_0^{(k)}(0) = \boldsymbol{e}_{m+1}^{(k-1)}(0)$, we can apply Theorem 2.4 and Lemma 3.2(i) inductively to obtain that

$$\left\|\boldsymbol{e}_0^{(k)}(0)\right\|_{A(\theta_0 I - A)^{-1}} \leq \left(1 - \frac{\lambda_{\min}}{\theta_0}\right)^{k-1} \xi_m^{k-1} |\rho_{m+1}^{(1)}(t) \cdots \rho_{m+1}^{(k-1)}(t)| \left\|\boldsymbol{e}_0^{(1)}(0)\right\|_{A(\theta_0 I - A)^{-1}}.$$

Further note that $\rho_{m+1}^{(i)} \equiv \rho_0^{(i+1)}$; therefore, by repeated application of Lemma 3.2(ii),

$$|\rho_{m+1}^{(1)}(t) \cdots \rho_{m+1}^{(k-1)}(t)| \leq |\rho_0(t)|^{k-1} = 1,$$

since ρ_0 is as in Theorem 3.3. Combining all these pieces, we have that

$$\left\|f(A)\boldsymbol{b} - \boldsymbol{f}_{m+1}^{(k)}\right\|_{A^{-1}(\theta_0 I - A)^{-1}} \leq \int_0^\infty \frac{1}{\lambda_{\min} + t} \left(1 - \frac{\lambda_{\min}}{\theta_0}\right)^k \xi_m^k \left\|\boldsymbol{e}_0^{(1)}(0)\right\|_{A(\theta_0 I - A)^{-1}} \mathrm{d}\mu(t).$$

Using (3.8) again, we obtain the desired result. ∎

3.2. Stable evaluation of the residual norms. For implementing the restarted method, which relies on a quadrature-based evaluation of an integral representation of the error function, it is important to be able to evaluate the residual norms stably, i.e. the collinearity factors $\rho_{m+1}(t)$ from Lemma 3.2, to high accuracy. The representation given in Lemma 3.2, while being important from a theoretical point of view, is numerically less stable; but as we explain in the forthcoming paper [5], $\rho_{m+1}(t)$ can alternatively and stably be obtained by solving (triangular) HPD systems. The result is similar to a known result for the standard Arnoldi method, see, e.g., [14, Proposition 6.7] or [4].

4. Numerical results. To demonstrate that our method is indeed an improvement over the standard restarted Lanczos method, we present the following two examples. In both cases, we compute $f(A)\boldsymbol{b}$, where $f(z) = \frac{1}{\sqrt{z}}$ and $\boldsymbol{b}$ is the normalized vector of all ones. The inverse square root is an important function in lattice quantum chromodynamics computations; see, e.g., [4, 15]. Such f is also a Stieltjes function of the form (3.1). In fact, for all $\sigma \in (0, 1)$,

$$z^{-\sigma} = \frac{\sin(\sigma\pi)}{\pi} \int_0^\infty \frac{t^{-\sigma}}{t + z} \, \mathrm{d}t$$

is a Stieltjes function; see, e.g., [9]. For the Radau-Lanczos method, we fix $\theta_0 = \lambda_{\max} + \lambda_{\min}$. As for the cycle length, we fix $m = 10$.

4.1. A diagonal matrix with large condition number. We first consider a diagonal matrix A of size $n = 1000$. Half of the eigenvalues of A are evenly distributed in the interval $[10^{-2}, 10^{-1}]$, and the other half lie in $[10^2, 10^3]$, resulting in a condition number of 10^5.

As one can see in Figure 4.1, the Radau-Lanczos method converges well before the standard one, which cannot even attain a competitive error as it stagnates around cycle 1500. Furthermore, the Radau-Lanczos method requires fewer iterations than the standard method.

4.2. The two-dimensional Laplacian. We next examine how our method performs on the discretized two-dimensional Laplacian operator of dimension $n = 1600$ with condition number $\mathcal{O}(10^2)$. This example shows that, again, the Radau-Lanczos methid is capable of attaining a significantly lower error norm than the standard one, and in fewer iterations, as seen in Figure 4.2.

5. Conclusions and future work. We have presented a new method for computing $f(A)\boldsymbol{b}$ and derived convergence bounds for this method in the case when A is an HPD matrix and f is a Stieltjes function. We developed these results in analogy to CG, even using classical CG results to obtain error bounds for when $f(z) = z^{-1}$. We demonstrated that our method improves upon the restarted standard method with two supporting numerical experiments.

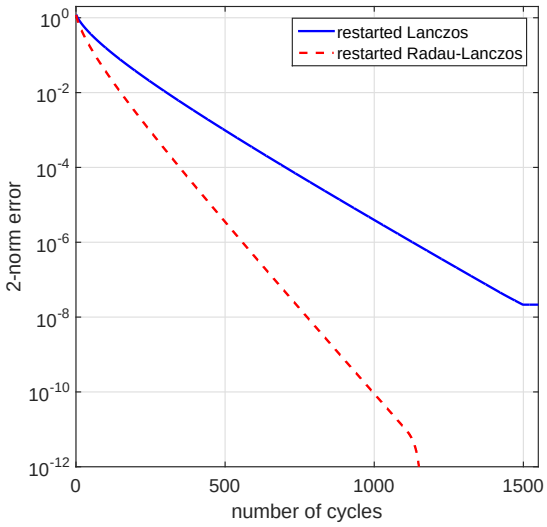

FIG. 4.1. *Convergence curves of example 4.1, where A is a diagonal matrix.*

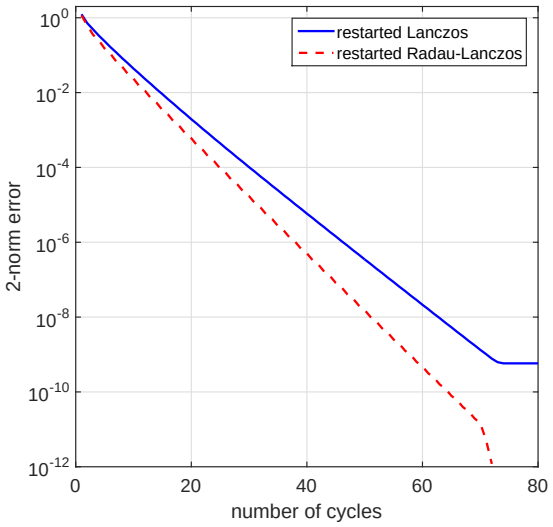

FIG. 4.2. *Convergence curves of example 4.2, where A is the discretized two-dimensional Laplacian operator.*

As noted throughout the paper, we plan to provide more detail on various aspects of the Radau-Lanczos method in the forthcoming paper [5]. In particular, we will present an algorithm which accounts for the stable and efficient computation of the residual norms and error matrix functions, which are essential for the restarted version of our method. We will further examine the behavior of our method in comparison to the standard one when applied to other matrices and functions.

REFERENCES

[1] MARTIN AFANASJEW, MICHAEL EIERMANN, OLIVER G. ERNST, AND STEFAN GÜTTEL, *Implementation of a restarted Krylov subspace method for the evaluation of matrix functions*, Linear Algebra and its Applications, 429 (2008), pp. 229–314.

[2] MICHAEL EIERMANN AND OLIVER G. ERNST, *A restarted Krylov subspace method for the evaluation of matrix functions*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 2481–2504.

[3] ANDREAS FROMMER, STEFAN GÜTTEL, AND MARCEL SCHWEITZER, *Convergence of restarted Krylov subspace methods for Stieltjes functions of matrices*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 1602–1624.

[4] ———, *Efficient and stable Arnoldi restarts for matrix functions based on quadrature*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 661–683.

[5] ANDREAS FROMMER, KATHRYN LUND-NGUYEN, MARCEL SCHWEITZER, AND DANIEL B. SZYLD, *The Radau-Arnoldi method for matrix functions*, 2016. In preparation.

[6] WALTER GAUTSCHI, *Gauss-Radau formulae for Jacobi and Laguerre weight functions*, Math. Comput. Simul., 54 (2000), pp. 403–412.

[7] GENE H. GOLUB, *Some modified matrix eigenvalue problems*, SIAM Review, 15 (1973), pp. 318–334.

[8] GENE H. GOLUB AND GÉRARD MEURANT, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton, 2010.

[9] PETER HENRICI, *Applied and Computational Complex Analysis, volume 2*, John Wiley & Sons, New York, 1977.

[10] NICHOLAS HIGHAM, *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, 2008.

[11] MILOS ILIĆ, IAN W. TURNER, AND DANIEL P. SIMPSON, *A restarted Lanczos approximation to functions of a symmetric matrix*, IMA Journal of Numerical Analysis, 30 (2010), pp. 1044–1061.

[12] DAVID G. LUENBERGER, *Linear and Nonlinear Programming*, Adison-Wesley, Reading, Massachusetts, 2nd ed., 1984.

[13] YOUSEF SAAD, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM Journal on Numerical Analysis, 29 (1992), pp. 209–228.

[14] ———, *Iterative Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, 2nd ed., 2003.

[15] JASPER VAN DEN ESHOF, ANDREAS FROMMER, THOMAS LIPPERT, KLAUS SCHILLING, AND HENK A. VAN DER VORST, *Numerical methods for the QCD overlap operator. I. Sign-function and error bounds*, Computer Physics Communications, 146 (2002), pp. 203–224.