Convergence of restarted Krylov subspace methods for Stieltjes functions of matrices

Andreas, Frommer and Stefan, Güttel and Marcel, Schweitzer

2014

MIMS EPrint: **2014.28**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

CONVERGENCE OF RESTARTED KRYLOV SUBSPACE METHODS FOR STIELTJES FUNCTIONS OF MATRICES*

ANDREAS FROMMER†, STEFAN GÜTTEL‡, AND MARCEL SCHWEITZER†

Abstract. To approximate $f(A)\boldsymbol{b}$—the action of a matrix function on a vector—by a Krylov subspace method, restarts may become mandatory due to storage requirements for the Arnoldi basis or due to the growing computational complexity of evaluating f on a Hessenberg matrix of growing size. A number of restarting methods have been proposed in the literature in recent years and there has been substantial algorithmic advancement concerning their stability and computational efficiency. However, the question under which circumstances convergence of these methods can be guaranteed has remained largely unanswered. In this paper we consider the class of Stieltjes functions and a related class, which contains important functions like the (inverse) square root and the matrix logarithm. For these classes of functions we present new theoretical results which prove convergence for Hermitian positive definite matrices A and arbitrary restart lengths. We also propose a modification of the Arnoldi approximation which guarantees convergence for the same classes of functions and any restart length if A is not necessarily Hermitian but positive real.

Key words. matrix functions, Krylov subspace methods, restarted Arnoldi method, conjugate gradient method, shifted linear systems, shifted GMRES method, harmonic Ritz values

AMS subject classifications. 65F60, 65F50, 65F10, 65F30

1. Introduction.

When approximating

$$f(A)\boldsymbol{b},$$

the action of a function of a matrix $A \in \mathbb{C}^{n\times n}$ on a vector $\boldsymbol{b} \in \mathbb{C}^n$ by m iterations of *Arnoldi's method* [8, 21, 25, 30], one computes an *Arnoldi decomposition*

$$AV_m = V_m H_m + h_{m+1,m}\boldsymbol{v}_{m+1}\hat{\boldsymbol{e}}_m^T, \quad \hat{\boldsymbol{e}}_m = (0,\ldots,0,1)^T \in \mathbb{R}^m, \tag{1.1}$$

where $H_m = (h_{ij}) \in \mathbb{C}^{m\times m}$ is an upper Hessenberg matrix and the columns of $V_m = [\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m] \in \mathbb{C}^{n\times m}$ form an orthonormal basis of $\mathcal{K}_m(A,\boldsymbol{b})$, the mth Krylov subspace corresponding to A and $\boldsymbol{b}$. One then approximates $f(A)\boldsymbol{b}$ in $\mathcal{K}_m(A,\boldsymbol{b})$ by

$$\boldsymbol{f}_m = V_m f(H_m) V_m^H \boldsymbol{b} = \|\boldsymbol{b}\|_2 V_m f(H_m)\hat{\boldsymbol{e}}_1, \quad \hat{\boldsymbol{e}}_1 = (1,0,\ldots,0)^T \in \mathbb{R}^m. \tag{1.2}$$

This procedure requires storing the full Arnoldi basis and evaluating $f(H_m)$, a function of an $m \times m$ matrix. When a large number m of iterations is necessary to approximate $f(A)\boldsymbol{b}$ with the desired accuracy, e.g., when A has eigenvalues near a singularity of f, the approximation (1.2) may be practically infeasible due to storage limitations and/or high computational complexity. To overcome this problem a number of restarting approaches have been proposed in the literature, where—similarly to the techniques for linear systems—after a certain number of iterations the Arnoldi basis is discarded and a new Arnoldi cycle is started to approximate the error of the last iterate, cf. [1, 2, 10, 11, 16, 23, 35]. While much work has been devoted to tuning the methods towards numerical stability and efficiency, there are only few theoretical

*This work was partially supported by Deutsche Forschungsgemeinschaft through Cooperative Research Centre SFB TRR55 "Hadron Physics from Lattice QCD".

†Department of Mathematics, Bergische Universität Wuppertal, 42097 Wuppertal, Germany, {frommer,schweitzer}@math.uni-wuppertal.de

‡School of Mathematics, The University of Manchester, M13 9PL Manchester, United Kingdom, stefan.guettel@manchester.ac.uk

results concerning the convergence of these methods. Finite termination at the exact solution is guaranteed—at least in exact arithmetic—for the unrestarted Arnoldi method in the sense that there exists $m \leq n$ with $f(A)\boldsymbol{b} = \|\boldsymbol{b}\|_2 V_m f(H_m)\hat{\boldsymbol{e}}_1$. This is not the case for the restarted variant where it might happen that the approximants do not converge to the solution at all. Such unwanted behavior is well known in the context of restarted GMRES for linear systems; see, e.g., [24,32,37,38]. We illustrate by an example that this is also the case for the restarted Arnoldi method.

EXAMPLE 1.1. For odd $n \in \mathbb{N}$ consider the matrix

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}, \tag{1.3}$$

which is non-singular, $\det(A) = 2$. Let $\hat{\boldsymbol{e}}_i$ denote the ith canonical unit vector and take $\boldsymbol{b} = \hat{\boldsymbol{e}}_1$ and the function $f(z) = z^{-1}$. This means that $f(A)\boldsymbol{b} = A^{-1}\boldsymbol{b}$ is the solution of the linear system $A\boldsymbol{x} = \boldsymbol{b}$, and the restarted Arnoldi method reduces to the restarted full orthogonalization method (FOM) [30,34]. The solution of the linear system is given by

$$\boldsymbol{x}_i = \begin{cases} \frac{1}{2} & \text{if } i \text{ is odd,} \\ -\frac{1}{2} & \text{if } i \text{ is even.} \end{cases} \tag{1.4}$$

When the restarted Arnoldi method with restart length $m < n$ is applied to the linear system $A\boldsymbol{x} = \boldsymbol{b}$, the first Arnoldi basis $V_m^{(1)}$ and the upper Hessenberg matrix $H_m^{(1)}$ in (1.1) are given by

$$V_m^{(1)} = [\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_m] \quad \text{and} \quad H_m^{(1)} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & 1 \end{bmatrix} \in \mathbb{R}^{m\times m}. \tag{1.5}$$

The matrix $H_m^{(1)}$ is clearly non-singular, thus the Arnoldi approximation $\boldsymbol{x}_m^{(1)} = V_m^{(1)}\big(H_m^{(1)}\big)^{-1}\hat{\boldsymbol{e}}_1$ from (1.2) is defined. A simple calculation shows that the corresponding residual $\boldsymbol{r}_m^{(1)} = \boldsymbol{b} - A\boldsymbol{x}_m^{(1)}$ satisfies $\boldsymbol{r}_m^{(1)} = \hat{\boldsymbol{e}}_{m+1}$. To describe the situation for further restart cycles let us agree that the lower indices ℓ in the canonical unit vectors $\hat{\boldsymbol{e}}_\ell$ are to be understood modulo n, i.e., $\hat{\boldsymbol{e}}_{n+1} \equiv \hat{\boldsymbol{e}}_1$, and so on. The second restart cycle computes the Arnoldi basis $V_m^{(2)} = [\hat{\boldsymbol{e}}_{m+1}, \ldots, \hat{\boldsymbol{e}}_{2m}]$, the same Hessenberg matrix $H_m^{(2)} = H_m^{(1)}$, and the residual $\boldsymbol{r}_m^{(2)} = \hat{\boldsymbol{e}}_{2m+1}$. Continuing in this manner we obtain that in all restart cycles the Hessenberg matrices are equal to the one from (1.5) and that in the kth cycle the Arnoldi basis consists of the canonical unit vectors $\hat{\boldsymbol{e}}_{(k-1)m+1}, \ldots, \hat{\boldsymbol{e}}_{km}$, and $\boldsymbol{r}_m^{(k)} = \hat{\boldsymbol{e}}_{km+1}$.

As a consequence, restarted FOM does not converge, since its residuals do not tend to $\boldsymbol{0}$, for any restart value $m \in \{1, \ldots, n-1\}$. We note that in this example A is normal and has the eigenvalues $1 + \exp(2\pi \mathrm{i} j/n)$ $(j = 1, \ldots, n)$ which all lie in the right half plane. We refer to [33] for further investigations of the arbitrary convergence behavior of restarted FOM.

Example 1.1 illustrates that we cannot guarantee the convergence of a restarted Arnoldi method unless additional assumptions on A and/or f are made. We restrict ourselves to the class of *Stieltjes functions* (see, e.g., [4, 5, 18]), which can be represented as a Riemann–Stieltjes integral

$$f(z) = \int_0^\infty \frac{1}{t+z}\,\mathrm{d}\mu(t), \tag{1.6}$$

with respect to a function $\mu(t)$ which is monotonically increasing and positive on $[0,\infty)$ and satisfies the condition

$$\int_0^\infty \frac{1}{t+1}\,\mathrm{d}\mu(t) < \infty. \tag{1.7}$$

Simple examples of Stieltjes functions arise if $\mu(t)$ is a piecewise constant step function with a finite number of (positive) jumps ξ_i at points t_i, i.e.,

$$f(z) = \sum\nolimits_{i=0}^{r} \frac{\xi_i}{t_i+z}.$$

Many other Stieltjes functions correspond to a differentiable function $\mu(t)$. In this case (1.6) can be written as an ordinary (improper) Riemann integral

$$\int_0^\infty \frac{\mu'(t)}{t+z}\,\mathrm{d}t, \tag{1.8}$$

cf., e.g., [18]. Two important examples of Stieltjes functions of this type are

$$z^{-\alpha} = \frac{\sin(\alpha\pi)}{\pi}\int_0^\infty \frac{t^{-\alpha}}{t+z}\,\mathrm{d}t \quad \text{for } \alpha \in (0,1)$$

and

$$\frac{\log(1+z)}{z} = \int_1^\infty \frac{t^{-1}}{t+z}\,\mathrm{d}t. \tag{1.9}$$

More examples of Stieltjes functions can be found in, e.g., [4, 5, 18].

The few theoretical results concerning the convergence of the restarted Arnoldi method for matrix functions available in the literature [2, 10] are based on approximation theory and make use of bounds for the error of interpolating polynomials for certain classes of analytic functions, using the connection between Krylov subspace methods and polynomial interpolation (as explained in, e.g., [12, 30]). Here we take a different approach, using the intimate relation between Arnoldi's method for matrix functions and FOM for families of shifted linear systems [34], as well as a similar relation with the "shifted GMRES" method from [15].

The remainder of this paper is organized as follows. In section 2 we briefly review the ideas behind restarted Krylov subspace methods for $f(A)\boldsymbol{b}$ without going into algorithmic details. We also reproduce known convergence results for the restarted Arnoldi method in this section and formulate a new result on the monotonic decrease of the Euclidean norm of the error. In section 3 we discuss the relation between the Arnoldi approximation of matrix functions and FOM for families of shifted linear systems. In section 4 we prove convergence of the restarted Arnoldi method with arbitrary restart length for Stieltjes functions and a class of related functions when

A is Hermitian positive definite. Thick restarts as a means towards accelerating convergence are briefly discussed in section 5. A modification of Arnoldi's method using harmonic Ritz values as interpolation points and the corresponding convergence theory for non-Hermitian, positive real matrices are presented in section 6. We end with results of some numerical experiments in section 7.

2. Restarted Arnoldi approximations for $f(A)\boldsymbol{b}$. The main idea when restarting Arnoldi's method for matrix functions is to apply an *additive correction*

$$\boldsymbol{f}_m^{(2)} = \boldsymbol{f}_m^{(1)} + \boldsymbol{e}_m^{(1)},$$

where $\boldsymbol{f}_m^{(1)}$ denotes the approximation (1.2) obtained from m iterations of the standard Arnoldi method and $\boldsymbol{e}_m^{(1)}$ is an approximation of the error $f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(1)}$ obtained by m iterations of a new Arnoldi cycle. Repeated application of this additive correction leads to approximations $\boldsymbol{f}_m^{(k)}$ for $f(A)\boldsymbol{b}$ with

$$\boldsymbol{f}_m^{(k)} = \boldsymbol{f}_m^{(k-1)} + \boldsymbol{e}_m^{(k-1)}, \quad k = 2, 3, \ldots .$$

A requirement for employing this approach is to have a representation of the error $f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(1)}$ in the form $e_m^{(1)}(A)\boldsymbol{v}^{(1)}$, with a new function $e_m^{(1)}(z)$ and a new vector $\boldsymbol{v}^{(1)}$, in order to again apply Arnoldi's method for approximating it. A first result in this direction was given in [10], characterizing the restart function $e_m^{(1)}(z)$ as the mth order divided difference [6] of $f(z)$ with respect to the Ritz values, i.e., the eigenvalues of H_m; see also [23]. For functions representable by a "Cauchy-type" integral, an integral representation of the restart function instead of a representation using divided differences was given in [16], which was then used to develop a numerically stable restart procedure. We rephrase this result for Stieltjes functions.

THEOREM 2.1. *Let f be a Stieltjes function. Assume* $\operatorname{spec}(A) \cap (-\infty, 0] = \emptyset$ *and denote by $\boldsymbol{f}_m$ the mth Arnoldi approximation* (1.2) *to $f(A)\boldsymbol{b}$. Assume that* $\operatorname{spec}(H_m) = \{\theta_1, \ldots, \theta_m\}$ *satisfies* $\operatorname{spec}(H_m) \cap (-\infty, 0] = \emptyset$ *and define*

$$e_m(z) = (-1)^{m+1} \|\boldsymbol{b}\|_2 \gamma_m \int_0^\infty \frac{1}{w_m(t)} \cdot \frac{1}{z+t} \, \mathrm{d}\mu(t), \quad z \notin (-\infty, 0], \tag{2.1}$$

where $w_m(t) = (t+\theta_1)\cdots(t+\theta_m)$ and $\gamma_m = \prod_{i=1}^m h_{i+1,i}$. Then

$$f(A)\boldsymbol{b} - \boldsymbol{f}_m = e_m(A)\boldsymbol{v}_{m+1}, \tag{2.2}$$

where $\boldsymbol{v}_{m+1}$ is the $(m+1)$st Arnoldi vector.

This result is contained in [16, Thm 3.2] for the case when μ is differentiable and thus f is of the form (1.8).* For general μ the proof of (2.2) can be done in an identical manner, so we refrain from reproducing it here. Inductively, Theorem 2.1 also yields a representation for the error $f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}$ of the restarted Arnoldi approximation $\boldsymbol{f}_m^{(k)}$ after k restart cycles. For its formulation we continue to systematically use an upper index in brackets to distinguish quantities from the various cycles. We then have

$$f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)} = e_m^{(k)}(A)\boldsymbol{v}_{m+1}^{(k)}, \tag{2.3}$$

*In [16] the integration interval is $(-\infty, 0]$, i.e., the integration variable t is transformed as $t \to -t$.

where

$$e_m^{(k)}(z) = (-1)^{k(m+1)} \|\boldsymbol{b}\|_2 \left(\prod_{j=1}^{k} \gamma_m^{(j)} \right) \int_0^\infty \frac{1}{\prod_{j=1}^{k} w_m^{(j)}(t)} \cdot \frac{1}{z+t} \, \mathrm{d}\mu(t). \tag{2.4}$$

For f a Stieltjes function and A Hermitian positive definite, the Ritz values θ_i are all real and positive. The conditions $\mathrm{spec}(A) \cap (-\infty, 0] = \emptyset$ and $\mathrm{spec}(H_m) \cap (-\infty, 0] = \emptyset$ are then always fulfilled, and the nodal polynomial $w_m(t) = (t+\theta_1) \cdots (t+\theta_m)$ is positive for $t \geq 0$. Therefore, $1/w_m(t)$ is also positive, and there exists a constant $\alpha > 0$ such that $1/w_m(t) \leq \frac{\alpha}{1+t}$ for $t \geq 0$. Together with (1.7) this shows that

$$\widetilde{\mu}(t) = \int_0^t \frac{1}{w_m(\tau)} \, \mathrm{d}\mu(t)$$

is defined and has a finite value for all $t \geq 0$. Since

$$\mathrm{d}\widetilde{\mu}(t) = \frac{1}{w_m(t)} \, \mathrm{d}\mu(t),$$

this yields the following proposition.

PROPOSITION 2.2. *Let f be a Stieltjes function and $A \in \mathbb{C}^{n \times n}$ be Hermitian positive definite. Then the error function $e_m(z)$ from* (2.1) *is a scalar multiple of a new Stieltjes function,*

$$e_m(z) = (-1)^{m+1} \|\boldsymbol{b}\|_2 \gamma_m \int_0^\infty \frac{1}{z+t} \, \mathrm{d}\widetilde{\mu}(t). \tag{2.5}$$

In Algorithm 1 we summarize a generic form of the restarted Arnoldi method based on an error function representation. We refer the reader to [16] for a detailed description of the method when the restart function from (2.2) is evaluated via numerical quadrature, or to [2] for a different approach using a rational approximation of f.

Algorithm 1: Restarted Arnoldi method for $f(A)\boldsymbol{b}$ from [10] (generic version).

Given: A, $\boldsymbol{b}$, f, m
Compute the Arnoldi decomposition $AV_m^{(1)} = V_m^{(1)} H_m^{(1)} + h_{m+1,m}^{(1)} \boldsymbol{v}_{m+1}^{(1)} \hat{\boldsymbol{e}}_m^T$
    with respect to A and $\boldsymbol{b}$.
Set $\boldsymbol{f}_m^{(1)} := \|\boldsymbol{b}\| V_m^{(1)} f(H_m^{(1)}) \hat{\boldsymbol{e}}_1$.
for $k = 2, 3, \ldots$ until convergence **do**
    Determine the error function $e_m^{(k-1)}(z)$.
    Compute the Arnoldi decomposition $AV_m^{(k)} = V_m^{(k)} H_m^{(k)} + h_{m+1,m}^{(k)} \boldsymbol{v}_{m+1}^{(k)} \hat{\boldsymbol{e}}_m^T$
        with respect to A and $\boldsymbol{v}_{m+1}^{(k-1)}$.
    Set $\boldsymbol{f}_m^{(k)} := \boldsymbol{f}_m^{(k-1)} + V_m^{(k)} e_m^{(k-1)}(H_m^{(k)}) \hat{\boldsymbol{e}}_1$.

There are two main results on the convergence of the restarted Arnoldi method in the literature so far. In [10] it is proven that for entire functions of order one (this class contains the exponential function as an important special case), superlinear convergence can be guaranteed for all restart lengths m.

THEOREM 2.3 (Theorem 4.2 and Corollary 4.3 in [10]). *Let $A \in \mathbb{C}^{n\times n}$, $\boldsymbol{b} \in \mathbb{C}^n$, and let f be an entire function of order one. Then there exist constants C and γ independent of m and k such that*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_2 \leq C\frac{\gamma^{km-1}}{(km-1)!}\|\boldsymbol{b}\|_2 \quad \text{for all } k. \tag{2.6}$$

The proof of Theorem 2.3 relies on convergence results for polynomials of best uniform approximation to entire functions of order one, see [13].

The second main convergence result, proven in [1], relies on a different approach. It uses the fact that the sequence of Ritz values being generated by the restarted Arnoldi method with restart length $m = 1$ applied to a Hermitian positive definite matrix A asymptotically alternates between only two values ϑ_1 and ϑ_2, so that the corresponding approximation can be asymptotically characterized as an interpolation process with only two nodes. Using several additional results on the asymptotic behavior of the Arnoldi basis vectors, it was then proven in [1] that the restarted Arnoldi approximation with $m = 1$ converges at least linearly for a class of functions which contains the Stieltjes functions.

THEOREM 2.4 (Corollary 5.5 in [1]). *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian positive definite, $\boldsymbol{b} \in \mathbb{C}^n$, and let $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalue of A, respectively. Let f be a function analytic in $[\lambda_{\min}, \lambda_{\max}]$ which has no singularities in $\mathbb{C} \setminus \mathbb{R}$. Then the restarted Arnoldi method with restart length $m = 1$ converges to $f(A)\boldsymbol{b}$ with asymptotic convergence factor at least*

$$\frac{\lambda_{\max} - \lambda_{\min}}{|\zeta - \lambda_{\max}| + |\zeta - \lambda_{\min}|}, \tag{2.7}$$

where ζ is a singularity of f which is closest to $[\lambda_{\min}, \lambda_{\max}]$.

Note that Arnoldi with restart length $m = 1$ can be seen as a generalization of the method of steepest descent for matrix functions (see [2]). While the two-cyclic behavior of Ritz values can also be observed for larger restart lengths, a generalization of Theorem 2.4 for $m > 1$ is currently unknown.

A third, weaker but noteworthy result is from [14], see also [7].

THEOREM 2.5. *Let f be a Stieltjes function, $A \in \mathbb{C}^{n\times n}$ Hermitian positive definite and let $\boldsymbol{f}_m$ denote the approximation obtained with m iterations of Arnoldi's method (without restarts). Then*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_{m+1}\|_2 < \|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|_2 \quad \text{for all } m \text{ with } \boldsymbol{f}_m \neq f(A)\boldsymbol{b},$$

i.e., the Euclidean norm of the error decreases strictly monotonically.

In the light of Proposition 2.2 this generalizes to the restarted Arnoldi method.

COROLLARY 2.6. *Under the assumptions of Theorem 2.5, the approximations $\boldsymbol{f}_m^{(k)}$ obtained via the restarted Arnoldi method satisfy*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k+1)}\|_2 < \|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_2 \quad \text{for all } k \text{ unless } \boldsymbol{f}_m^{(k)} = f(A)\boldsymbol{b}.$$

Corollary 2.6 only shows that the errors are monotonically decreasing in the Euclidean norm, but it does not state that the errors tend to 0. However, it complements the asymptotic result from Theorem 2.4 for Stieltjes functions (and restart length

$m = 1$) as it shows that the convergence is monotone in the Euclidean norm from the very first cycle.

The next two sections will show that moving from the Euclidean norm to the energy norm will allow us to give stronger convergence results for Stieltjes functions of Hermitian positive definite matrices.

3. Approximating $f(A)\boldsymbol{b}$ and the FOM iterates for shifted linear systems. For $t \geq 0$ let $\boldsymbol{x}_m(t) = \|\boldsymbol{b}\|_2 V_m (H_m + tI)^{-1} \hat{\boldsymbol{e}}_1$ denote the mth FOM iterate for the shifted linear system

$$(A + tI)\boldsymbol{x} = \boldsymbol{b}$$

with initial guess $\boldsymbol{x}_0(t) = \boldsymbol{0}$ and $\operatorname{spec}(A) \cap (-\infty, 0] = \emptyset$. For the Arnoldi approximation $\boldsymbol{f}_m$ for $f(A)\boldsymbol{b}$ with a Stieltjes function $f(z) = \int_0^\infty \frac{1}{t+z}\,\mathrm{d}\mu(t)$ we then have

$$\boldsymbol{f}_m = \|\boldsymbol{b}\|_2 V_m f(H_m)\hat{\boldsymbol{e}}_1 = \int_0^\infty \|\boldsymbol{b}\|_2 V_m (H_m + tI)^{-1}\hat{\boldsymbol{e}}_1\,\mathrm{d}\mu(t) = \int_0^\infty \boldsymbol{x}_m(t)\,\mathrm{d}\mu(t).$$

The respective FOM residuals $\boldsymbol{r}_m(t) = \boldsymbol{b} - (A + tI)\boldsymbol{x}_m(t)$ satisfy

$$\boldsymbol{r}_m(t) = \frac{(-1)^{m+1}\|\boldsymbol{b}\|_2 \gamma_m}{w_m(t)}\boldsymbol{v}_{m+1}. \tag{3.1}$$

This is known from [31], and it also follows by applying (2.1) to the function $f(z) = (z + t)^{-1}$ represented as a Stieltjes function with the step function μ having exactly one jump of size 1 at t. Using Theorem 2.1 we see that the error $\boldsymbol{e}_m = f(A)\boldsymbol{b} - \boldsymbol{f}_m$ can be represented as

$$\boldsymbol{e}_m = e_m(A)\boldsymbol{v}_{m+1} = \int_0^\infty (A + tI)^{-1}\boldsymbol{r}_m(t)\,\mathrm{d}\mu(t) = \int_0^\infty \boldsymbol{e}_m(t)\,\mathrm{d}\mu(t), \tag{3.2}$$

where $\boldsymbol{e}_m(t) = (A+tI)^{-1}\boldsymbol{b} - \boldsymbol{x}_m(t)$ is the error of the mth FOM iterate for the system $(A+tI)\boldsymbol{x} = \boldsymbol{b}$. A similar result holds for analytic functions using the Cauchy integral representation; see, e.g., [11, 21, 30].

We proceed to show that this representation carries over to the restarted case, thus following [11] where this was done for functions with a Cauchy integral representation. Restarting FOM for the system $(A + tI)\boldsymbol{x} = \boldsymbol{b}$ after k cycles of length m means that given the current approximation $\boldsymbol{x}_m^{(k)}$ and its residual $\boldsymbol{r}_m^{(k)}$ we compute a correction for $\boldsymbol{x}_m^{(k)}$ by applying m iterations of FOM to the system $(A + tI)\boldsymbol{e}^{(k)}(t) = \boldsymbol{r}_m^{(k)}(t)$, yielding

$$\boldsymbol{x}_m^{(k+1)}(t) = \boldsymbol{x}_m^{(k)}(t) + \boldsymbol{e}_m^{(k)}(t) \text{ with } \boldsymbol{e}_m^{(k)}(t) = \|\boldsymbol{b}\|_2 V_m^{(k)}(H_m^{(k)} + tI)^{-1}\hat{\boldsymbol{e}}_1.$$

Inductively, using (3.1), we see that for all $t \geq 0$ the residual $\boldsymbol{r}_m^{(k)}(t)$ satisfies

$$\boldsymbol{r}_m^{(k)}(t) = (-1)^{k(m+1)}\|\boldsymbol{b}\|_2 \frac{\prod_{j=1}^k \gamma_m^{(j)}}{\prod_{j=1}^k w^{(j)}(t)}\boldsymbol{v}_{m+1}^{(k)}.$$

Using (2.3) and (2.4), this shows that for the errors $\boldsymbol{e}_m^{(k)} = f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}$ os the restarted Arnoldi approximations $\boldsymbol{f}_m^{(k)}$ we have

$$\boldsymbol{e}_m^{(k)} = e_m^{(k)}(A)\boldsymbol{v}_{m+1}^{(k)} = \int_0^\infty (A + tI)^{-1}\boldsymbol{r}_m^{(k)}(t)\,\mathrm{d}\mu(t) = \int_0^\infty \boldsymbol{e}_m^{(k)}(t)\,\mathrm{d}\mu(t). \tag{3.3}$$

Our approach to prove convergence of the restarted Arnoldi method for Stieltjes functions is based on known convergence results for the (restarted) FOM iterates for $(A+tI)\boldsymbol{x} = \boldsymbol{b}$. For A Hermitian positive definite and $t \geq 0$, the matrix $A+tI$ is Hermitian positive definite as well and (restarted) FOM is mathematically equivalent to (restarted) CG. Of course, restarted CG seems not to be a method with great practical use. However, the convergence theory available for CG can also be applied to restarted CG, and this will allow us to establish the convergence of the restarted Arnoldi method for Stieltjes functions of Hermitian positive definite matrices in the next section. We end the present section by recalling the classical convergence result for CG which uses the energy norm

$$\|\boldsymbol{e}\|_A = \sqrt{\boldsymbol{e}^H A \boldsymbol{e}}.$$

THEOREM 3.1 (see, e.g., [31]). *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian positive definite and $\boldsymbol{x}_0, \boldsymbol{b} \in \mathbb{C}^n$. Further, let $\boldsymbol{x}^*$ denote the solution of the linear system $A\boldsymbol{x} = \boldsymbol{b}$ and let $\boldsymbol{x}_m$ be the mth CG iterate with initial guess $\boldsymbol{x}_0$. Let $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$ denote the condition number of A and define*

$$c = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \quad \text{and} \quad \alpha_m = \frac{1}{\cosh(m \ln c)}. \tag{3.4}$$

Then the energy norm of the error in the CG method satisfies

$$\|\boldsymbol{x}^* - \boldsymbol{x}_m\|_A \leq \alpha_m \|\boldsymbol{x}^* - \boldsymbol{x}_0\|_A.$$

4. Convergence of restarted Arnoldi for Stieltjes matrix functions. Using the tools from section 2 and section 3, we will now derive our main results for positive definite A. Our starting point is the following lemma, which provides a first generalization of Theorem 3.1 to the matrix function case.

LEMMA 4.1. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian positive definite, $\boldsymbol{b} \in \mathbb{C}^n$, and let f be a function of the form (1.6) and $\boldsymbol{f}_m^{(k)}$ the approximation to $f(A)\boldsymbol{b}$ from k cycles of Arnoldi's method with restart length m. Let $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalue of A, respectively, and define the functions*

$$\kappa(t) = \frac{\lambda_{\max}+t}{\lambda_{\min}+t},\ c(t) = \frac{\sqrt{\kappa(t)}-1}{\sqrt{\kappa(t)}+1}, \quad \text{and } \alpha_m(t) = \frac{1}{\cosh(m \ln c(t))}. \tag{4.1}$$

The energy norm of the error of $\boldsymbol{f}_m^{(k)}$ is then bounded by

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A \leq \|\boldsymbol{b}\|_2 \sqrt{\lambda_{\max}} \int_0^\infty \frac{\alpha_m(t)^k}{\sqrt{\lambda_{\min}+t}\cdot\sqrt{\lambda_{\max}+t}}\, \mathrm{d}\mu(t). \tag{4.2}$$

Proof. By using (3.3), we can write

$$f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)} = \int_0^\infty \boldsymbol{e}_m^{(k)}(t)\, \mathrm{d}\mu(t),$$

where $\boldsymbol{e}_m^{(k)}(t)$ denotes the error of the approximation $\boldsymbol{x}_m^{(k)}(t)$ from k cycles of restarted CG with restart length m for the shifted linear system $(A+tI)\boldsymbol{x} = \boldsymbol{b}$. This yields

$$\begin{aligned}\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A &\leq \int_0^\infty \|\boldsymbol{e}_m^{(k)}(t)\|_A\, \mathrm{d}\mu(t) \\ &\leq \int_0^\infty \frac{\sqrt{\lambda_{\max}}}{\sqrt{\lambda_{\max}+t}} \|\boldsymbol{e}_m^{(k)}(t)\|_{A+tI}\, \mathrm{d}\mu(t),\end{aligned}$$

where we used that $\|\boldsymbol{v}\|_A \leq \sqrt{\lambda_{\max}/(\lambda_{\max}+t)} \cdot \|\boldsymbol{v}\|_{A+tI}$ holds for all $t \in [0,\infty)$ since $\boldsymbol{v}^H(A+tI)\boldsymbol{v} = \boldsymbol{v}^H A\boldsymbol{v} + t\boldsymbol{v}^H\boldsymbol{v}$ and $\boldsymbol{v}^H A\boldsymbol{v} \leq \lambda_{\max}\boldsymbol{v}^H\boldsymbol{v}$. We now apply Theorem 3.1 for the shifted matrices $A+tI$ which are positive definite for $t \in [0,\infty)$. Using the fact that the kth cycle of restarted CG can be interpreted as performing m iterations of CG with initial guess $\boldsymbol{x}_m^{(k-1)}(t)$, the result of the previous cycle, and denoting the solution of $(A+tI)\boldsymbol{x} = \boldsymbol{b}$ by $\boldsymbol{x}^*(t)$, we obtain

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A \leq \sqrt{\lambda_{\max}} \int_0^\infty \frac{\alpha_m(t)}{\sqrt{\lambda_{\max}+t}} \|\boldsymbol{x}^*(t) - \boldsymbol{x}_m^{(k-1)}(t)\|_{A+tI} \,\mathrm{d}\mu(t),$$

with $\alpha_m(t)$ from (4.1). Repeatedly applying the CG estimate for all t throughout all restart cycles and using the fact that the initial guess of the first restart cycle is $\boldsymbol{x}_0(t) = \boldsymbol{0}$ for all t, we conclude that

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A \leq \sqrt{\lambda_{\max}} \int_0^\infty \frac{\alpha_m(t)^k}{\sqrt{\lambda_{\max}+t}} \|\boldsymbol{x}^*(t)\|_{A+tI} \,\mathrm{d}\mu(t). \tag{4.3}$$

As $\boldsymbol{x}^*(t) = (A+tI)^{-1}\boldsymbol{b}$, a straightforward calculation shows that

$$\|\boldsymbol{x}^*(t)\|_{A+tI} \leq \frac{\|\boldsymbol{b}\|_2}{\sqrt{\lambda_{\min}+t}}. \tag{4.4}$$

Inserting (4.4) into (4.3) completes the proof. ∎

Lemma 4.1 gives a bound for the energy norm of the error when approximating a Stieltjes matrix function by the restarted Arnoldi method with arbitrary restart length m. It is, however, not immediately clear whether the right-hand side of (4.2) tends to zero as $k \to \infty$, or even whether it is finite at all. We now further investigate the integral on the right-hand side of (4.2) and prove that it is always finite and tends to zero, thus proving convergence for all Stieltjes functions. To do so we need the following auxiliary result on the monotonicity of the function $\alpha_m(t)$ from (4.1).

PROPOSITION 4.2. *The function $\alpha_m(t)$ from (4.1) is monotonically decreasing on $[0,\infty)$.*

Proof. As a function of $t \in [0,\infty)$, κ decreases monotonically from $\kappa(0)$ to 1, c increases monotonically from $c(\kappa(0))$ to 1 as a function of $\kappa \in [\kappa(0),\infty)$, and α_m increases monotonically as a function of $c \in [c(\kappa(0)),1)$. Altogether, thus, α_m decreases monotonically as a function of t. ∎

With these prerequisites, we are prepared to give our first main result.

THEOREM 4.3. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian positive definite, $\boldsymbol{b} \in \mathbb{C}^n$, f a function of the form (1.6), and $\boldsymbol{f}_m^{(k)}$ the approximation from k cycles of Arnoldi's method with restart length m. Further, let $\alpha_m(t)$ be defined as in (4.1) and let $t_0 \geq 0$ be the left endpoint of the support of μ. The energy norm of the error of $\boldsymbol{f}_m^{(k)}$ can then be bounded as*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A \leq C\alpha_m(t_0)^k, \tag{4.5}$$

where

$$C = \|\boldsymbol{b}\|_2\sqrt{\lambda_{\max}} \cdot f\big(\sqrt{\lambda_{\min}\lambda_{\max}}\big) \tag{4.6}$$

is a constant independent of m and k, and $0 \leq \alpha_m(t_0) < 1$. In particular, the restarted Arnoldi method converges for all restart lengths $m \geq 1$.

Proof. We begin by using Lemma 4.1 and Proposition 4.2 to estimate

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A \le \|\boldsymbol{b}\|_2\sqrt{\lambda_{\max}}\int_0^\infty \frac{\alpha_m(t)^k}{\sqrt{\lambda_{\min}+t}\cdot\sqrt{\lambda_{\max}+t}}\,\mathrm{d}\mu(t)$$
$$\le \|\boldsymbol{b}\|_2\alpha_m(t_0)^k\sqrt{\lambda_{\max}}\int_0^\infty \frac{1}{\sqrt{\lambda_{\min}+t}\cdot\sqrt{\lambda_{\max}+t}}\,\mathrm{d}\mu(t)\,. \quad (4.7)$$

The geometric mean $\sqrt{\lambda_{\min}\lambda_{\max}}$ of $\lambda_{\min}$ and $\lambda_{\max}$ satisfies

$$\frac{1}{\sqrt{\lambda_{\min}+t}\cdot\sqrt{\lambda_{\max}+t}} \le \frac{1}{\sqrt{\lambda_{\min}\lambda_{\max}}+t}.$$

Using this in (4.7) we obtain

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A \le \|\boldsymbol{b}\|_2\alpha_m(t_0)^k\sqrt{\lambda_{\max}}\int_0^\infty \frac{1}{\sqrt{\lambda_{\min}\lambda_{\max}}+t}\,\mathrm{d}\mu(t).$$

The integral on the right-hand side is $f(\sqrt{\lambda_{\min}\lambda_{\max}})$, which completes the proof. ∎

Using $\|\boldsymbol{v}\|_2 \le \frac{1}{\sqrt{\lambda_{\min}}}\|\boldsymbol{v}\|_A$ for all $\boldsymbol{v}\in\mathbb{C}^n$, Theorem 4.3 implies the bound

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_2 \le \widetilde{C}\alpha_m(t_0)^k, \text{ where } \widetilde{C} = \|\boldsymbol{b}\|_2\sqrt{\kappa(0)}f(\sqrt{\lambda_{\min}\lambda_{\max}}) \quad (4.8)$$

for the *Euclidean* norm of the error.

We proceed by discussing two "extremal" cases of the error bounds given in (4.5) and (4.8). The first one is restart length $m = 1$, the situation considered in [1]. Then

$$\alpha_1(t_0) = \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}+2t_0} = \frac{\lambda_{\max}-\lambda_{\min}}{|-t_0-\lambda_{\max}|+|-t_0-\lambda_{\min}|},$$

so that we recover exactly the asymptotic convergence factor from Theorem 2.4 since for a Stieltjes function f and A Hermitian positive definite, the singularity of f nearest to $[\lambda_{\min},\lambda_{\max}]$ is always $\zeta = -t_0$.

The second extremal case is $k = 1$, i.e., the unrestarted Arnoldi method.

COROLLARY 4.4. *Let the assumptions of Theorem 4.3 hold and let $\boldsymbol{f}_m$ be the approximation to $f(A)\boldsymbol{b}$ after m iterations of the unrestarted Arnoldi method. The energy norm of the error of $\boldsymbol{f}_m$ can then be bounded as*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|_A \le C\alpha_m(t_0), \quad (4.9)$$

where C is the constant from (4.6).

The bound from (4.9) is, up to the factor C, the same bound as the standard bound for CG convergence for the linear system $(A+t_0I)\boldsymbol{x} = \boldsymbol{b}$ obtained by bounding the CG polynomials with the Chebyshev polynomials. This bound does not incorporate superlinear convergence effects due to spectral adaptation. We will give a result in this direction (Theorem 6.6) at the end of section 6.

It is possible to generalize the result of Theorem 4.3 to another class of functions closely related to Stieltjes functions, namely functions of the form $f(z) = z\widetilde{f}(z)$, where $\widetilde{f}$ is a Stieltjes function. Notable examples are the positive fractional powers $z^\alpha = zz^{\alpha-1}$ for $\alpha\in(0,1)$, including the square root, and the logarithm $\log(1+z) = z\frac{\log(1+z)}{z}$. The error representation for these functions is closely related to the one for Stieltjes functions. The result was given in [16] for functions with a Cauchy-type integral representation. As before, we rephrase it in terms of Stieltjes functions.

PROPOSITION 4.5 (Corollary 3.4 in [16]). *Let the assumptions of Theorem 2.1 hold for the function $\widetilde{f}$ and let $f(z) = z\widetilde{f}(z)$. Denote by $\boldsymbol{f}_m$ the mth Arnoldi approximation* (1.2) *to $f(A)\boldsymbol{b}$. Then*

$$f(A)\boldsymbol{b} - \boldsymbol{f}_m = Ae_m(A)\boldsymbol{v}_{m+1} - h_{m+1,m}\left(\hat{\boldsymbol{e}}_m^T f(H_m)\hat{\boldsymbol{e}}_1\right)\boldsymbol{v}_{m+1}, \tag{4.10}$$

where e_m is the error function for $\tilde{f}$ from Theorem 2.1.

We assume that the restart procedure is implemented as suggested in [16], which means that the second term in (4.10) is evaluated exactly before each restart (all necessary quantities are known) and then added to the Arnoldi iterate, giving the approximation $\hat{\boldsymbol{f}}_m$. Then the error is given by

$$f(A)\boldsymbol{b} - \hat{\boldsymbol{f}}_m = (-1)^{m+1}\|\boldsymbol{b}\|_2\gamma_m A\int_0^\infty \frac{1}{w_m(t)}(A+tI)^{-1}\boldsymbol{v}_{m+1}\,\mathrm{d}\mu(t). \tag{4.11}$$

We can now formulate the following convergence result for this restarted Arnoldi method. It follows by using exactly the same techniques as in the proofs of Lemma 4.1 and Theorem 4.3, now applied to the error representation (4.11), and it makes use of the fact that $\|A\|_A = \lambda_{\max}$.

THEOREM 4.6. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian positive definite, $\boldsymbol{b} \in \mathbb{C}^n$, $f(z) = z\widetilde{f}(z)$ where $\widetilde{f}$ is a Stieltjes function as in* (1.6). *Let $\hat{\boldsymbol{f}}_m^{(k)}$ be the approximation from k cycles of Arnoldi's method with restart length m for $f(A)$, let $\alpha_m(t)$ be defined as in* (4.1), *and let $t_0 \geq 0$ be the left endpoint of the support of μ. Then, if the restart method is implemented as described above,*

$$\|f(A)\boldsymbol{b} - \hat{\boldsymbol{f}}^{(k)}\|_A \leq \lambda_{\max} C\alpha_m(t_0)^k$$

and

$$\|f(A)\boldsymbol{b} - \hat{\boldsymbol{f}}^{(k)}\|_2 \leq \lambda_{\max} \widetilde{C}\alpha_m(t_0)^k,$$

where C and $\widetilde{C}$ are the constants from Theorem 4.3 and from (4.8), *respectively. In particular, the restarted Arnoldi method for $f(z) = z\widetilde{f}(z)$ converges for all restart lengths $m \geq 1$.*

5. Thick restarts. Thick restarts were introduced in [28] and then [11] as a means to accelerate the convergence of the Arnoldi method similar in spirit to deflated restarts in Krylov subspace methods for solving linear systems; see, e.g., [26, 27]. Without going too far into the technical details, we now sketch how the previous convergence results carry over to thick restarts.

The idea of thick restarts is to use information on the spectrum of A gained in cycle k to effectively deflate eigenvalues close to a singularity of f. For A Hermitian and positive definite and f a Stieltjes function, this means that we target at deflating the smallest eigenvalues. Technically, instead of the standard Arnoldi decomposition (1.1), we now compute for each cycle $k > 1$ an Arnoldi-like relation of the form

$$AW_{\ell+m}^{(k)} = W_{\ell+m}^{(k)}G_{\ell+m}^{(k)} + \boldsymbol{v}^{(k)}\hat{\boldsymbol{e}}_{\ell+m}^T,\ \ W_{\ell+m}^{(k)} \in \mathbb{C}^{n\times(\ell+m)},\ \ G_{\ell+m}^{(k)} \in \mathbb{C}^{(\ell+m)\times(\ell+m)},$$

where $W_{\ell+m}^{(k)}$ has orthonormal columns which span the sum of a deflation subspace of dimension ℓ and the Krylov subspace $\mathcal{K}_m(A, \boldsymbol{v}^{(k-1)})$. For a Stieltjes function f this implies that we get a relation analogous to (3.3)

$$\boldsymbol{e}_{\ell+m}^{(k)} = e_{\ell+m}^{(k)}(A)\boldsymbol{v}^{(k)} = \int_0^\infty (A+tI)^{-1}\boldsymbol{r}_{\ell+m}^{(k)}(t)\,\mathrm{d}\mu(t) = \int_0^\infty \boldsymbol{e}_{\ell+m}^{(k)}(t)\,\mathrm{d}\mu(t),$$

where all residuals $\boldsymbol{r}_{\ell+m}^{(k)}(t)$ are collinear with $\boldsymbol{v}^{(k)}$ and $\boldsymbol{e}_{\ell+m}^{(k)}(t) = (A+tI)^{-1}\boldsymbol{r}_{\ell+m}^{(k)}(t)$ is the error of the restarted deflated FOM iterate $\boldsymbol{x}_{\ell+m}^{(k)}(t)$ for $(A+tI)\boldsymbol{x} = \boldsymbol{b}$ after cycle k, i.e.,

$$\boldsymbol{x}_{\ell+m}^{(k)}(t) = \boldsymbol{x}_{\ell+m}^{(k-1)} + W_{\ell+m}^{(k-1)}\big(G_{\ell+m}^{(k-1)} + tI\big)^{-1}\big(W_{\ell+m}^{(k-1)}\big)^H \boldsymbol{r}_{\ell+m}^{(k-1)}(t).$$

Noting that $\boldsymbol{x}_{\ell+m}^{(k-1)} + W_{\ell+m}^{(k-1)}\big(G_{\ell+m}^{(k-1)} + tI\big)^{-1}\big(W_{\ell+m}^{(k-1)}\big)^H \boldsymbol{r}_{\ell+m}^{(k-1)}(t)$ minimizes the energy norm $\|\boldsymbol{x} - (A+tI)^{-1}\boldsymbol{r}_{\ell+m}^{(k-1)}(t)\|_{A+tI}$ over $\boldsymbol{x}_{\ell+m}^{(k-1)} + \operatorname{colspan}(W_{\ell+m}^{(k)})$ and that $\operatorname{colspan}(W_{\ell+m}^{(k)})$ contains $\mathcal{K}_m(A, \boldsymbol{v}^{(k-1)})$, we obtain

$$\|\boldsymbol{e}_{\ell+m}^{(k)}(t)\|_{A+tI} \le \alpha_m^k(t)\,\|(A+tI)^{-1}\boldsymbol{b}\|_{A+tI}$$

with $\alpha_m(t)$ from (4.1). We thus arrive at the following generalization of Theorem 4.3.

THEOREM 5.1. *Under the same assumptions as in Theorem 4.3, the approximations $\boldsymbol{f}_{\ell+m}^{(k)}$ for $f(A)\boldsymbol{b}$ obtained with the thick restarting approach satisfy*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_{\ell+m}^{(k)}\|_A \le C\alpha_m(t_0)^k, \quad C = \|\boldsymbol{b}\|_2\sqrt{\lambda_{\max}}\cdot f(\sqrt{\lambda_{\min}\lambda_{\max}}). \tag{5.1}$$

The interest of this theorem is that it shows that the thick restarting method converges. Its drawback is that (5.1) does not at all reflect the acceleration of convergence which thick restarts are aiming at. A more precise analysis would have to take into account how well $\operatorname{colspan}(W_{\ell+m}^{(k)})$ approximates the smallest eigenvectors of A, and elaborating on this is beyond the scope of this paper. In fact, a rigorous analysis of these effects would need to take into account not only the extremal eigenvalues of A, but the distribution of *all* eigenvalues and the components of $\boldsymbol{b}$ in the eigenvector basis of A. Practically, one usually observes that after some number of, say, k_0 cycles, the eigenvectors corresponding to the ℓ smallest eigenvalues are very well approximated by a subspace of $\operatorname{colspan}(W_{\ell+m}^{(k)})$, so that from then on one can replace $\alpha_m(t_0)$ in (5.1) by its *effective* counterpart

$$\alpha_m^{\text{eff}}(t_0) = \frac{1}{\cosh(m\ln c^{\text{eff}})}, \quad c^{\text{eff}} = \frac{\sqrt{\kappa^{\text{eff}}}-1}{\sqrt{\kappa^{\text{eff}}}-1}, \quad \kappa^{\text{eff}} = \frac{\lambda_{\max}+t_0}{\lambda_{\min}^{\text{eff}}+t_0},$$

where $\lambda_{\min}^{\text{eff}}$ is the $(m+1)$st smallest eigenvalue of A. The bound (5.1) is then to be replaced by the estimate

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_{\ell+m}^{(k)}\|_A \lesssim C\big(\alpha_m(t_0)\big)^{k_0}\big(\alpha_m^{\text{eff}}(t_0)\big)^{k-k_0}.$$

6. Restarted harmonic Arnoldi for Stieltjes matrix functions. Any matrix function $f(A)$ is identical to a polynomial matrix function $q_{f,A}(A)$, where $q_{A,f}$ is the polynomial of degree $\le n-1$ which interpolates f at $\operatorname{spec}(A)$; see, e.g., [19, 22]. Interpolation is to be understood in the Hermite sense; more precisely, at each eigenvalue λ all derivatives of f and $q_{A,f}$ have to coincide up to order $\operatorname{ind}(\lambda)-1$, where $\operatorname{ind}(\lambda)$ is the size of the largest Jordan block for λ. By results from [12, 30], we have $q(A)\boldsymbol{b} = V_m q(H_m)V_m^H\boldsymbol{b}$ for any polynomial q of degree $\le m-1$, hence the Arnoldi approximation $\boldsymbol{f}_m$ from (1.2) can also be characterized as $\boldsymbol{f}_m = q_{H_m,f}(A)\boldsymbol{b}$, where $q_{H_m,f}$ interpolates f at the eigenvalues of H_m which are the (standard) Ritz values of A with respect to the Krylov subspace $\mathcal{K}_m(A,\boldsymbol{b})$.

The purpose of this section is to show that an alternative Krylov subspace approximation, based on polynomial interpolation at the *harmonic* Ritz values instead of the standard Ritz values, allows us to enlarge the class of matrices for which we can prove convergence of a restarting method. This class will be the positive real matrices, i.e., matrices A for which $\mathrm{Re}(\boldsymbol{v}^H A\boldsymbol{v}) > 0$ for all $\boldsymbol{v} \in \mathbb{C}^n,\ \boldsymbol{v} \neq 0$. The matrix A from Example 1.1 is positive real, and we showed that even for the simple Stieltjes function $f(z) = z^{-1}$ there is no convergence with the standard restarted Arnoldi method, i.e., when we interpolate at the standard Ritz values.

DEFINITION 6.1. *The* harmonic *Ritz values of $A \in \mathbb{C}^{n\times n}$ with respect to a subspace $\mathcal{U} \subseteq \mathbb{C}^n$ are those numbers $\vartheta \in \mathbb{C}$ for which there exists $\boldsymbol{x} \in \mathcal{U},\ \boldsymbol{x} \neq 0$ such that*

$$A\boldsymbol{x} - \vartheta\boldsymbol{x} \perp A\mathcal{U}.$$

We note that the standard Ritz values θ are those for which $A\boldsymbol{x}' - \theta\boldsymbol{x}' \perp \mathcal{U}$ for some $\boldsymbol{x}' \neq 0$ and that for non-singular A the value ϑ is a harmonic Ritz value with respect to $\mathcal{U}$ if and only if ϑ^{-1} is a standard Ritz value for A^{-1} with respect to $A\mathcal{U}$. From now on we will consider Krylov subspaces $\mathcal{U} = \mathcal{K}_m(A, b)$ and no longer explicitly mention the dependence of the (harmonic) Ritz values on the subspace. The following known results on harmonic Ritz values are key to formulating a restarted *harmonic* Arnoldi method for matrix functions and its analysis for Stieltjes functions. We also refer to [20] where interpolation at harmonic Ritz values has been discussed in the context of the unrestarted Arnoldi method.

LEMMA 6.2. *Consider the Arnoldi decomposition* (1.1).

(i) The harmonic Ritz values of A with respect to $\mathcal{K}_m(A, \boldsymbol{b})$ are the eigenvalues of the matrix

$$\widetilde{H}_m = H_m + \left(h_{m+1,m}H_m^{-1}\hat{\boldsymbol{e}}_m\right)\hat{\boldsymbol{e}}_m^T.$$

(ii) Let $\boldsymbol{u} \in \mathbb{C}^n$ be any vector and $\hat{H}_m = H_m + \boldsymbol{u}\hat{\boldsymbol{e}}_m^H$. Then for any polynomial q of degree $\leq m-1$ we have

$$V_m q(\hat{H}_m)V_m^H\boldsymbol{b} = q(A)\boldsymbol{b}.$$

(iii) With $q_{\widetilde{H}_m+tI,f}$ the polynomial which interpolates f at the spectrum of $\widetilde{H}_m+tI$, $f(\widetilde{H}_m + tI) = q_{\widetilde{H}_m+tI,f}(\widetilde{H}_m + tI)$, we have for all $t \in \mathbb{C}$

$$V_m q_{\widetilde{H}_m+tI,f}(\widetilde{H}_m + tI)V_m^H\boldsymbol{b} = q_{\widetilde{H}_m+tI,f}(A + tI)\boldsymbol{b}.$$

(iv) For $f(z) = z^{-1}$, let $q_{\widetilde{H}_m,()^{-1}}$ be the polynomial interpolating f at $\mathrm{spec}(\widetilde{H}_m)$, and let $p_m(z) = 1 - zq_{\widetilde{H}_m,()^{-1}}(z)$. Put $\tilde{\boldsymbol{x}}_m(t) = V_m(\widetilde{H}_m + tI)^{-1}V_m^H\boldsymbol{b}$. Then

$$\tilde{\boldsymbol{r}}_m(t) := \boldsymbol{b} - (A + tI)\tilde{\boldsymbol{x}}_m(t) = \eta_m(t)\tilde{\boldsymbol{r}}_m(0),$$

where

$$\eta_m(t) = \frac{1}{p_m(-t)},\quad \tilde{\boldsymbol{r}}_m(0) = p_m(A)\boldsymbol{b}. \tag{6.1}$$

Proof. Part (i) can be found in, e.g., [29]. For $\boldsymbol{u} = \mathbf{0}$ part (ii) was shown in [30]; for the matrix $\widetilde{H}_m$ (i.e., $\boldsymbol{u} = h_{m+1,m} H_m^{-1} \hat{\boldsymbol{e}}_m$) the result was given in [36]. The proof for the general case follows in the same manner, i.e., by inductively establishing that $V_m \hat{H}_m^j V_m^H \boldsymbol{b} = A^j \boldsymbol{b}$ for $j = 0, 1, \ldots, m-1$. For part (iii) we use the shift invariance of the Arnoldi relation (1.1),

$$AV_m = V_m H_m + h_{m+1,m} \boldsymbol{v}_{m+1} \hat{\boldsymbol{e}}_m^T \Rightarrow (A + tI)V_m = V_m(H_m + tI) + h_{m+1,m} \boldsymbol{v}_{m+1} \hat{\boldsymbol{e}}_m^T$$

for all $t \in \mathbb{C}$, and apply (ii) to the matrix $A+tI$ and the rank-1 modification $(H_m+tI)+(h_{m+1,m}H_m^{-1}\hat{\boldsymbol{e}}_m)\hat{\boldsymbol{e}}_m^T$ of $H_m + tI$. We note in passing that for $t \neq 0$ the harmonic Ritz values of $A + tI$ are *not* the eigenvalues of $\widetilde{H}_m + tI$. To show (iv) let $q_{\widetilde{H}_m+tI,()^{-1}}(z)$ interpolate $1/z$ at $\operatorname{spec}(\widetilde{H}_m + tI) = \operatorname{spec}(\widetilde{H}_m) + t$. Then $\tilde{\boldsymbol{r}}_m(t) = p_{m,t}(A + tI)\boldsymbol{b}$ with $p_{m,t}(z) = 1 - z q_{\widetilde{H}_m+t,()^{-1}}(z)$. For each t the polynomial $p_{m,t}$ of exact degree m interpolates the zero function at $\operatorname{spec}(\widetilde{H}_m+tI)$ and is normalized to satisfy $p_{m,t}(0) = 1$. In particular, with ϑ_i $(i = 1, \ldots, m)$ denoting the (possibly multiple) eigenvalues of $\widetilde{H}_m$, we have

$$p_{m,0}(z) = \prod_{i=1}^m \left(1 - \frac{z}{\vartheta_i}\right) \tag{6.2}$$

and

$$p_{m,t}(z) = \prod_{i=1}^m \left(1 - \frac{z}{\vartheta_i + t}\right) = \frac{1}{p_{m,0}(-t)} p_{m,0}(z - t), \tag{6.3}$$

the last equality holding because the polynomial on the right-hand side has the same zeros as $p_{m,t}$ and attains the value 1 at $z = 0$. ∎

Let us reserve the term *harmonic Arnoldi approximation* for the approximation $\tilde{\boldsymbol{f}}_m := q_{\widetilde{H}_m,f}(A)\boldsymbol{b}$ for $f(A)\boldsymbol{b}$, where the polynomial $q_{\widetilde{H}_m,f}$ interpolates f at the harmonic Ritz values of A. By Lemma 6.2 we have

$$\tilde{\boldsymbol{f}}_m = V_m f(\widetilde{H}_m) V_m^H \boldsymbol{b},$$

and with f a Stieltjes function, $f(z) = \int_0^\infty \frac{1}{t+z}\,\mathrm{d}\mu(t)$, we obtain the representation

$$\tilde{\boldsymbol{f}}_m = V_m \int_0^\infty (\widetilde{H}_m + tI)^{-1}\,\mathrm{d}\mu(t)\, V_m^H \boldsymbol{b}.$$

Similarly to (3.2), the error of the harmonic Arnoldi approximation is

$$\tilde{\boldsymbol{e}}_m = \int_0^\infty (A + tI)^{-1} \tilde{\boldsymbol{r}}_m(t)\,\mathrm{d}\mu(t) = \int_0^\infty \eta_m(t)(A + tI)^{-1}\,\mathrm{d}\mu(t) \cdot \tilde{\boldsymbol{r}}_m(0)$$

with $\eta_m(t)$ from (6.1). We thus have $\tilde{\boldsymbol{e}}_m = \tilde{e}_m(A)\tilde{\boldsymbol{r}}_m(0)$ with the error function

$$\tilde{e}_m(z) = \int_0^\infty \frac{\eta_m(t)}{z + t}\,\mathrm{d}\mu(t). \tag{6.4}$$

Alternative representations for $\tilde{e}_m(z)$ are possible, for example as a divided difference with respect to the harmonic Ritz values (see, e.g., [10] where a divided difference representation was derived for arbitrary interpolating polynomials). However, the integral representation (6.4) for Stieltjes functions has the same two crucial advantages as has (2.1) and (2.2) for the standard Arnoldi approximation: it allows for a stable

numerical implementation of restarts based on numerical quadrature, and it is the starting point of our convergence analysis for the case that A is positive real. Note that we have

$$\tilde{\boldsymbol{e}}_m = \int_0^\infty \tilde{\boldsymbol{e}}_m(t)\,\mathrm{d}\mu(t) \quad \text{with } \tilde{\boldsymbol{e}}_m(t) = (A+tI)^{-1}\boldsymbol{b} - \tilde{\boldsymbol{x}}_m(t),$$

where $\tilde{\boldsymbol{e}}_m(t)$ is the error of the Krylov subspace approximation $\tilde{\boldsymbol{x}}_m(t) = V_m(\widetilde{H}_m + tI)^{-1}V_m^H\boldsymbol{b}$ to the solution of the system $(A+tI)\boldsymbol{x} = \boldsymbol{b}$. For $t = 0$, the vector $\tilde{\boldsymbol{x}}_m(0)$ is thus just the mth GMRES iterate for the system $A\boldsymbol{x} = \boldsymbol{b}$. For $t > 0$ the vectors $\tilde{\boldsymbol{x}}_m(t)$ are *not* the GMRES iterates for $(A+tI)\boldsymbol{x} = \boldsymbol{b}$; instead they are Krylov subspace approximations characterized by a residual collinear to that of the GMRES iterates for $t = 0$. The collinear residual approach was introduced in [15] for developing variants of restarted GMRES for families of shifted linear systems. We refer to this work for an efficient computational method based on a single QR-factorization (of H_m) to obtain all the vectors $\tilde{\boldsymbol{x}}_m(t)$ for several values of t simultaneously. The paper [15] also showed that the restarted shifted GMRES approach produces convergent iterates $\tilde{\boldsymbol{x}}_m(t)$ for all $t \geq 0$ if A is positive real, and Lemma 6.3 below is a quantitative refinement of results presented therein.

The four quantities

$$\begin{aligned}
\rho &:= \min\big\{\mathrm{Re}\big(\frac{\boldsymbol{v}^H A^{-1}\boldsymbol{v}}{\boldsymbol{v}^H\boldsymbol{v}}\big) : \boldsymbol{v} \in \mathbb{C}^n, \boldsymbol{v} \neq 0\big\},\\
\delta &:= \min\big\{\big|\frac{\boldsymbol{v}^H A\boldsymbol{v}}{\boldsymbol{v}^H\boldsymbol{v}}\big| : \boldsymbol{v} \in \mathbb{C}^n, \boldsymbol{v} \neq 0\big\},\\
\delta' &:= \min\big\{\big|\frac{\boldsymbol{v}^H A^{-1}\boldsymbol{v}}{\boldsymbol{v}^H\boldsymbol{v}}\big| : \boldsymbol{v} \in \mathbb{C}^n, \boldsymbol{v} \neq 0\big\},\\
\nu_{\max} &:= \max\big\{\frac{(A\boldsymbol{v})^H(A\boldsymbol{v})}{\boldsymbol{v}^H\boldsymbol{v}} : \boldsymbol{v} \in \mathbb{C}^n, \boldsymbol{v} \neq 0\big\}
\end{aligned}$$

will be useful in the analysis to follow. Since with A the matrix A^{-1} is also positive real, the numbers ρ, δ, δ' and $\nu_{\max}$ are all positive if A is positive real. Clearly, $\rho \leq \delta'$ with equality holding if the field of values of A is symmetric with respect to the real axis, for example, when $A \in \mathbb{R}^{n\times n}$.

LEMMA 6.3. *Let A be positive real. Then*

$$\|\tilde{\boldsymbol{r}}_m(0)\|_2 \leq \big(1 - \delta\delta'\big)^{m/2}\|\boldsymbol{b}\|_2, \tag{6.5}$$

and the collinearity factors $\eta_m(t)$ from (6.1) *satisfy*

$$|\eta_m(t)| \leq \left(\frac{1}{1+t\rho}\right)^m \leq 1. \tag{6.6}$$

Proof. The bound (6.5) on the GMRES residuals for A positive real was derived in [9, Cor. 6.2]. It improves over the classical textbook bound; see, e.g., [17,31]. From the definition of $\eta_m(t)$ in Lemma 6.2 we have

$$\eta_m(t) = \frac{1}{\prod_{i=1}^m (1 + t/\vartheta_i)},$$

with ϑ_i being the harmonic Ritz values of A. Since the harmonic Ritz values of A are the inverses of the Ritz values of A^{-1}, we have $\vartheta_i^{-1} = \boldsymbol{v}_i^H A^{-1}\boldsymbol{v}_i / \boldsymbol{v}_i^H\boldsymbol{v}_i$ for some

vector $\boldsymbol{v}_i \in \mathbb{C}^n$ and thus $\mathrm{Re}(\vartheta_i^{-1}) \geq \rho$. Therefore, for any $t \geq 0$ we have

$$|1 + t\vartheta_i^{-1}| \geq 1 + t\mathrm{Re}(\vartheta_i^{-1}) \geq 1 + t\rho \quad \text{for } i = 1, \ldots, m,$$

which gives (6.6). ∎

We need a further auxiliary result relating the energy norms with respect to $A^H A$ and $(A + tI)^H(A + tI)$.

LEMMA 6.4. *Let A be positive real.*

(i) For all $\boldsymbol{v} \in \mathbb{C}^n$ and $t \geq 0$ we have

$$\|\boldsymbol{v}\|_{A^H A}^2 \leq \frac{1}{\nu_{\max}^{-1} t^2 + 2\rho t + 1} \|\boldsymbol{v}\|_{(A+tI)^H(A+tI)}^2 .$$

(ii) For $t \geq 0$ we have

$$\frac{1}{\nu_{\max}^{-1} t^2 + 2\rho t + 1} \leq \frac{\nu_{\max}}{(t + \rho\nu_{\max})^2} .$$

Proof. For part (i) we expand

$$\|\boldsymbol{v}\|_{(A+tI)^H(A+tI)}^2 = \|\boldsymbol{v}\|_{A^H A}^2 + 2t\mathrm{Re}(\boldsymbol{v}^H A\boldsymbol{v}) + t^2\|\boldsymbol{v}\|_2^2 .$$

The inequality now follows from $\|\boldsymbol{v}\|_2^2 \geq \frac{1}{\nu_{\max}} \|\boldsymbol{v}\|_{A^H A}^2$ and

$$\mathrm{Re}(\boldsymbol{v}^H A\boldsymbol{v})/(\boldsymbol{v}^H A^H A\boldsymbol{v}) = \mathrm{Re}(\boldsymbol{w}^H A^{-1}\boldsymbol{w})/(\boldsymbol{w}^H\boldsymbol{w}) \geq \rho, \text{ where } A\boldsymbol{w} = \boldsymbol{v}.$$

The inequality in part (ii) is equivalent to $(t + \rho\nu_{\max})^2 \leq t^2 + 2\rho\nu_{\max} t + \nu_{\max}$, i.e., to $\rho^2\nu_{\max} \leq 1$, which can be established as follows: let $\boldsymbol{v}$ be the normalized eigenvector of $(AA^H)^{-1}$ corresponding to the smallest eigenvalue which is $1/\nu_{\max}$, since the spectra of $A^H A$ and AA^H are identical. Then

$$\rho \leq \delta' \leq |\boldsymbol{v}A^{-1}\boldsymbol{v}| \leq \|\boldsymbol{v}\|_2 \|A^{-1}\boldsymbol{v}\|_2 = \frac{1}{\nu_{\max}^{1/2}} .$$

∎

We are now in a position to state the following result on the convergence of the restarted harmonic Arnoldi method as an analogon to Lemma 4.1 and Theorem 4.3 for the standard Arnoldi approximations.

THEOREM 6.5. *Let $A \in \mathbb{C}^{n\times n}$ be positive real, $\boldsymbol{b} \in \mathbb{C}^n$, f a function of the form (1.6), and $\tilde{\boldsymbol{f}}_m^{(k)}$ the approximation from k cycles of the harmonic Arnoldi approximation method with restart length m. Define*

$$\tilde{\alpha}_m(t) := \left(\frac{\sqrt{1 - \delta\delta'}}{1 + t\rho} \right)^m , \quad t \geq 0.$$

Let $t_0 \geq 0$ be the left endpoint of the support of μ. Then the $A^H A$-energy norm of the error of $\tilde{\boldsymbol{f}}_m^{(k)}$ satisfies

$$\|f(A)\boldsymbol{b} - \tilde{\boldsymbol{f}}_m^{(k)}\|_{A^H A} \leq \|\boldsymbol{r}_m^{(k)}(0)\|_2 \int_0^\infty \frac{(1 + t\rho)^{-mk}}{\sqrt{\nu_{\max}^{-1} t^2 + 2\rho t + 1}} \, \mathrm{d}\mu(t) \tag{6.7}$$

$$\leq \|\boldsymbol{b}\|_2 \int_0^\infty \frac{\tilde{\alpha}_m(t)^k}{\sqrt{\nu_{\max}^{-1} t^2 + 2\rho t + 1}} \, \mathrm{d}\mu(t) \tag{6.8}$$

$$\leq C\tilde{\alpha}_m(t_0)^k, \tag{6.9}$$

where $0 \le \tilde{\alpha}_m(t_0) < 1$ *and*

$$C = \|\boldsymbol{b}\|_2 \cdot \sqrt{\nu_{\max}} \cdot f(\rho\nu_{\max}). \tag{6.10}$$

In particular, the restarted harmonic Arnoldi method converges for all restart lengths $m \ge 1$.

Proof. As the proof is very similar to that of Lemma 4.1 and Theorem 4.3 we only give a sketch. Using an upper index, as before, to distinguish the quantities belonging to the different cycles we have

$$f(A)\boldsymbol{b} - \tilde{\boldsymbol{f}}_m^{(k)} = \int_0^\infty \tilde{\boldsymbol{e}}_m^{(k)}(t)\,\mathrm{d}\mu(t) = \int_0^\infty (A+tI)^{-1}\tilde{\boldsymbol{r}}_m^{(k)}(t)\,\mathrm{d}\mu(t).$$

Using Lemma 6.4(i) together with the equality $\|\tilde{\boldsymbol{e}}_m^{(k)}(t)\|_{(A+tI)^H(A+tI)} = \|\tilde{\boldsymbol{r}}_m^{(k)}(t)\|_2$ and the collinearity of these residuals as stated in Lemma 6.2(iv) one obtains

$$\|f(A)\boldsymbol{b} - \tilde{\boldsymbol{f}}_m^{(k)}\|_{A^HA} \le \int_0^\infty \frac{|\eta_m^{(1)}(t)\cdots\eta_m^{(k)}(t)|}{\sqrt{\nu_{\max}^{-1}t^2 + 2\rho t + 1}}\|\tilde{\boldsymbol{r}}_m^{(k)}(0)\|_2\,\mathrm{d}\mu(t).$$

Inequality (6.7) now follows by bounding each factor $|\eta_m^{(j)}(t)|$ via Lemma 6.3(ii). The second relation (6.8) is obtained by using the bound for $\|\tilde{\boldsymbol{r}}_m^{(k)}(0)\|_2$ from Lemma 6.3(i). To get (6.9) and (6.10) one then uses the fact that $\tilde{\alpha}_m(t)$ is monotonically decreasing as a function of t and Lemma 6.4(ii). ∎

Using $(1+t\rho) \le 1$ for all $t \ge 0$, (6.7) yields

$$\|f(A)\boldsymbol{b} - \tilde{\boldsymbol{f}}_m^{(k)}\|_{A^HA} \le \|\boldsymbol{r}_m^{(k)}(0)\|_2 \int_0^\infty \frac{(1)}{\sqrt{\nu_{\max}^{-1}t^2 + 2\rho t + 1}}\,\mathrm{d}\mu(t) \le C_1\|\boldsymbol{r}_m^{(k)}(0)\|_2, \tag{6.11}$$

where $C_1 = \sqrt{\nu_{\max}}f(\rho\nu_{\max})$, the last inequality holding because of Lemma 6.4(ii). This estimate indicates that the restarted harmonic Arnoldi method converges at least as fast as restarted GMRES for the system $A\boldsymbol{x} = \boldsymbol{b}$. In particular, if (restarted or unrestarted) GMRES shows superlinear convergence, so does (restarted or unrestarted) harmonic Arnoldi for Stieltjes functions.

The estimate (6.11) admits a counterpart for the standard Arnoldi method (and A Hermitian positive definite), to be derived in exactly the same manner: one makes use of the fact that the shifted CG residuals are all collinear and that they can be expressed via polynomials similar to (6.2) and (6.3), except that the harmonic Ritz values need to be replaced by the standard Ritz values, all of which lie in $[\lambda_{\min}, \lambda_{\max}]$. We state this result as our last theorem.

THEOREM 6.6. *Let* $A \in \mathbb{C}^{n\times n}$ *be Hermitian positive definite,* $\boldsymbol{b} \in \mathbb{C}^n$, f *a function of the form* (1.6), *and* $\boldsymbol{f}_m^{(k)}$ *the approximation from* k *cycles of Arnoldi's method with restart length* m. *Further, let* $t_0 \ge 0$ *be the left endpoint of the support of* μ. *The* A^2*-energy norm of the error of* $\boldsymbol{f}_m^{(k)}$ *can then be bounded as*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_{A^2} \le C\|\boldsymbol{r}_m^{(k)}(t_0)\|_2,$$

where $\|\boldsymbol{r}_m^{(k)}(t_0)\|_2$ *is the 2-norm of the residual of the (restarted) CG iterate for the system* $(A+t_0I)\boldsymbol{x} = \boldsymbol{b}$ *and*

$$C = \lambda_{\max}\cdot f\big(\sqrt{\lambda_{\max}}\big).$$

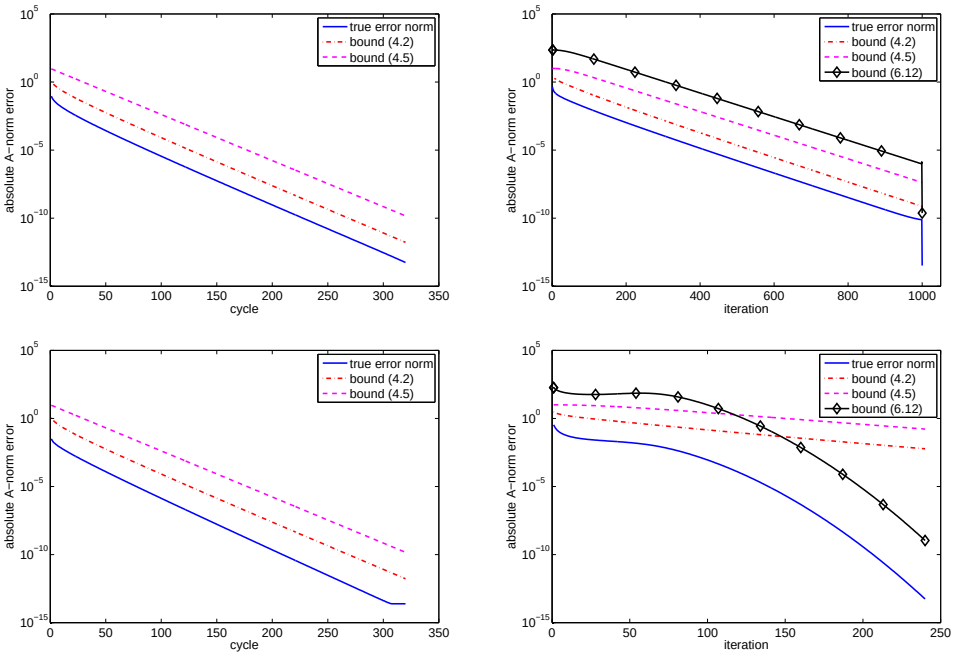

FIG. 7.1. *Comparison of the norm of the error and the error bounds from Lemma 4.1, Theorem 4.3 and Theorem 6.6 for diagonal A, $f(z) = z^{-1/2}$, restart length $m = 20$ (left) and unrestarted Arnoldi (right). Top row: the eigenvalues of A are the Chebyshev points in $[10^{-2}, 10^2]$. Bottom row: the eigenvalues of A are linearly spaced in $[10^{-2}, 10^2]$.*

Consequently,

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{(k)}\|_A \leq \frac{C}{\sqrt{\lambda_{\min}}} \|\boldsymbol{r}_m^{(k)}(t_0)\|_2. \tag{6.12}$$

We cannot expect (6.12) to be tight, however it indicates that the convergence behavior of the (restarted) Arnoldi method is dominated by the convergence of CG for $A\boldsymbol{x} = \boldsymbol{b}$. For example, if (unrestarted) CG shows superlinear convergence, the (unrestarted) Arnoldi approximations for $f(A)\boldsymbol{b}$ can be expected to converge superlinearly, too.

To end this section, we remark that it is also possible to prove convergence of the restarted harmonic Arnoldi approach for A positive real and $f(z) = z\hat{f}(z)$ with $\hat{f}(z)$ a Stieltjes function, analogously to Theorem 4.6, and for an appropriate thick restarting variant. Moreover, it is possible to define shifted harmonic Ritz values ϑ_i with respect to a subspace $\mathcal{U} \subseteq \mathbb{C}^n$ and a "target" t_0 other than 0 (which we have implicitly used here; cf. Definition 6.1): these shifted harmonic Ritz values satisfy

$$(A + t_0 I)\boldsymbol{x}_i - (\vartheta_i + t_0)\boldsymbol{x}_i \perp (A + t_0 I)\mathcal{U}$$

with non-zero vectors $\boldsymbol{x}_i \in \mathbb{C}^n$ (see, e.g., [20]). The shifted harmonic Ritz values would be more adequate if the left endpoint t_0 of the support of μ is different from zero and all results presented in this section can be modified to account for that.

7. Numerical experiments. The purpose of this paper was to advance the convergence theory for restarted Arnoldi-type methods. We end with a few (academic) numerical examples illustrating the theory.

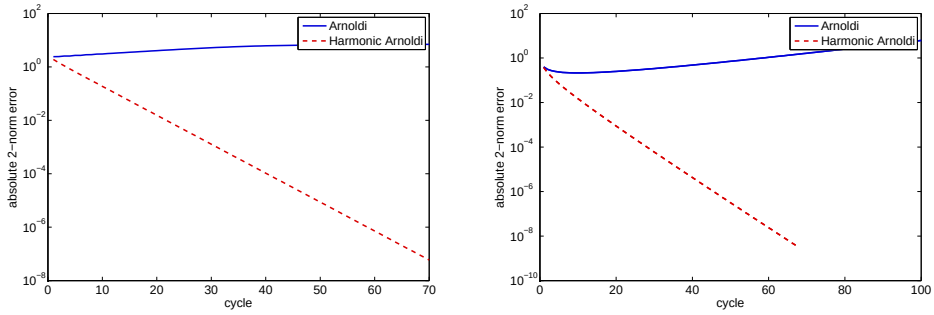

FIG. 7.2. *Convergence curves of the restarted Arnoldi and the restarted harmonic Arnoldi for the matrix from* (7.1) *for the functions* $f(z) = z^{-1}$ *(left) and* $f(z) = z^{-1/2}$ *(right), respectively.*

We first consider a Hermitian positive definite (diagonal) matrix $A \in \mathbb{R}^{1000\times 1000}$ with its eigenvalues chosen as the Chebyshev points in $[10^{-2}, 10^2]$ and compute $f(A)\boldsymbol{b}$ with $f = z^{-1/2}$ and $\boldsymbol{b}$ the (normalized) vector of all ones. In the top row of Figure 7.1 we compare the energy norm of the error with the bound (4.2) (which has been evaluated using adaptive Gauß–Kronrod quadrature) and the bound (4.5). On the left-hand side of Figure 7.1 we give the results for restart length $m = 20$, while on the right-hand side we report the results for the unrestarted Arnoldi method (i.e., $k = 1$ and $m = 1000$) and also show the error norm predicted by bound (6.12) (which is of more interest in the unrestarted case as superlinear convergence effects do typically not take place for small restart lengths). In both cases, the convergence rate is predicted quite accurately, with (4.2) overestimating the true norm of the error by about one order of magnitude and (4.5) by about two orders of magnitude. Of course, the convergence to (approximately) machine precision in the last iteration of the unrestarted Arnoldi method is not captured by the error bounds except by (6.12).

When the eigenvalues are distributed differently, the bounds (4.2) and (4.5) may capture the actual convergence behavior less well. To illustrate this we choose f and $\boldsymbol{b}$ as before, but now the eigenvalues of A are linearly spaced in $[10^{-2}, 10^2]$. The results are given in the bottom row of Figure 7.1. For the small restart length $m = 20$, the bounds again reflect the actual convergence well, although now the rate of the bounds is slightly too pessimistic. There is a substantial difference for the unrestarted Arnoldi method: while in an initial phase the slope of all curves is approximately the same, after some iterations the superlinear convergence behavior of Arnoldi's method (see, e.g., [3]) appears and this is only captured by the error bound (6.12). This indicates that the error bounds (4.2) and (4.5) are not well suited for monitoring the progress if superlinear convergence due to spectral adaptation occurs, as it can happen for the unrestarted Arnoldi method or the restarted Arnoldi method with a *large* restart length m (relatively to the matrix size n). Identifying superlinear convergence effects is, however, possible by keeping track of the residual norm of a CG iterate for the system $(A+t_0 I)\boldsymbol{x} = \boldsymbol{b}$ as indicated by the behavior of error bound (6.12) on the right-hand side of Figure 7.1. While the superlinear convergence behavior is predicted quite accurately, the bound itself substantially overestimates the magnitude of the error.

To illustrate the behavior of the harmonic Arnoldi approach we first extend the

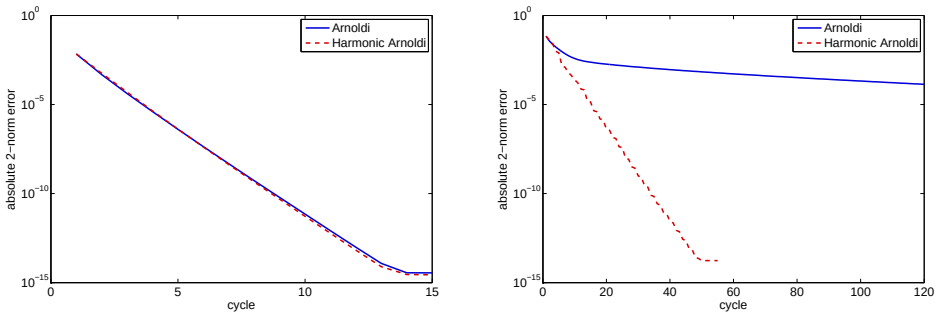

FIG. 7.3. *Convergence curves of the restarted Arnoldi method and the restarted harmonic Arnoldi method for a diagonal matrix A and $f(z) = z^{-1/2}$. The eigenvalues of A are chosen randomly in the disk of radius 1 centered at $1 + 10^{-1}$ (left) and $1 + 10^{-5}$ (right).*

example from section 1 and consider the matrix

$$A = \begin{bmatrix} \alpha & 0 & \cdots & 0 & 1 \\ 1 & \alpha & 0 & \cdots & 0 \\ 0 & 1 & \alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \alpha \end{bmatrix} \in \mathbb{R}^{n \times n}, \tag{7.1}$$

where $\alpha \in \mathbb{R}$ is a parameter and n is odd. The matrix A is normal (so that its field of values is the convex hull of the eigenvalues) and its eigenvalues are the nth roots of unity shifted by α. For $\cos(2\pi/n) < \alpha$, all eigenvalues of A lie in the right half plane, hence A is positive real. Figure 7.2 shows the convergence curves of the restarted Arnoldi and restarted harmonic Arnoldi methods for $n = 21$, $\alpha = 0.995$, $\boldsymbol{b} = \hat{\boldsymbol{e}}_1$, restart length $m = 10$, and the functions $f(z) = z^{-1}$ (left) and $f(z) = z^{-1/2}$ (right). For both Stieltjes functions the restarted harmonic Arnoldi method converges as guaranteed by Theorem 6.5 while the standard restarted Arnoldi method diverges. The decrease of the error norm in the first few cycles of the standard Arnoldi method for approximating $A^{-1/2}$ can be explained as follows: for larger values of t, the shifted linear systems with $A + tI$ do converge, so that the divergence only stems from the behavior of the systems for t near 0. In the first few iterations, the error components implicitly corresponding to the linear systems with larger shifts are reduced and the divergence of the systems with small shifts is not yet visible.

Next we compute $f(A)\boldsymbol{b}$ for $f(z) = z^{-1/2}$ and a diagonal matrix A of dimension 1000 with eigenvalues $\alpha + 1 + re^{2\pi \mathrm{i}\theta}$, where r and θ are two random variables which are independently and uniformly distributed in $[0, 1]$, and $\alpha > 1$. As A is normal and its eigenvalues lie in the right half plane it is again positive real. Figure 7.3 shows the convergence of the restarted standard and harmonic Arnoldi methods with restart length $m = 20$ for $\alpha = 10^{-1}$ (left) and $\alpha = 10^{-5}$ (right). In the first case the matrix is quite well conditioned and both methods behave very similarly and converge quite fast. For $\alpha = 10^{-5}$ the spectrum of A is much closer to the origin. Both restarted methods still converge, but now the harmonic Arnoldi method performs significantly faster. Note that the convergence of the restarted harmonic Arnoldi method is guaranteed by the bound from Theorem 6.5. We do not report its value here because the bound is very pessimistic, neither capturing the order of magnitude of the error nor the

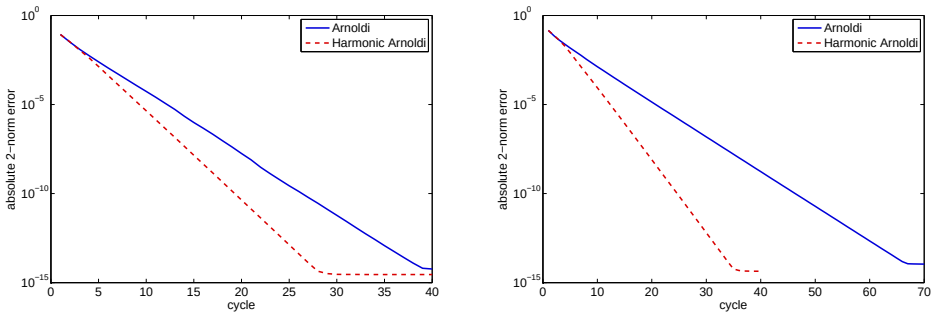

FIG. 7.4. *Convergence curves of the restarted Arnoldi method and the restarted harmonic Arnoldi method, where A is non-diagonalizable with 2×2 Jordan blocks, and $f(z) = z^{-1/2}$. The details on the spectrum of A for the left and right plots are given in the text.*

convergence slope accurately (which is no surprise, as this phenomenon is also known for the GMRES bound it is based upon).

As our last example we take a non-normal, non-diagonalizable matrix A. It is of dimension 1000 and consists of 500 2×2 Jordan blocks of the form

$$\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}.$$

Such a block is positive real if $\mathrm{Re}(\lambda) > 0.5$. For our test we produced A randomly, the imaginary part of λ being uniformly distributed in $[-10, 10]$ and the real part being uniformly distributed in $[0.6, 0.8]$ (left part of Figure 7.4) and $[0.5001, 0.5099]$ (right part of Figure 7.4). We again took $f(z) = z^{-1/2}$ and $\boldsymbol{b} = \hat{\boldsymbol{e}}_1$ and used restart length $m = 20$. We see that in both cases the restarted harmonic Arnoldi method converges reasonably fast, while the standard Arnoldi method requires about twice as many iterations as the harmonic Arnoldi approach for the less well conditioned system.

Acknowledgment. The authors would like to thank Valeria Simoncini and Daniel B. Szyld for stimulating discussions which contributed to the improvement of a preliminary version of Lemma 4.1 and thus Theorem 4.3.

REFERENCES

[1] M. AFANASJEW, M. EIERMANN, O. G. ERNST, AND S. GÜTTEL, *A generalization of the steepest descent method for matrix functions*, Electron. Trans. Numer. Anal., 28 (2008), pp. 206–222.

[2] ———, *Implementation of a restarted Krylov subspace method for the evaluation of matrix functions*, Linear Algebra Appl., 429 (2008), pp. 2293–2314.

[3] B. BECKERMANN AND S. GÜTTEL, *Superlinear convergence of the rational Arnoldi method for the approximation of matrix functions*, Numer. Math., 121 (2012), pp. 205–236.

[4] C. BERG, *Stieltjes-Pick-Bernstein-Schoenberg and their connection to complete monotonicity*, in Positive Definite Functions. From Schoenberg to Space-Time Challenges, S. Mateu and E. Porcu, eds., Dept. of Mathematics, University Jaume I, Castellón de la Plana, Spain, 2008.

[5] C. BERG AND G. FORST, *Potential Theory on Locally Compact Abelian Groups*, Springer-Verlag, Heidelberg, 1975.

[6] C. DE BOOR, *Divided differences*, Surv. Approx. Theory, 1 (2005), pp. 46–69.

[7] V. DRUSKIN, *On monotonicity of the Lanczos approximation to the matrix exponential*, Linear Algebra Appl., 429 (2008), pp. 1679–1683.

[8] V. DRUSKIN AND L. KNIZHNERMAN, *Two polynomial methods of calculating functions of symmetric matrices*, Comput. Math. Math. Phys., 29 (1989), pp. 112–121.

[9] M. Eiermann and O. G. Ernst, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.
[10] M. Eiermann and O. G. Ernst, *A restarted Krylov subspace method for the evaluation of matrix functions*, SIAM J. Numer. Anal., 44 (2006), pp. 2481–2504.
[11] M. Eiermann, O. G. Ernst, and S. Güttel, *Deflated restarting for matrix functions*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 621–641.
[12] T. Ericsson, *Computing functions of matrices using Krylov subspace methods*, tech. report, Department of Computer Science, Chalmers University of Technology and the University of Göteborg, 1990.
[13] M. Freund and E. Görlich, *Polynomial approximation of entire functions and rate of growth of Taylor coefficients*, Proc. Edinb. Math. Soc., 28 (1985), pp. 341–348.
[14] A. Frommer, *Monotone convergence of the Lanczos approximations to matrix functions of Hermitian matrices*, Electron. Trans. Numer. Anal., 35 (2009), pp. 118–128.
[15] A. Frommer and U. Glässner, *Restarted GMRES for shifted linear systems*, SIAM J. Sci. Comput., 19 (1998), pp. 15–26.
[16] A. Frommer, S. Güttel, and M. Schweitzer, *Efficient and stable Arnoldi restarts for matrix functions based on quadrature*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 661–683.
[17] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, vol. 17 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[18] P. Henrici, *Applied and Computational Complex Analysis, Vol. 2*, John Wiley & Sons, New York, 1977.
[19] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, 2008.
[20] M. Hochbruck and M. E. Hochstenbach, *Subspace extraction for matrix functions*, preprint, Case Western Reserve University, Department of Mathematics, Cleveland, OH, 2005.
[21] M. Hochbruck and C. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.
[22] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.
[23] M. Ilić, I. W. Turner, and D. P. Simpson, *A restarted Lanczos approximation to functions of a symmetric matrix*, IMA J. Numer. Anal., 30 (2010), pp. 1044–1061.
[24] W. Joubert, *On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems*, Numer. Linear Algebra Appl., 1 (1994), pp. 427–447.
[25] L. Knizhnerman, *Calculation of functions of unsymmetric matrices using Arnoldi's method*, Comput. Math. Math. Phys., 31 (1991), pp. 1–9.
[26] R. B. Morgan, *Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1112–1135.
[27] ———, *GMRES with deflated restating*, SIAM J. Sci. Comput., 24 (2002), pp. 20–37.
[28] J. Niehoff, *Projektionsverfahren zur Approximation von Matrixfunktionen mit Anwendungen auf die Implementierung exponentieller Integratoren*, Dissertation, Mathematisch-Naturwissenschaftliche Fakultät der Heinrich–Heine Universität Düsseldorf, 2006.
[29] C. C. Paige, B. N. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.
[30] Y. Saad, *Analysis of some Krylov subspace approximations to the exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228.
[31] ———, *Iterative Methods for Sparse Linear Systems, 2nd ed.*, SIAM, 2003.
[32] Y. Saad and M. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
[33] M. Schweitzer, *Any cycle-convergence curve is possible for restarted FOM*, tech. report, in preparation, 2014.
[34] V. Simoncini, *Restarted full orthogonalization method for shifted linear systems*, BIT, 43 (2003), pp. 459–466.
[35] H. Tal-Ezer, *On restart and error estimation for Krylov approximation of $w = f(A)v$*, SIAM J. Sci. Comput., 29 (2007), pp. 2426–2441.
[36] J. van den Eshof, A. Frommer, T. Lippert, K. Schilling, and H. A. van der Vorst, *Numerical methods for the QCD overlap operator, I. Sign-function and error bounds*, Comput. Phys. Commun., 146 (2002), pp. 203–224.
[37] E. Vecharynski and J. Langou, *Any admissible cycle-convergence behavior is possible for restarted GMRES at its initial cycles*, Numer. Linear Algebra Appl., 18 (2011), pp. 499–511.
[38] I. Zavorin, D. P. OLeary, and H. Elman, *Complete stagnation of GMRES*, Linear Algebra Appl., 367 (2003), pp. 165–183.