



Krylov subspace methods for estimating operator-vector multiplications in Hilbert spaces

Yuka Hashimoto^{1,2} · Takashi Nodera³

Received: 17 April 2020 / Revised: 26 January 2021 / Accepted: 2 February 2021 /

Published online: 21 February 2021

© The Author(s) 2021

Abstract

The Krylov subspace method has been investigated and refined for approximating the behaviors of finite or infinite dimensional linear operators. It has been used for approximating eigenvalues, solutions of linear equations, and operator functions acting on vectors. Recently, for time-series data analysis, much attention is being paid to the Krylov subspace method as a viable method for estimating the multiplications of a vector by an unknown linear operator referred to as a transfer operator. In this paper, we investigate a convergence analysis for Krylov subspace methods for estimating operator-vector multiplications.

Keywords Krylov subspace method · Operator-vector multiplication · Unbounded operator · Transfer operator · Nonlinear dynamical system · Infinite dimensional Hilbert space

Mathematics Subject Classification 65F60 · 37M10

1 Introduction

Linear operators are used in various tasks in engineering and scientific research such as simulations and data analysis. A classical example of linear operators is a differential operator used for describing various natural phenomena. Krylov

✉ Yuka Hashimoto
yuka.hashimoto.rw@hco.ntt.co.jp; yukahashimoto@math.keio.ac.jp

Takashi Nodera
nodera@math.keio.ac.jp

¹ NTT Network Technology Laboratories, NTT Corporation, 3-9-11, Midori-cho, Musashino, Tokyo 180-8585, Japan

² School of Fundamental Science and Technology, Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa 223-8522, Japan

³ Department of Mathematics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa 223-8522, Japan

subspace methods have been actively researched in numerical linear algebra for approximating the behavior of such *given* operators, such as approximation of eigenvalues, solutions of linear equations, and operator functions acting on vectors, which provide approximations of solutions or information on the solutions of differential equations [16, 19, 24, 27, 31–33, 35, 36, 43, 45]. In many cases, problems in infinite dimensional spaces such as differential equations are discretized by, for example, a finite difference method [8] or finite element method [1] and are transformed into finite dimensional problems with matrices, after which the Krylov subspace methods are applied to the matrices. On the other hand, Krylov subspace methods for operators in infinite dimensional Hilbert spaces without discretization have also been investigated and more general results than those for matrices have been developed [6, 9–12, 15, 26, 28, 30].

Meanwhile, linear operators that represent time evolutions in dynamical systems, called transfer operators, are being investigated in relation to various fields such as machine learning, physics, molecular dynamics, and control engineering [18, 21, 25, 40–42]. *Unknown* transfer operators are estimated through data generated by dynamical systems. Since the transfer operators are generally linear even if the dynamical systems are nonlinear, Krylov subspace methods can be used to understand nonlinear dynamical systems [3, 4, 14, 20]. To make the algorithm computable only with the data, transfer operators are often discussed in relation to RKHSs (reproducing kernel Hilbert spaces). The Arnoldi and shift-invert Arnoldi methods have been proposed as Krylov subspace methods for transfer operators in RKHSs. The Arnoldi method is a standard Krylov subspace method [3, 35], but for its convergence, operators applied to it have to be bounded. However, not all transfer operators are bounded [17]. For example, transfer operators defined in the RKHS associated with the Gaussian kernel are unbounded if the dynamical system is nonlinear and deterministic. Thus, the shift-invert Arnoldi method was also considered [14]. When we apply the shift-invert Arnoldi method, a shifted and inverted operator $(\gamma I - K)^{-1}$ for some γ which is not contained in the spectrum of K was considered instead of an unbounded K .

The main difference between the classical settings assumed for the Krylov subspace methods mentioned in the first paragraph and those for transfer operators is whether the information of the model is given or not. In the classical setting, a differential operator is given, and a *model-driven* approach with the operator is applied. On the other hand, in the above setting for transfer operators, neither a dynamical system nor a transfer operator is given. Instead, data generated by the system is given. A *data-driven* approach is applied in this case. The purpose of applying the Krylov subspace method also differs in some situations. When we apply it to a transfer operator, denoted as K , one important task is to estimate $K^n v$ for a given vector v and some $n \in \mathbb{N}$ because K is unknown. On the other hand, in the classical setting for a given operator such as a differential operator, denoted as A , Av for a given vector v is already known, because both A and v are known. The main task of the Krylov subspace method in such a setting is to estimate $f(A)v$ for a given vector v and a function f such as $f(z) = z^{-1}$ and $f(z) = e^z$, except for $f(z) = z^n$. For this reason, Krylov subspace methods for estimating operator-vector multiplications have not been discussed in the classical setting of numerical

linear algebra. Meanwhile, although Krylov subspace methods for estimating operator-vector multiplications have been proposed in machine learning, the convergence analysis for them has not been fully investigated.

The objective of this paper is to analyze the convergence of such Krylov subspace methods for estimating operator-vector multiplications. We define a “residual” for approximating $K^n v$ for a vector v and analyze the convergence of the residuals of the Krylov approximations. The classical Krylov subspace methods for estimating $f(A)v$ are frequently associated with residuals. For $f(z) = z^{-1}$, for example, the GMRES (generalized minimal residual method) approximation minimizes the residual in a Krylov subspace and the convergence of the residual is superlinear [9, 28, 44]. Moreover, in BiCG (biconjugate gradient) type approximations, the residual or a value relevant to the residual is orthogonal to the Krylov subspace [43]. For a more general f , a generalized residual is proposed for evaluating the convergence of approximations [2, 13, 16, 34].

In our case, we show that the Arnoldi approximation converges to the minimizer of the residual. To illustrate this point, an error bound for a Krylov approximation of an operator function acting on a vector is used [10–12, 15, 27]. For the shift-invert Arnoldi method, the convergence analysis is not straightforward. At first glance, the problem of estimating $K^n v$ seems to be the same as that of estimating the operator function $f(K)$ acting on the vector v , where $f(z) = z^n$. However, the situation is different from that of the classical Krylov subspace methods in terms of operator functions. The existing error bound for the Krylov approximation of $f(K)v$ requires an assumption of the holomorphicity of f on the spectrum of K . On the other hand, the function $f(z) = (\gamma - z^{-1})^n$, where “ $f((\gamma I - K)^{-1}) = K^n$ ” holds formally, is not holomorphic at 0, but 0 is contained in the spectrum of $(\gamma I - K)^{-1}$ if K is unbounded. We resolve this problem through the factor K^{-n} that appears in the residual.

This paper is structured in the following manner. In Sect. 2, to explain why operator-vector multiplications need to be estimated for data analysis, we review the definition of a transfer operator and the Krylov subspace methods for it. In Sect. 3, we generalize the problem to Krylov approximations for estimating operator-vector multiplications for linear operators in a Hilbert space and investigate a convergence analysis. In Sect. 4, we empirically confirm the results investigated in Sect. 3. Section 5 is the conclusion.

1.1 Notations

Linear operators are denoted with standard capital letters, except for $m \times m$ matrices, which are denoted in bold. Calligraphic capital letters and Italicized Greek capital letters denote sets. The inner product and norm are denoted as $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively.

2 Background

In this section, we briefly review the definition of Perron–Frobenius operators and Krylov subspace methods for Perron–Frobenius operators [14, 20]. Perron–Frobenius operators are transfer operators often discussed in relation to RKHSs, and

their Krylov subspaces naturally appear [18, 20, 22]. The adjoint operators of Perron–Frobenius operators are referred to as the Koopman operators [23], which are also transfer operators and have been researched for data-driven approaches [3, 4, 21, 40, 42].

2.1 Perron–Frobenius operator in RKHS

Consider the following dynamical system with random noise [14]:

$$X_{t+1} = h(X_t) + \xi_t, \quad (1)$$

where $t \in \mathbb{Z}_{\geq 0}$, (Ω, \mathcal{F}) is a measurable space, $(\mathcal{X}, \mathcal{B})$ is a Borel measurable and locally compact Hausdorff vector space, X_t and ξ_t are random variables from Ω to \mathcal{X} , and $h : \mathcal{X} \rightarrow \mathcal{X}$ is a generally nonlinear map. Assume $\{\xi_t\}_{t \in \mathbb{Z}}$ is an i.i.d. stochastic process and that ξ_t is also independent of X_t . The random variable ξ_t corresponds to random noise in \mathcal{X} .

Let P be a probability measure in Ω . The nonlinear time evolution of X_t in a dynamical system (1) is regarded as a linear time evolution of the push forward measure $X_{t*}P$, defined by $X_{t*}P(B) = P(X_t^{-1}(B))$ for $B \in \mathcal{B}$. To describe the time evolution in a Hilbert space, RKHS [37] is used. An RKHS is a Hilbert space constructed by a map $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{C}$ called a positive definite kernel. For $x \in \mathcal{X}$, a map $\phi : \mathcal{X} \rightarrow \mathbb{C}^{\mathcal{X}}$ defined as $\phi(x) = k(x, \cdot)$ is called a feature map. Let $\mathcal{H}_{k,0}$ be a vector space defined as

$$\mathcal{H}_{k,0} = \left\{ \sum_{i=1}^n c_i \phi(x_i) \mid n \in \mathbb{N}, c_1, \dots, c_n \in \mathbb{C}, x_1, \dots, x_n \in \mathcal{X} \right\}.$$

In $\mathcal{H}_{k,0}$, the inner product associated with k is defined, and the completion of $\mathcal{H}_{k,0}$, which is denoted as \mathcal{H}_k , is called an RKHS. An observation $z \in \mathcal{X}$ is regarded as a vector $\phi(z)$ in \mathcal{H}_k through ϕ . Moreover, if k is bounded, continuous, and c_0 -universal, then the space of all the complex-valued finite regular Borel measures on \mathcal{X} , which is denoted as $\mathcal{M}(\mathcal{X})$, is densely embedded into \mathcal{H}_k . That is, a map $\Phi : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{H}_k$ defined as $\mu \mapsto \int_{x \in \mathcal{X}} \phi(x) d\mu(x)$ is injective [39] and $\Phi(\mathcal{M}(\mathcal{X}))$ is dense in \mathcal{H}_k [14]. Here, c_0 -universal means that \mathcal{H}_k is dense in the space of all continuous functions that vanish at infinity. The map Φ is called a kernel mean embedding [29]. For example, the Gaussian kernel $e^{-c\|x-y\|_2^2}$ and Laplacian kernel $e^{-c\|x-y\|_1}$ with $c > 0$ for $x, y \in \mathbb{R}^d$ are bounded and continuous c_0 -universal kernels. Therefore, a complex-valued finite regular Borel measure μ is regarded as a vector $\Phi(\mu)$ in the dense subset of Hilbert space \mathcal{H}_k . Since the map $\Phi : \mathcal{M}(\mathcal{X}) \rightarrow \mathcal{H}_k$ is linear, it is possible to define a linear operator $K : \Phi(\mathcal{M}(\mathcal{X})) \rightarrow \mathcal{H}_k$, which is called a *Perron–Frobenius operator*, in \mathcal{H}_k as follows:

$$K\Phi(\mu) = \Phi(\beta_{t*}(\mu \otimes P)), \quad (2)$$

where $\beta_t : \mathcal{X} \times \Omega \rightarrow \mathcal{X}$ is defined as $(x, \omega) \mapsto h(x) + \xi_t(\omega)$. Since ξ_t and X_t are independent, $\Phi(\beta_{t*}(X_{t*}P \otimes P)) = \Phi((h(X_t) + \xi_t)_*P)$ holds, and K maps $\Phi(X_{t*}P)$ to

$\Phi(X_{t+1*}P)$. In addition, since $\{\xi_t\}_{t \in \mathbb{Z}}$ is an i.i.d. process, it can be shown that K does not depend on t .

2.2 Krylov subspace methods for Perron–Frobenius operators

Let $\{x_0, x_1, \dots\} \subseteq \mathcal{X}$ be observed time-series data from the dynamical system (1), i.e., $x_t = X_t(\omega_0)$ for some $\omega_0 \in \Omega$. By using Krylov subspace methods, we estimate $K^n\phi(x_t)$ for $x_t \in \mathcal{X}$ to predict $\phi(x_{t+n})$ through available data. In the m th Krylov step, the data is split into S datasets. Examples of the choice for S are $S = m + 1$ and $S = M$ for a sufficiently large natural number M . Let $\mu_{t,N}^S = 1/N \sum_{i=0}^{N-1} \delta_{x_{t+iS}}$ ($t = 0, \dots, m$) be empirical measures with the datasets, where $N \in \mathbb{N}$ and δ_x denotes the Dirac measure at $x \in \mathcal{X}$. It is assumed that $\mu_{t,N}^S$ weakly converges to a finite regular Borel measure μ_t^S as $N \rightarrow \infty$ for $t = 0, \dots, m$.

A Krylov subspace is constructed with $\Phi(\mu_t^S)$. To construct the Krylov subspace only with the observed data $\{x_0, x_1, \dots\}$, the following equality of the average of noise ξ_t is assumed for any measurable and integrable function f :

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} \int_{\omega \in \Omega} f(h(x_{t+iS}) + \xi_t(\omega)) dP(\omega) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} f(h(x_{t+iS}) + \xi_{t+iS}(\eta)) \text{ a.s. } \eta \in \Omega. \end{aligned} \quad (3)$$

The left- and right-hand sides of Eq. (3) represent the space average and time average of ξ_t , respectively. The same types of assumptions as Eq. (3) have also been considered in other studies [4, 42]. By applying the above settings and assumptions, the Arnoldi and shift-invert Arnoldi approximations for the Perron–Frobenius operator K , defined as Eq. (2), are computed as explained in the following subsections.

2.3 The Arnoldi method

In this section, the Perron–Frobenius operator K is assumed to be bounded. Under the assumption (3), the following equation is derived since Φ is continuous:

$$\lim_{N \rightarrow \infty} K\Phi(\mu_{t,N}^S) = \Phi(\mu_{t+1}^S) \quad (t = 0, \dots, m-1). \quad (4)$$

Thus, if K is bounded, $K\Phi(\mu_t^S) = \Phi(\mu_{t+1}^S)$ holds. Therefore, if the set of the vectors $\{\Phi(\mu_0^S), \dots, \Phi(\mu_{m-1}^S)\}$ is linearly independent, the following space, denoted as $\mathcal{K}_m(K, \Phi(\mu_0^S))$, is an m -dimensional Krylov subspace of the operator K and vector $\Phi(\mu_0^S)$:

$$\mathcal{K}_m(K, \Phi(\mu_0^S)) = \text{Span}\{\Phi(\mu_0^S), \dots, \Phi(\mu_{m-1}^S)\}.$$

Remark 2.1 If S depends on m , the initial vector $\Phi(\mu_0^S)$ depends on m . In this case, the inclusion $\mathcal{K}_{m-1}(K, \Phi(\mu_0^S)) \subseteq \mathcal{K}_m(K, \Phi(\mu_0^S))$ does not always hold. On the other

hand, if S does not depend on m , $\Phi(\mu_0^S)$ does not depend on m and the inclusion $\mathcal{K}_{m-1}(K, \Phi(\mu_0^S)) \subseteq \mathcal{K}_m(K, \Phi(\mu_0^S))$ holds.

Let q_1, \dots, q_m be an orthonormal basis of the Krylov subspace $\mathcal{K}_m(K, \Phi(\mu_0^S))$ obtained through the Gram–Schmidt orthonormalization and $Q_m : \mathbb{C}^m \rightarrow \mathcal{H}_k$ be defined as $[c_1, \dots, c_m] \mapsto \sum_{i=1}^m c_i q_i$. Note that $Q_m Q_m^*$, where $*$ means adjoint, is a projection operator onto the Krylov subspace. There exists an invertible matrix $\mathbf{R}_m \in \mathbb{C}^{m \times m}$ such that $[\Phi(\mu_0^S), \dots, \Phi(\mu_{m-1}^S)] = Q_m \mathbf{R}_m$. This makes it possible to compute the following *Arnoldi approximation* of $K\phi(z)$ for an observable $z \in \mathcal{X}$ only with observed data $\{x_0, x_1, \dots\}$:

$$\begin{aligned} K\phi(z) &\approx Q_m Q_m^* K Q_m Q_m^* \phi(z) \\ &= Q_m Q_m^* [\Phi(\mu_1^S), \dots, \Phi(\mu_m^S)] \mathbf{R}_m^{-1} Q_m^* \phi(z) \\ &= Q_m \tilde{\mathbf{K}}_m Q_m^* \phi(z), \end{aligned}$$

where $\tilde{\mathbf{K}}_m = Q_m^* K Q_m = Q_m^* [\Phi(\mu_1^S), \dots, \Phi(\mu_m^S)] \mathbf{R}_m^{-1}$.

2.4 The shift-invert Arnoldi method

The convergence of the Arnoldi method along m is not always attained if K is unbounded [14]. According to Ikeda et al. [17], not all the Perron–Frobenius operators are bounded. For this reason, the shift-invert Arnoldi method is also considered.

Let $\gamma \notin \Lambda(K)$ be fixed, where $\Lambda(K)$ is the spectrum of K under the assumption of $\Lambda(K) \neq \mathbb{C}$, and consider using bounded bijective operator $(\gamma I - K)^{-1}$. Under the assumption (3), the following equation is derived:

$$\lim_{N \rightarrow \infty} (\gamma I - K)^{-1} u_{t+1, N}^S = u_t^S, \quad (5)$$

where $u_{t, N}^S = \sum_{i=0}^t \binom{t}{i} (-1)^i \gamma^{t-i} \Phi(\mu_{i, N}^S)$ and $u_t^S = \sum_{i=0}^t \binom{t}{i} (-1)^i \gamma^{t-i} \Phi(\mu_i^S)$. Since $(\gamma I - K)^{-1}$ is bounded, $(\gamma I - K)^{-1} u_{t+1}^S = u_t^S$ holds. Therefore, if the set of the vectors $\{u_1^S, \dots, u_m^S\}$ is linearly independent, then the space spanned by $\{u_1^S, \dots, u_m^S\}$ is an m -dimensional Krylov subspace of the operator $(\gamma I - K)^{-1}$ and vector u_m^S . Similar to the Arnoldi method, let q_1, \dots, q_m be an orthonormal basis of the Krylov subspace $\mathcal{K}_m((\gamma I - K)^{-1}, u_m^S)$ obtained through the Gram–Schmidt orthonormalization and $Q_m : \mathbb{C}^m \rightarrow \mathcal{H}_k$ be defined as $[c_1, \dots, c_m] \mapsto \sum_{i=1}^m c_i q_i$. There exists an invertible matrix $\mathbf{R}_m \in \mathbb{C}^{m \times m}$ satisfying $[u_1^S, \dots, u_m^S] = Q_m \mathbf{R}_m$. If K is unbounded, Kv for $v \in \mathcal{H}_k$ is not always defined. However, if $v \in \Phi(\mathcal{M}(\mathcal{X}))$, Kv is defined, in which case Kv is represented as $(\gamma I - ((\gamma I - K)^{-1})^{-1})v$. On the basis of this observation, the following shift-invert Arnoldi approximation of $K\phi(z)$ for $z \in \mathcal{X}$ is deduced if $\tilde{\mathbf{L}}_m$, which is defined as $\tilde{\mathbf{L}}_m = Q_m^* (\gamma I - K)^{-1} Q_m = Q_m^* [u_0^S, \dots, u_{m-1}^S] \mathbf{R}_m^{-1}$, is invertible:

$$\begin{aligned} K\phi(z) &\approx Q_m f_\gamma(Q_m^*(\gamma I - K)^{-1} Q_m) Q_m^* \phi(z) \\ &= Q_m f_\gamma(Q_m^*[u_0^S, \dots, u_{m-1}^S] \mathbf{R}_m^{-1}) Q_m^* \phi(z) \\ &= Q_m \tilde{\mathbf{K}}_m Q_m^* \phi(z), \end{aligned}$$

where $f_\gamma(z) = \gamma - z^{-1}$ for $z \in \mathbb{C}$ and $\tilde{\mathbf{K}}_m = f_\gamma(\tilde{\mathbf{L}}_m)$.

3 Convergence analysis

In this section, we provide a convergence analysis of the Arnoldi method and shift-invert Arnoldi method described in Sect. 2. The problem is generalized to a separable complex Hilbert space \mathcal{H} and linear operator K on \mathcal{H} by setting $v = \phi(z)$, $v_0 = \Phi(\mu_0^S)$, and $v_i = K^i v_0$ for $i = 1, \dots, m$.

In Sect. 3.1, we generalize the problem. In Sect. 3.2, we define a residual of an approximation of $K^n v$. Then, we investigate the relationship between the two methods and the residuals in Sects. 3.3 and 3.4.

3.1 The general setting for Krylov subspace methods for estimating operator-vector multiplications

Let \mathcal{H} be a separable complex Hilbert space, let $K : \mathcal{D} \rightarrow \mathcal{H}$ be an unknown linear map, where \mathcal{D} is a dense subset of \mathcal{H} , and let v and v_0 be given vectors in \mathcal{H} . We assume $K^i v_0 \in \mathcal{D}$ for any natural number i since by the definition of the Perron–Frobenius operator K , $Kv \in \mathcal{D}$ holds for $\mathcal{D} = \Phi(\mathcal{M}(\mathcal{X}))$. The purpose of the Krylov subspace method is to estimate $K^n v$ only with v, v_0, \dots, v_m , where $v_i = K^i v_0$.

Assume the dimension of $\mathcal{K}_m(K, v_0)$ is m . Let q_1, \dots, q_m be an orthonormal basis of the Krylov subspace $\mathcal{K}_m(K, v_0)$ obtained through the Gram–Schmidt orthonormalization and $Q_m : \mathbb{C}^m \rightarrow \mathcal{H}$ be defined as $[c_1, \dots, c_m] \mapsto \sum_{i=1}^m c_i q_i$. Then, there exists an invertible matrix $\mathbf{R}_m \in \mathbb{C}^{m \times m}$ such that $[v_0, \dots, v_{m-1}] = Q_m \mathbf{R}_m$. The *Arnoldi approximation* of $K^n v$, which is denoted as a_m^{Arnoldi} , is defined as

$$a_m^{\text{Arnoldi}} = Q_m \tilde{\mathbf{K}}_m^n Q_m^* v,$$

where $\tilde{\mathbf{K}}_m = Q_m^* K Q_m$, which can be represented as $Q_m^*[v_1, \dots, v_m] \mathbf{R}_m^{-1}$.

Analogously, let $\gamma \notin \Lambda(K)$, let q_1, \dots, q_m be an orthonormal basis of the Krylov subspace $\mathcal{K}_m((\gamma I - K)^{-1}, u_m)$ obtained through the Gram–Schmidt orthonormalization, where $u_m = \sum_{i=0}^m \binom{m}{i} (-1)^i \gamma^{m-i} v_i$, and let $Q_m : \mathbb{C}^m \rightarrow \mathcal{H}$ be defined as $[c_1, \dots, c_m] \mapsto \sum_{i=1}^m c_i q_i$. Then, there exists an invertible matrix $\mathbf{R}_m \in \mathbb{C}^{m \times m}$ such that $[u_1, \dots, u_m] = Q_m \mathbf{R}_m$. Let $\tilde{\mathbf{L}}_m = Q_m^* (\gamma I - K)^{-1} Q_m$, which can be represented as $Q_m^*[u_0, \dots, u_{m-1}] \mathbf{R}_m^{-1}$. If $\tilde{\mathbf{L}}_m$ is invertible, the *shift-invert Arnoldi approximation* of $K^n v$, which is denoted as a_m^{SIA} , is defined as

$$a_m^{\text{SIA}} = Q_m \tilde{\mathbf{K}}_m^n Q_m^* v,$$

where $\tilde{\mathbf{K}}_m = f_\gamma(\tilde{\mathbf{L}}_m)$ and $f_\gamma(z) = \gamma - z^{-1}$. Here, we give the different expression of $\tilde{\mathbf{K}}_m$ for the shift-invert Arnoldi method from the Arnoldi method.

3.2 A residual of an approximation of operator-vector multiplication

Assume $0 \notin \Lambda(K)$. We define a *residual* of an approximation a_m of $K^n v$ as follows:

$$\text{res}(a_m) = v - K^{-n} a_m. \quad (6)$$

Although the approximation error $K^n v - a_m$ is generally not available since $K^n v$ is unknown, $K^{-n} a_m$ is available in some cases. For example, if K is a Perron–Frobenius operator and we know past observations x_{-1}, \dots, x_{-n} , then we can calculate $K^{-n} \phi(\mu_t^\gamma)$ for $t = 0, \dots, m-1$. Then, we can also calculate $K^{-n} a_m^{\text{Arnoldi}}$. In fact, the residual (6) is a reasonable criterion for evaluating the convergence of the approximation for two reasons. First, the residual of an approximation a_m of the solution of a linear equation $Ax = b$ is defined as $b - Aa_m$. If the problem of approximating $K^n v$ is regarded as that of solving $K^{-n} x = v$, the residual of approximation a_m is $v - K^{-n} a_m$. Second, the following proposition shows that the value $v - K^{-n} a_m$ can be decomposed into a generalized residual of the Krylov approximation proposed by Saad [34] and Hochbruck et al. [16] and the error with respect to projecting v into a Krylov subspace.

Proposition 3.1 Assume $0 \notin \Lambda(K)$. Let $a_m = Q_m \tilde{\mathbf{K}}_m^n Q_m^* v$ be the Arnoldi or shift-invert Arnoldi approximation of $K^n v$ and let $f(z) = z^{-1}$. In addition, let r_m be the generalized residual of a_m with respect to $f(K^{-n})v$, i.e.,

$$r_m = \frac{1}{2\pi i} \int_{z \in \Gamma} f(z) ((zI - K^{-n}) Q_m (zI - \tilde{\mathbf{K}}_m^{-n})^{-1} Q_m^* v - v) dz,$$

where i is the imaginary unit and Γ is a rectifiable Jordan curve enclosing $\Lambda(\tilde{\mathbf{K}}_m^{-1})$ but not enclosing 0. Then, the residual of a_m defined as (6) is decomposed as follows:

$$\text{res}(a_m) = r_m + (v - Q_m Q_m^* v).$$

Proof Since $0 \notin \Lambda(K)$ and $\Lambda(\tilde{\mathbf{K}}_m) = \Lambda(Q_m^* K Q_m) \subseteq \Lambda(K)$ hold, we have $0 \notin \Lambda(\tilde{\mathbf{K}}_m)$. Thus, we obtain $0 \notin \Lambda(\tilde{\mathbf{K}}_m^{-1})$, and there exists a rectifiable Jordan curve Γ where f is holomorphic in the region enclosed by Γ and continuous on Γ . Therefore, $\int_{z \in \Gamma} f(z) dz = 0$, and by the Cauchy's integral formula, the following equalities are derived:

$$\begin{aligned} r_m &= \frac{1}{2\pi i} \int_{z \in \Gamma} f(z) ((zI - K^{-n}) Q_m (zI - \tilde{\mathbf{K}}_m^{-n})^{-1} Q_m^* v - v) dz \\ &= Q_m f(\tilde{\mathbf{K}}_m^{-n}) \tilde{\mathbf{K}}_m^{-n} Q_m^* v - K^{-n} Q_m f(\tilde{\mathbf{K}}_m^{-n}) Q_m^* v \\ &= Q_m Q_m^* v - K^{-n} a_m, \end{aligned}$$

which completes the proof of the proposition. \square

3.3 Convergence analysis for the Arnoldi method

In this section, we assume K is bounded and $0 \notin \Lambda(K)$.

The Arnoldi approximation $a_m^{\text{Arnoldi}} = Q_m \tilde{\mathbf{K}}_m^n Q_m^* v$ is obtained through two projections. First, the vector $v \in \mathcal{H}$ is projected onto the Krylov subspace $\mathcal{K}_m(K, v_0)$. Then, K acts on the projected vector in $\mathcal{K}_m(K, v_0)$ and is projected back to the Krylov subspace again. Note that we do not need the first projection in the classical Krylov subspace method for approximating $f(A)v$ for a given linear operator A , vector v , and function f since we can compute $A^i v$ for $i = 1, \dots, m-1$ and construct the Krylov subspace of A and v . On the other hand, we cannot construct the Krylov subspace of K and v in our current case since K is unknown and only $K^i v_0$, not $K^i v$, for $i = 1, \dots, m-1$ and a vector v_0 are given. This prevents us evaluating the convergence speed of the approximation error or residual directly since the convergence speed of the approximation depends on that of the projected vector $Q_m Q_m^* v$ to the original vector v . Therefore, we first consider the minimizer of the residual in a Krylov subspace and evaluate the difference between the Arnoldi approximation and minimizer.

In fact, since the projection $Q_m Q_m^*$ is orthogonal, the projected vector $Q_m Q_m^* v$ minimizes the difference from the original vector v , i.e.,

$$\arg \min_{u \in \mathcal{K}_m(K, v_0)} \|v - u\| = Q_m Q_m^* v. \quad (7)$$

Since each $u \in \mathcal{K}_m(K, v_n)$ satisfies $K^{-n}u \in \mathcal{K}_m(K, v_0)$, Eq. (7) implies that the inequality $\|v - K^{-n}\tilde{a}_m\| \leq \|v - K^{-n}u\|$ holds, where $\tilde{a}_m = K^n Q_m Q_m^* v \in \mathcal{K}_m(K, v_n)$. Therefore, \tilde{a}_m minimizes $\|v - K^{-n}u\|$ for all $u \in \mathcal{K}_m(K, v_n)$, i.e.,

$$\arg \min_{u \in \mathcal{K}_m(K, v_n)} \|v - K^{-n}u\| = \tilde{a}_m.$$

However, in practice, $\mathcal{K}_m(K, v_n)$ is unavailable only with v, v_0, \dots, v_m . Therefore, \tilde{a}_m is also unavailable. Thus, a_m^{Arnoldi} , instead of \tilde{a}_m is used for estimating $K^n v$.

We evaluate the difference between a_m^{Arnoldi} and \tilde{a}_m . Let $\mathbb{D}_\rho = \{z \in \mathbb{C} \mid |z| \leq \rho\}$ be the disk of diameter $\rho > 0$, let $\mathcal{W}(K) = \{\langle v, Kv \rangle \mid v \in \mathcal{D}, \|v\| = 1\}$ be the numerical range of K , and let $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ be the extended complex plane. Moreover, let α_ρ be a conformal map from $\mathbb{C} \setminus \mathcal{W}(K)$ to $\overline{\mathbb{C}} \setminus \mathbb{D}_\rho$ that satisfies $\alpha_\rho(\infty) = \infty$ and $\lim_{z \rightarrow \infty} \alpha_\rho(z)/z = 1$, and let Γ_r be the region enclosed by the contour $\{z \in \mathbb{C} \mid |\alpha_\rho(z)| = r\}$ for $r > \rho$. Here, $\mathcal{W}(K)$ is the closure of $\mathcal{W}(K)$ and by the Riemann mapping theorem, the map α_ρ exists. The following theorem is deduced.

Theorem 3.2 *Let $n < m$, and let p_{m-n-1} and \tilde{p}_{n-1} be polynomials of order $m-n-1$ and $n-1$ that satisfy $a_m^{\text{Arnoldi}} = K^n p_{m-n-1}(K)v_0 + \tilde{p}_{n-1}(K)v_0$. Assume the set $\{v_0, \dots, v_{m-1}\}$ is linearly independent. If the function f_m defined as $f_m(z) = z^{-n}\tilde{p}_{n-1}(z)$ is holomorphic in Γ_r , the residual of a_m^{Arnoldi} is evaluated as follows:*

$$\|\text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m)\| \leq 2C_1 C_2(m) \|v_0\| \frac{(\rho/r)^m}{1 - (\rho/r)},$$

where $C_1 > 0$ is a constant and $C_2(m) > 0$ depends on m .

We use the following lemma for deriving Theorem 3.2.

Lemma 3.3 *Let $n < m$. Assume $0 \notin \mathcal{W}(K)$ and the set $\{v_0, \dots, v_{m-1}\}$ is linearly independent. Then, the following equality is deduced:*

$$\text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m) = Q_m \tilde{\mathbf{K}}_m^{-n} \tilde{p}_{n-1}(\tilde{\mathbf{K}}_m) Q_m^* v_0 - K^{-n} \tilde{p}_{n-1}(K) v_0. \quad (8)$$

Proof The identity $p(K)v_0 = Q_m p(\tilde{\mathbf{K}}_m) Q_m^* v_0$ holds for any polynomial p of an order less than or equal to $m-1$. In addition, by the assumption of $0 \notin \mathcal{W}(K)$ and the inclusion $\mathcal{W}(\tilde{\mathbf{K}}_m) \subseteq \mathcal{W}(K)$, $\tilde{\mathbf{K}}_m$ is invertible. As a result, the following equalities are derived:

$$\begin{aligned} \text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m) &= K^{-n} \tilde{a}_m - K^{-n} a_m^{\text{Arnoldi}} \\ &= Q_m \tilde{\mathbf{K}}_m^{-n} \tilde{\mathbf{K}}_m^n Q_m^* v - (p_{m-n-1}(K)v_0 + K^{-n} \tilde{p}_{n-1}(K)v_0) \\ &= Q_m \tilde{\mathbf{K}}_m^{-n} Q_m^* (K^n p_{m-n-1}(K)v_0 + \tilde{p}_{n-1}(K)v_0) \\ &\quad - (p_{m-n-1}(K)v_0 + K^{-n} \tilde{p}_{n-1}(K)v_0) \\ &= Q_m \tilde{\mathbf{K}}_m^{-n} (\tilde{\mathbf{K}}_m^n p_{m-n-1}(\tilde{\mathbf{K}}_m) Q_m^* v_0 + \tilde{p}_{n-1}(\tilde{\mathbf{K}}_m) Q_m^* v_0) \\ &\quad - (Q_m p_{m-n-1}(\tilde{\mathbf{K}}_m) Q_m^* v_0 + K^{-n} \tilde{p}_{n-1}(K)v_0) \\ &= Q_m \tilde{\mathbf{K}}_m^{-n} \tilde{p}_{n-1}(\tilde{\mathbf{K}}_m) Q_m^* v_0 - K^{-n} \tilde{p}_{n-1}(K) v_0, \end{aligned}$$

which completes the proof of the lemma. \square

The vector $Q_m \tilde{\mathbf{K}}_m^{-n} \tilde{p}_{n-1}(\tilde{\mathbf{K}}_m) Q_m^* v_0$ in the right-hand side of Eq. (8) is equivalent to the Arnoldi approximation of the operator function $f_m(K)$ acting on the vector v_0 [12, 15]. Note that since \tilde{p}_{n-1} depends on m , f_m depends on m . By using this fact, we now prove Theorem 3.2.

Proof (Proof of Theorem 3.2) Let \mathcal{P}_{m-1} be the set of all polynomials of orders less than or equal to $m-1$. By Crouzeix et al. [5], the following bound is deduced for any $p \in \mathcal{P}_{m-1}$:

$$\begin{aligned} \|\text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m)\| &= \|Q_m f_m(\tilde{\mathbf{K}}_m) Q_m^* v_0 - f_m(K) v_0\| \\ &\leq \|Q_m f_m(\tilde{\mathbf{K}}_m) Q_m^* v_0 - Q_m p(\tilde{\mathbf{K}}_m) Q_m^* v_0\| + \|f_m(K) v_0 - p(K) v_0\| \\ &\leq 2C_1 \|v_0\| \|f_m - p\|_{\infty, \mathcal{W}(K)}, \end{aligned}$$

where $0 < C_1 \leq 1 + \sqrt{2}$. In addition, for a linear operator K and a map $f : \mathbb{C} \rightarrow \mathbb{C}$ that is holomorphic in the interior of $\mathcal{W}(K)$ and continuous in $\mathcal{W}(K)$, the norm $\|f\|_{\infty, \mathcal{W}(K)}$ is defined as $\|f\|_{\infty, \mathcal{W}(K)} = \sup_{z \in \mathcal{W}(K)} |f(z)|$. By taking the infimum among $p \in \mathcal{P}_{m-1}$, we obtain

$$\|\text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m)\| \leq 2C_1 \|v_0\| \inf_{p \in \mathcal{P}_{m-1}} \|f_m - p\|_{\infty, \mathcal{W}(K)}. \quad (9)$$

In fact, the infimum in the inequality (9) can be taken among $p \in \{\tilde{p} \in \mathcal{P}_{m-1} \mid \|\tilde{p}\|_{\infty, \mathcal{W}(K)} \leq 2\|f_m\|_{\infty, \mathcal{W}(K)}\}$, which is a compact space. Indeed, for a polynomial $p \in \mathcal{P}_{m-1}$ satisfying $\|p\|_{\infty, \mathcal{W}(K)} > 2\|f_m\|_{\infty, \mathcal{W}(K)}$, we have

$$\|f_m - p\|_{\infty, \mathcal{W}(K)} > \|f_m\|_{\infty, \mathcal{W}(K)} = \|f_m - 0\|_{\infty, \mathcal{W}(K)},$$

and $0 \in \mathcal{P}_{m-1}$. Therefore, the infimum in the inequality (9) can be replaced with the minimum. By Ellacott [7, Corollary 2.2], this factor is bounded as

$$\min_{p \in \mathcal{P}_{m-1}} \|f_m - p\|_{\infty, \mathcal{W}(K)} \leq C_2(m) \frac{(\rho/r)^m}{1 - (\rho/r)},$$

where $C_2(m) = \max_{z \in \Gamma_r} |f_m(z)|$, which completes the proof of the theorem. \square

In fact, the following proposition guarantees the order of the increase in the factor $C_2(m)$ is at least m in the case where the initial vector v_0 does not depend on m (see Remark 2.1). Thus, in this case, the Arnoldi approximation a_m^{Arnoldi} approaches \tilde{a}_m , the minimizer of the residual, in the order of $m\alpha^m$ for some $0 < \alpha < 1$.

Proposition 3.4 *Assume the set $\{v_0, \dots, v_{m-1}\}$ is linearly independent and v_0 does not depend on m . If the function f_m is holomorphic in Γ_r , then the factor $C_2(m)$ is bounded as*

$$C_2(m) \leq C_2(1) + (m-1)C_3,$$

for some constant $C_3 > 0$ does not depend on m .

Proof We first evaluate the coefficients of the polynomial \tilde{p}_{n-1} . Since p_{m-n-1} and \tilde{p}_{n-1} depend on m , we denote them as p_{m-n-1}^m and \tilde{p}_{n-1}^m in this proof. Let $\tilde{p}_{n-1}^m(z) = \sum_{i=0}^{n-1} c_i(m)z^i$ and $p_{m-n-1}^m(z) = \sum_{i=n}^{m-1} c_i(m)z^{i-n}$, where $c_i(m) \in \mathbb{C}$. Then, by the definitions of \tilde{p}_{n-1}^m and p_{m-n-1}^m , we have

$$\begin{aligned} \|a_m^{\text{Arnoldi}} - a_{m-1}^{\text{Arnoldi}}\| &= \left\| \sum_{i=0}^{m-1} c_i(m)v_i - \sum_{i=0}^{m-2} c_i(m-1)v_i \right\| \\ &\geq \left| \left\langle \tilde{q}_i, \left(\sum_{i=0}^{m-1} c_i(m)v_i - \sum_{i=0}^{m-2} c_i(m-1)v_i \right) \right\rangle \right| = |\langle \tilde{q}_i, v_i \rangle| |c_i(m) - c_i(m-1)|, \end{aligned}$$

for $i = 0, \dots, n$, where \tilde{q}_i is a normalized vector in the orthogonal complement of the space spanned by $\{v_0, \dots, v_{i-1}, v_{i+1}, \dots\}$. In addition, we have

$$\|a_m^{\text{Arnoldi}} - a_{m-1}^{\text{Arnoldi}}\| = \|\mathcal{Q}_m \tilde{\mathbf{K}}_m^n \mathcal{Q}_m^* v - \mathcal{Q}_{m-1} \tilde{\mathbf{K}}_{m-1}^n \mathcal{Q}_{m-1}^* v\| \leq 2\|K^n\| \|v\|.$$

As a result, the following inequality is derived:

$$|c_i(m) - c_i(m-1)| \leq \frac{2\|K^n\| \|v\|}{|\langle \tilde{q}_i, v_i \rangle|}. \quad (10)$$

We now evaluate $C_2(m)$. By the inequality (10) and the holomorphicity of f_m , we obtain

$$\begin{aligned}
 C_2(m) &= \sup_{z \in \Gamma_r} |z^{-n} \tilde{p}_n^m(z)| \\
 &\leq \sup_{z \in \Gamma_r} |z^{-n} \tilde{p}_n^m(z) - z^{-n} \tilde{p}_n^{m-1}(z)| + \sup_{z \in \Gamma_r} |z^{-n} \tilde{p}_n^{m-1}(z)| \\
 &= \sup_{z \in \Gamma_r} \left| \sum_{i=0}^n z^{-n+i} (c_i(m) - c_i(m-1)) \right| + C_2(m-1) \\
 &\leq C_3 + C(m-1),
 \end{aligned} \tag{11}$$

where

$$C_3 = \sum_{i=1}^n \frac{2\|K^n\|\|v\|}{|\langle \tilde{q}_i, v_i \rangle|} \sup_{z \in \Gamma_r} |z^{-n+i}|.$$

Applying the inequality (11) recursively completes the proof of the proposition. \square

The decrease in the value $\|\text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m)\|$ is confirmed numerically in Sect. 4.2.

3.4 Convergence analysis for the shift-invert Arnoldi method

The convergence of the Arnoldi method is not guaranteed when K is unbounded. Moreover, although Theorem 3.2 requires the assumption about the numerical range of K , it is generally hard to calculate the numerical range of K , a linear operator in an infinite dimensional space. Therefore, we also consider the shift-invert Arnoldi method.

The shift-invert Arnoldi approximation $a_m^{\text{SIA}} = Q_m \tilde{K}_m^n Q_m^* v$ can also be obtained through two projections similar to the Arnoldi method. However, in this case, instead of K , the polynomial of $(\gamma I - K)^{-1}$ that approximates K acts on the vector which is the projection of v onto $\mathcal{K}_m((\gamma I - K)^{-1}, u_m)$.

Let $n < m$. To address K^{-n} in the residual, we slightly modify the Krylov subspace $\mathcal{K}_m((\gamma I - K)^{-1}, u_m)$ and define a space $\tilde{\mathcal{K}}_m((\gamma I - K)^{-1}, w_{m-n})$ as follows:

$$\begin{aligned}
 \tilde{\mathcal{K}}_m((\gamma I - K)^{-1}, w_{m-n}) &:= \text{Span}\{w_1, \dots, w_{m-n}, K^{-1}w_{m-n}, \dots, K^{-n}w_{m-n}\} \\
 &= \text{Span}\{(\gamma I - K)^{-m+n+1}w_{m-n}, \dots, (\gamma I - K)^{-1}w_{m-n}, \\
 &\quad w_{m-n}, K^{-1}w_{m-n}, \dots, K^{-n}w_{m-n}\} \\
 &= \text{Span}\{(\gamma I - K)^{-m+n+1}K^{-n}w_{m-n}, \dots, (\gamma I - K)^{-1}K^{-n}w_{m-n}, \\
 &\quad w_{m-n}, K^{-1}w_{m-n}, \dots, K^{-n}w_{m-n}\} \\
 &= \text{Span}\{K^{-n}w_1, \dots, K^{-n}w_{m-n-1}, w_{m-n}, K^{-1}w_{m-n}, \dots, K^{-n}w_{m-n}\},
 \end{aligned}$$

where $w_t = \sum_{i=0}^t \binom{t}{i} (-1)^i \gamma^{t-i} v_{i+n}$. The third equality is obtained because by applying the partial fraction decomposition principle, $(\gamma I - K)^i K^{-n} w_{m-n}$ for $i = -1, \dots, -m + n + 1$ is represented as a linear combination of $(\gamma I - K)^i w_{m-n}, \dots, (\gamma I - K)^{-1} w_{m-n}, K^{-n} w_{m-n}, \dots, K^{-1} w_{m-n}$ and the coefficient of $(\gamma I - K)^i w_{m-n}$ is not 0. Then, $(\gamma I - K)^{-1} w_{t+1} = w_t$ holds. Assume the dimension of $\tilde{\mathcal{K}}_m((\gamma I - K)^{-1}, w_{m-n})$ is m . Let q_1, \dots, q_m be the orthonormal basis of the space $\tilde{\mathcal{K}}_m((\gamma I - K)^{-1}, w_{m-n})$ obtained through the Gram–Schmidt orthonormalization and let $Q_m : \mathbb{C}^m \rightarrow \mathcal{H}$ be defined as $[c_1, \dots, c_m] \mapsto \sum_{i=1}^m c_i q_i$. Then, there exists an invertible matrix $\mathbf{R}_m \in \mathbb{C}^{m \times m}$ such that $[w_1, \dots, w_{m-n}, K^{-1} w_{m-n}, \dots, K^{-n} w_{m-n}] = Q_m \mathbf{R}_m$. Let $a_m^{\text{SIA}} = Q_m \tilde{\mathbf{K}}_m^n Q_m^* v$, where $\tilde{\mathbf{K}}_m = f_\gamma(\tilde{\mathbf{L}}_m)$ and $\tilde{\mathbf{L}}_m = Q_m^* (\gamma I - K)^{-1} Q_m$.

Since the projection $Q_m Q_m^*$ is orthogonal, the projected vector $Q_m Q_m^* v$ minimizes the difference from the original vector v , i.e.,

$$\arg \min_{u \in \tilde{\mathcal{K}}_m((\gamma I - K)^{-1}, w_{m-n})} \|v - u\| = Q_m Q_m^* v. \quad (12)$$

Since $(\gamma I - K)^i K^{-n} w_{m-n}$ can be represented as a linear combination of $K^{-n+i} w_{m-n}, \dots, K^{-n} w_{m-n}$, for $i = 1, \dots, n$, we have

$$K^{-n} w_{m-n+i} = (\gamma I - K)^i K^{-n} w_{m-n} \in \tilde{\mathcal{K}}_m((\gamma I - K)^{-1}, w_{m-n}).$$

Thus, each $u \in \text{Span}\{w_1, \dots, w_m\}$ satisfies $K^{-n} u \in \tilde{\mathcal{K}}_m((\gamma I - K)^{-1}, w_{m-n})$. Therefore, Eq. (12) implies that $\|v - K^{-n} \tilde{a}_m\| \leq \|v - K^{-n} u\|$ holds, where $\tilde{a}_m = K^n Q_m Q_m^* v$. As a result, \tilde{a}_m minimizes $\|v - K^{-n} u\|$ for all $u \in \text{Span}\{w_1, \dots, w_m\}$, i.e.,

$$\arg \min_{u \in \text{Span}\{w_1, \dots, w_m\}} \|v - K^{-n} u\| = \tilde{a}_m.$$

However, in practice, $\text{Span}\{w_1, \dots, w_m\}$ is unavailable only with v, v_0, \dots, v_m . Therefore \tilde{a}_m is also unavailable. Thus, a_m^{SIA} , instead of \tilde{a}_m , is used for estimating $K^n v$.

Concerning the difference between a_m^{SIA} and \tilde{a}_m , the following theorem is deduced. Here, r , ρ , and Γ_r are defined in the same manner as those for the Arnoldi approximation by replacing $\mathcal{W}(K)$ with $\mathcal{W}((\gamma I - K)^{-1})$.

Theorem 3.5 *Let n satisfy $m \geq 2n + 1$, and let p_{m-n-1} be a polynomial of order $m - n - 1$ and \tilde{p}_n be a polynomial of order n without the term of order 0 that satisfies $a_m^{\text{SIA}} = p_{m-n-1}((\gamma I - K)^{-1}) w_{m-n} + \tilde{p}_n(K^{-1}) w_{m-n}$. Assume $0 \notin \Lambda(K)$ and the set $\{w_1, \dots, w_{m-n}, K^{-1} w_{m-n}, \dots, K^{-n} w_{m-n}\}$ is linearly independent. If $g_\gamma(z) := z^n(\gamma z - 1)^{-n}$ is holomorphic in Γ_r , the residual of a_m^{SIA} is evaluated as*

$$\|\text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m)\| \leq 2C \|\tilde{p}_n(K^{-1}) w_{m-n}\| \frac{(\rho/r)^{m-n}}{1 - (\rho/r)},$$

where $C > 0$ is a constant.

We use the following lemma for deriving Theorem 3.5.

Lemma 3.6 Let n satisfy $m \geq 2n + 1$. Assume $0 \notin \Lambda(K)$ and the set $\{w_1, \dots, w_{m-n}, K^{-1}w_{m-n}, \dots, K^{-n}w_{m-n}\}$ is linearly independent. If $1/\gamma \notin \mathcal{W}((\gamma I - K)^{-1})$, the following equality is deduced:

$$\begin{aligned} & \text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m) \\ &= Q_m g_\gamma(\tilde{\mathbf{L}}_m) Q_m^* \tilde{p}_n(K^{-1}) w_{m-n} - K^{-n} \tilde{p}_n(K^{-1}) w_{m-n}. \end{aligned} \quad (13)$$

Proof Let $\hat{w}_{m-n} = K^{-n} w_{m-n}$.

Then, we have

$$\begin{aligned} (\gamma I - K)^{-n} K^n \hat{w}_{m-n} &= (\gamma I - K)^{-n} (\gamma I - (\gamma I - K))^n \hat{w}_{m-n} \\ &= (\gamma(\gamma I - K)^{-1} - I)^n \hat{w}_{m-n}. \end{aligned}$$

Since $p((\gamma I - K)^{-1}) w_{m-n} = Q_m p(\tilde{\mathbf{L}}_m) Q_m^* w_{m-n}$ and $p((\gamma I - K)^{-1}) \hat{w}_{m-n} = Q_m p(\tilde{\mathbf{L}}_m) Q_m^* \hat{w}_{m-n}$ holds for any polynomial p of an order less than or equal to n , the following equality is deduced:

$$\tilde{\mathbf{L}}_m^n Q_m^* w_{m-n} = (\gamma \tilde{\mathbf{L}}_m - I)^n Q_m^* \hat{w}_{m-n}. \quad (14)$$

By the assumption of $1/\gamma \notin \mathcal{W}((\gamma I - K)^{-1})$ and the inclusion $\mathcal{W}(\tilde{\mathbf{L}}_m) \subseteq \mathcal{W}((\gamma I - K)^{-1})$, $\gamma \tilde{\mathbf{L}}_m - I$ is invertible. Therefore, by Eq. (14), the following equalities are deduced:

$$\begin{aligned} & \text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m) = K^{-n} \tilde{a}_m - K^{-n} a_m^{\text{SIA}} \\ &= Q_m g_\gamma(\tilde{\mathbf{L}}_m) \tilde{\mathbf{K}}_m^n Q_m^* v - K^{-n} (p_{m-n-1}((\gamma I - K)^{-1}) w_{m-n} + \tilde{p}_n(K^{-1}) w_{m-n}) \\ &= Q_m g_\gamma(\tilde{\mathbf{L}}_m) Q_m^* (p_{m-n-1}((\gamma I - K)^{-1}) w_{m-n} + \tilde{p}_n(K^{-1}) w_{m-n}) \\ &\quad - (p_{m-n-1}((\gamma I - K)^{-1}) K^{-n} w_{m-n} + K^{-n} \tilde{p}_n(K^{-1}) w_{m-n}) \\ &= Q_m g_\gamma(\tilde{\mathbf{L}}_m) (p_{m-n-1}(\tilde{\mathbf{L}}_m) Q_m^* w_{m-n} + Q_m^* \tilde{p}_n(K^{-1}) w_{m-n}) \\ &\quad - (Q_m p_{m-n-1}(\tilde{\mathbf{L}}_m) g_\gamma(\tilde{\mathbf{L}}_m) Q_m^* w_{m-n} + K^{-n} \tilde{p}_n(K^{-1}) w_{m-n}) \\ &= Q_m g_\gamma(\tilde{\mathbf{L}}_m) Q_m^* \tilde{p}_n(K^{-1}) w_{m-n} - K^{-n} \tilde{p}_n(K^{-1}) w_{m-n}, \end{aligned}$$

which completes the proof of the lemma. \square

The vector $Q_m g_\gamma(\tilde{\mathbf{L}}_m) Q_m^* \tilde{p}_n(K^{-1}) w_{m-n}$ in the right-hand side of Eq. (13) is equivalent to the shift-invert Arnoldi approximation of the operator function $g_\gamma((\gamma I - K)^{-1})$ acting on the vector $\tilde{p}_n(K^{-1}) w_{m-n}$. In the same manner as the Arnoldi method, Theorem 3.5 is proved as follows.

Proof (Proof of Theorem 3.5) By Crouzeix et al. [5] and Ellacott [7, Corollary 2.2], the following bound is deduced:

$$\begin{aligned}
& \|Q_m g_\gamma(\tilde{\mathbf{L}}_m) Q_m^* \tilde{p}_n(K^{-1}) w_{m-n} - K^{-n} \tilde{p}_n(K^{-1}) w_{m-n}\| \\
& \leq 2C_1 C_2 \|\tilde{p}_n(K^{-1}) w_{m-n}\| \min_{p \in \mathcal{P}_{m-n-1}} \|g_\gamma - p\|_{\infty, \mathcal{W}((\gamma I - K)^{-1})} \\
& \leq 2C_1 C_2 \|\tilde{p}_n(K^{-1}) w_{m-n}\| \frac{(\rho/r)^{m-n}}{1 - (\rho/r)},
\end{aligned}$$

where $0 < C_1 \leq 1 + \sqrt{2}$ and $C_2 = \max_{z \in \Gamma_r} |g_\gamma(z)|$, which completes the proof of the theorem. \square

Remark 3.7 If K is unbounded, 0 is contained in $A((\gamma I - K)^{-1})$. Since $\Lambda((\gamma I - K)^{-1}) \subseteq \mathcal{W}((\gamma I - K)^{-1}) \subseteq \Gamma_r$ holds, 0 is contained in Γ_r . This is why g_γ , which is always holomorphic at 0 and holomorphic in \mathbb{C} when $\gamma = 0$, is used instead of f_γ , which is not holomorphic at 0, for evaluating Eq. (13).

Remark 3.8 We can choose r arbitrary as long as g_γ is holomorphic in Γ_r . The choice of r corresponds to a trade-off between the decay rate ρ/r of $\|\text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m)\|$ and the magnitude of the constant C_2 . Indeed, the decay rate ρ/r is small if r is large. On the other hand, the larger r becomes, the smaller the distance between Γ_r and $1/\gamma$, the singular point of g_γ , becomes. Thus, the larger r becomes, the larger $C_2 = \max_{z \in \Gamma_r} |g_\gamma(z)|$ becomes.

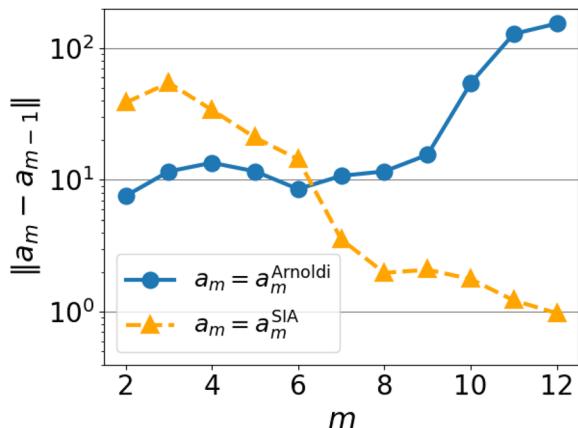
As a result, if γ is chosen so that g_γ is holomorphic in Γ_r , and if the factor $\|\tilde{p}_n(K^{-1}) w_{m-n}\|$ is bounded by some constant, the difference between residuals of a_m^{SIA} and \tilde{a}_m , the minimizer of the residual, exponentially decays as m becomes larger. Unfortunately, evaluating $\|\tilde{p}_n(K^{-1}) w_{m-n}\|$ theoretically is a challenging task. Thus, we numerically confirm that the factor $\|\tilde{p}_n(K^{-1}) w_{m-n}\|$ is bounded by a constant in Sect. 4.2.

4 Numerical experiments

Several typical numerical experiments are implemented in this section. These experiments are in a collection of problems to illustrate that the shift-invert Arnoldi method performs better than the Arnoldi method and to confirm the results in Theorems 3.2 and 3.5 numerically. All numerical computations of these experiments are executed with Python 3.6.

All the experiments are under the setting described in Sect. 2. Therefore, Krylov subspaces are subspaces of RKHSs and discrepancies are measured by norms in the RKHSs. In practical computations, all μ_i^S s in the algorithms are replaced by $\mu_{i,N}^S$ for $N \in \mathbb{N}$. The convergence of the approximation constructed with $\mu_{i,N}^S$ to the one constructed with μ_i^S is shown in [14, Section 4.3].

Fig. 1 The convergence behavior of a_m along the dimensions of Krylov subspace m



4.1 Comparison between the Arnoldi and shift-invert Arnoldi methods

To illustrate that the shift-invert Arnoldi method performs better than the Arnoldi method, the following dynamical system is considered in $\mathcal{X} \subseteq \mathbb{R}$:

$$X_t = 0.99X_{t-1} \cos(0.1X_{t-1}) + \xi_t, \quad (15)$$

where $X_0 = 0.5$ and ξ_t ($t = 0, 1, \dots$) are independent random variables with the Gaussian distribution with a mean of 0 and a standard deviation of 0.01. For $n = 1$, $N = 50$, $m = 2, 3, \dots, 12$, and $v = \phi(x_{1600})$, the discrepancy $\|a_m - a_{m-1}\|$, where $a_m = a_m^{\text{Arnoldi}}$ or $a_m = a_m^{\text{SIA}}$, are computed. The vector v_0 is set as $\Phi(\mu_{0,N}^S)$ as defined in Sect. 2.2. In addition, we set $\gamma = 1 + i$, $k(x, y) = e^{-\|x-y\|_2^2}$ and $S = m + 1$. The results of the mean values with 100 time-series data randomly generated by Eq. (15) are illustrated in Fig. 1. If K is bounded, and if the Krylov subspace converges to the dense subset of \mathcal{H}_k , $Q_m \tilde{K}_m Q_m^*$ converges strongly to K . Therefore, a_m converges to Kv in \mathcal{H}_k , and $\|a_m - a_{m-1}\|$ decreases as m becomes larger in this case. However, in Fig. 1, $\|a_m - a_{m-1}\|$ does not decrease as m becomes larger with the Arnoldi method. This is due to the unboundedness of K . On the other hand, that with the shift-invert Arnoldi method decreases as m becomes larger. The results indicate that the shift-invert Arnoldi method can address the unboundedness of K and is a better choice than the Arnoldi method in this case.

4.2 Confirmation of Theorems 3.2 and 3.5

The following deterministic dynamical system on $\mathcal{X} = \mathbb{R}$ is considered:

$$X_t = 0.99X_{t-1} + 0.2, \quad (16)$$

where $X_0 = 1$. In this example, we set $n = 1$, $N = 10$, $S = 15$, $v_0 = \Phi(\mu_{0,N}^S)$, and $v = \phi(x_{300})$.

To confirm the result of Theorem 3.5, we first check the assumptions of Theorem 3.5 and the behavior of $\|\tilde{p}_n(K^{-1})w_{m-n}\|$. Regarding the condition $0 \notin \Lambda(K)$, the function $h(x) = 0.99x + 0.2$ is bijective, and there is the inverse map h^{-1} of h , defined as $h^{-1}(x) = (x - 0.2)/0.99$. Therefore, the linear operator L defined as $L\phi(x) = \phi(h^{-1}(x))$ is the inverse of the Perron–Frobenius operator K with respect to the dynamical system (16), i.e., $L = K^{-1}$. Let $k(x, y) = e^{-|x-y|}$ be the Laplacian kernel and we put $\eta = 1/0.99$, the coefficient of x in h^{-1} . Then, K^{-1} is bounded. Indeed, for $v \in \mathcal{H}_k$, there exists a sequence $\{v_l\}_{l=1}^{\infty} \subseteq \mathcal{H}_{k,0}$ (see Sect. 2.1) such that $v = \lim_{l \rightarrow \infty} v_l$. Let v_l be represented as $v_l = \sum_{i=1}^{n(l)} \phi(x_i(l))c_i(l)$ for some $x_i(l) \in \mathcal{X}$ and $c_i(l) \in \mathbb{C}$. By the identity $e^{-|x|} = \int_{\omega \in \mathbb{R}} e^{-ix\omega} 4/(1+\omega^2)d\omega$ [38, Table 2], we have

$$\begin{aligned} \|K^{-1}v_l - K^{-1}v_{l'}\|^2 &= \int_{\omega \in \mathbb{R}} G(l, l') \frac{4}{1+\omega^2/\eta^2} d\omega \\ &\leq \int_{\omega \in \mathbb{R}} G(l, l') \frac{4\eta^2}{1+\omega^2} d\omega = \eta^2 \left\| \sum_{i=1}^{n(l)} \phi(x_i(l))c_i(l) - \sum_{i=1}^{n(l')} \phi(x_i(l'))c_i(l') \right\|^2 \\ &= \eta^2 \|v_l - v_{l'}\|^2, \end{aligned} \quad (17)$$

where

$$G(l, l') = \left| \sum_{i=1}^{n(l)} c_i(l) e^{-ix_i(l)\omega} - \sum_{i=1}^{n(l')} c_i(l') e^{-ix_i(l')\omega} \right|^2.$$

Since $\{v_l\}_{l=1}^{\infty}$ is a Cauchy sequence, Eq. (17) implies $\{K^{-1}v_l\}_{l=1}^{\infty}$ is a Cauchy sequence. Thus, the sequence $\{K^{-1}v_l\}_{l=1}^{\infty}$ converges. Therefore, K^{-1} is bounded.

Concerning the condition about the linearly independence, the orbit of the dynamical system (16) monotonically increases as i becomes large. Thus, the observations x_0, x_1, \dots satisfy $x_i \neq x_j$ for $i \neq j$. Therefore, since k is set as the Laplacian kernel, the set $\{\phi(x_0), \phi(x_1), \dots\}$ is linearly independent. As a result, the condition about the linearly independence is satisfied since v_0 is represented as $v_0 = \Phi(\mu_{0,N}^S) = 1/N \sum_{i=0}^{N-1} \phi(x_{iS})$ and by the definition of w_i .

Regarding the holomorphicity of g_{γ} , it has a singular point at $1/\gamma$. Moreover, by the Cauchy–Schwarz inequality, the following inequality holds for $v \in \mathcal{H}_k$, $\|v\| = 1$:

$$|\langle v, (\gamma I - K)^{-1}v \rangle| \leq \|(\gamma I - K)^{-1}\|.$$

Since the function $H(\gamma) := \|(\gamma I - K)^{-1}\|$ is continuous on $\mathbb{C} \setminus \Lambda(K)$, we consider $\|K^{-1}\|$ instead of $\|(\gamma I - K)^{-1}\|$ for $\gamma \approx 0$. By Eq. (17), $\|K^{-1}\| \leq \eta$ holds. Thus, for a sufficiently small γ , we have $\|(\gamma I - K)^{-1}\| \lesssim \eta$. This implies the numerical range $\mathcal{W}((\gamma I - K)^{-1})$ is contained in the ball \mathbb{D}_{ρ} where $\rho \approx \eta$. As a result, if we set γ as a sufficiently small value, $\rho \approx \eta$, and $\rho < r \lesssim 1/|\gamma|$, then g_{γ} is holomorphic in Γ_r .

As for the behavior of $\|\tilde{p}_n(K^{-1})w_{m-1}\|$, the following inequality is derived since $\tilde{p}_n(z) = c_m z$ for some $c_m \in \mathbb{C}$ in the case of $n = 1$:

$$\|\tilde{p}_n(K^{-1})w_{m-1}\| \leq |c_m| \|K^{-1}\| \|w_{m-1}\|.$$

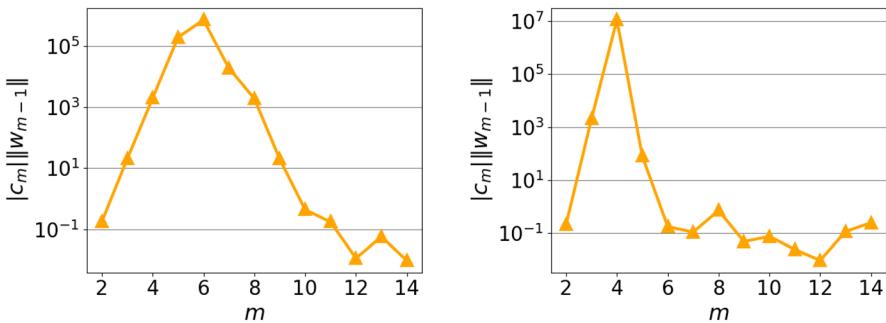


Fig. 2 The dimension of Krylov subspace m versus $|c_m| ||w_{m-1}||$ (left: $\gamma = 10^{-2}$, right: $\gamma = 10^{-4}$)

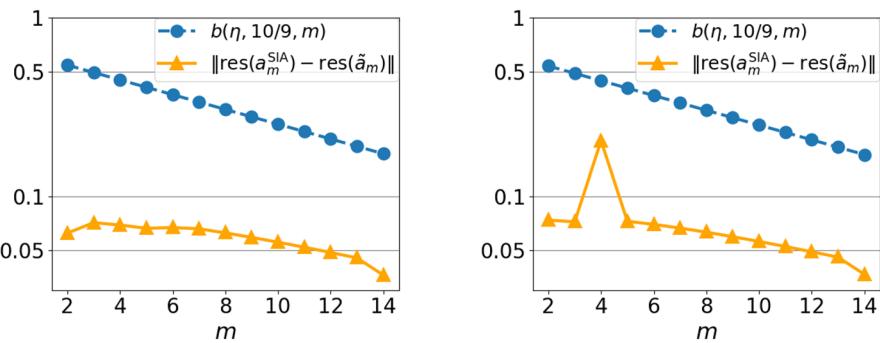


Fig. 3 The dimension of Krylov subspace m versus $\|\text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m)\|$ (left: $\gamma = 10^{-2}$, right: $\gamma = 10^{-4}$)

Thus, the behavior of $\|\tilde{p}_n(K^{-1})w_{m-n}\|$ is described by that of $|c_m| ||w_{m-1}||$. Figure 2 shows the value $|c_m| ||w_{m-1}||$ along m for $\gamma = 10^{-2}$ and $\gamma = 10^{-4}$. We can see $\|\tilde{p}_n(K^{-1})w_{m-1}\|$ are bounded by around 0.1 for sufficiently large m in both cases.

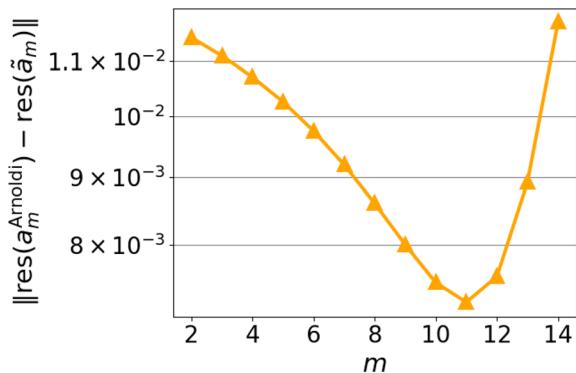
Next, we consider the decay rate of $\|\text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m)\|$. We can set $\rho \approx \eta$ and $\rho < r \lesssim 1/|\gamma|$. However, as we mentioned in Remark 3.8, choosing r as a larger value makes the constant C_2 in Theorem 3.5 larger. Figure 3 illustrates the value $\|\text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m)\|$ along m . We can see the decay rate is around 9/10. Thus, we set $\rho = \eta \approx 1$ and $r = 10/9$, and computed the following theoretical upper bound in accordance with the proof of Theorem 3.5:

$$b(\rho, r, m) := 2(1 + \sqrt{2})C_2(r) \cdot 0.1 \frac{(\rho/r)^{m-1}}{1 - (\rho/r)},$$

where $C_2(r) = \max_{z \in \Gamma_r} |g_\gamma(z)|$. We can see the theoretical upper bound describes the decay of the value $\|\text{res}(a_m^{\text{SIA}}) - \text{res}(\tilde{a}_m)\|$ correctly.

Finally, we confirm Theorem 3.2. In the same manner as the numerical range of K^{-1} , that of K is shown to be contained in the ball $\mathbb{D}_{0.99}$. Unfortunately, this evaluation is not sufficient for checking the holomorphicity of f_m defined in Theorem 3.2 and determine ρ and r since f_m has a singular point at 0. Figure 4 illustrates the

Fig. 4 The dimension of Krylov subspace m versus $\|\text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m)\|$



value $\|\text{res}(a_m^{\text{Arnoldi}}) - \text{res}(\tilde{a}_m)\|$ along m . It decays for small m , but grows for larger m , which implies the assumption of the holomorphicity of f_m in Theorem 3.2 would not be satisfied for this example.

4.3 Confirmation of the decrease in the residuals

To confirm the decrease in the residuals of the Arnoldi and shift-invert Arnoldi approximations, experiments with synthetic data generated by the Landau equation and real-world data are implemented.

4.3.1 Experiments with data generated by the Landau equation

The following Landau equation in $\mathcal{X} \subseteq [0, \infty)$ [4] is considered:

$$\frac{dr}{dt} = 0.5r - r^3. \quad (18)$$

Discretizing Eq. (18) and adding random noise results in the following discrete dynamical system:

$$X_t = X_{t-1} + \Delta t(0.5X_{t-1} - X_{t-1}^3 + \xi_t), \quad (19)$$

where $X_0 = 1$ and ξ_t ($t = 0, 1, \dots$) are independent random variables with the Gaussian distribution with a mean of 0 and a standard deviation of 0.01. For $n = 1, N = 50$, $m = 2, 3, \dots, 12$, $v_0 = \Phi(\mu_{0,N}^S)$, and $v = \phi(x_{1600})$, the residual $\|\text{res}(a_m)\|$, where $a_m = a_m^{\text{Arnoldi}}$ or $a_m = a_m^{\text{SIA}}$ are computed. In addition, we set $\Delta t = 0.01$, $\gamma = 1 + i$, $k(x, y) = e^{-|x-y|}$, and $S = m + 1, 15$. The computational results of mean values with 100 time-series data randomly generated by Eq. (19) are shown in Fig. 5. The results indicate that the residual $\|\text{res}(a_m)\|$ decreases as m increases, both with the Arnoldi

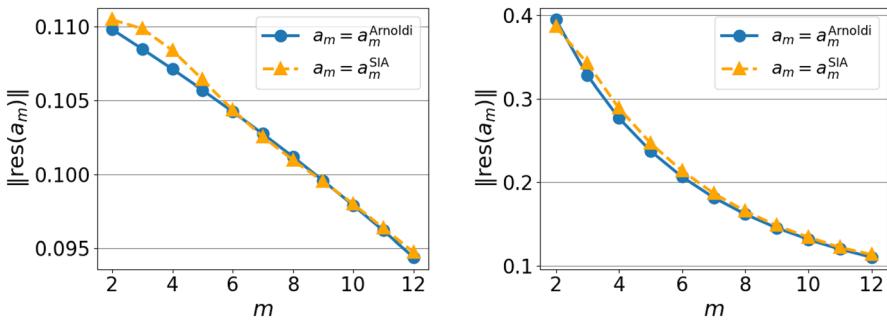
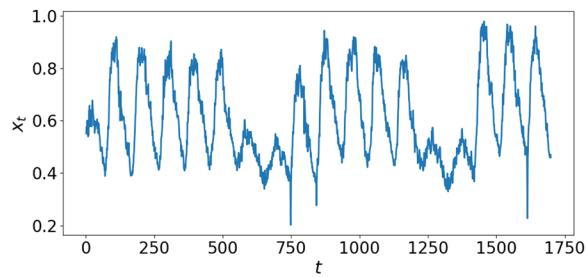


Fig. 5 The dimension of Krylov subspace m versus the residual $\| \text{res}(a_m) \|$ (left: $S = 15$, right: $S = m + 1$)

Fig. 6 Real-world teletraffic data at each t



and shift-invert Arnoldi methods and regardless of whether S depends on m or not. Although the residuals where $S = m + 1$ are larger than those where $S = 15$, they decrease faster than those where $S = 15$.

4.3.2 Experiments with real-world data

For the last experiment, real-world teletraffic data,¹ as shown in Fig. 6, is used. For $t = 0, 1, \dots$, x_t represents the amount of teletraffic (gbps) that passes through a certain node (ID 12) in a network composed of 23 nodes and 227 links at time t . The time width is 15 seconds. To extract the relationship between x_{t+1} and x_{t-p+1}, \dots, x_t , x_t is considered to be generated by a random variable Y_t , and $X_t = [Y_t, \dots, Y_{t-p+1}]$ is set for obtaining the relation (1) with $\mathcal{X} \subseteq \mathbb{R}^p$. For $n = 1, 2$, $N = 50$, $p = 1, 3$, $m = 2, 3, \dots, 12$, $v_0 = \Phi(\mu_{0,N}^S)$, and $v = \phi([x_{1600}, \dots, x_{1600+p-1}])$, the residual $\| \text{res}(a_m) \|$, where $a_m = a_m^{\text{Arnoldi}}$ or $a_m = a_m^{\text{SIA}}$, are computed. In addition, we set $\gamma = 1 + i$, $k(x, y) = e^{-|x-y|}$, and $S = 15$. The results are shown in Fig. 7. For $p = 1$, the residual $\| \text{res}(a_m) \|$ does not always decrease when m becomes larger with either the Arnoldi or shift-invert Arnoldi method. This implies that the relationship between x_{t+1} and x_t does not describe the behavior of the data fully. For $p = 3$, the residual of

¹ Available at <http://totem.run.montefiore.ulg.ac.be/files/data/traffic-matrices-anonymized-v2.tar.bz2>.

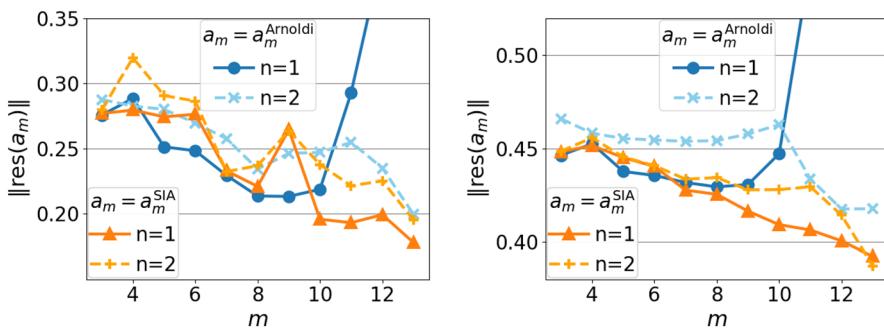


Fig. 7 The dimension of Krylov subspace m versus the residual $\|\text{res}(a_m)\|$ (left: $p = 1$, right: $p = 3$)

the shift-invert Arnoldi method decreases as m becomes larger, whereas that of the Arnoldi method does not decrease even if m becomes larger. This implies that the relationship between x_{t+1} and x_t, \dots, x_{t-2} describes the behavior of the data, but the unboundedness of the Perron–Frobenius operator prevents the Arnoldi method from constructing a proper approximation from the data or the assumption about the holomorphicity of f_m in Theorem 3.2 is not satisfied. The advantage of the shift-invert Arnoldi method over the Arnoldi method is also underlined in this example.

5 Conclusion

We investigated a convergence analysis for Krylov subspace methods for estimating operator-vector multiplications $K^n v$, where K is a linear operator and v is a vector in a Hilbert space, in this paper. The Arnoldi method and shift-invert Arnoldi method were considered. Although these methods have been proposed for time-series data analysis in machine learning, their convergence has not been thoroughly analyzed. We proved that the Arnoldi approximation converges to the minimizer of the residual. For the shift-invert Arnoldi method, the derivation was not straightforward since the function that appears in the estimated operator-vector multiplications was not holomorphic. This problem was addressed by applying the factor K^{-n} that appears in the residual. As a result, we showed the shift-invert Arnoldi approximation converges to the minimizer of the residual on the assumption that a factor related to the initial vector and K^{-1} is bounded by some constant. The aforementioned results were also confirmed numerically with synthetic and real-world data.

As future work, we will theoretically evaluate the aforementioned factor and give sufficient conditions for the convergence of the shift-invert Arnoldi approximation.

Acknowledgements We would like to thank the anonymous referees, whose comments helped improve the manuscript significantly, and editor Tomohiro Sogabe.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alberty, J., Carstensen, C., Funken, S.A.: Remarks around 50 lines of Matlab: short finite element implementation. *Numer. Algorithms* **20**, 117–137 (1999)
- Botchev, M.A., Grimm, V., Hochbruck, M.: Residual, restarting, and Richardson iteration for the matrix exponential. *SIAM J. Sci. Comput.* **35**(3), A1376–A1397 (2013)
- Budišić, M., Mohr, R., Mezić, I.: Applied Koopmanism. *Chaos* **22**, 047510 (2012)
- Črnjarić-Žic, N., Maćešić, S., Mezić, I.: Koopman operator spectrum for random dynamical systems. *J. Nonlinear Sci.* **20**, 20 (2019)
- Crouzeix, M., Palencia, C.: The numerical range is a $(1 + \sqrt{2})$ -spectral set. *SIAM J. Matrix Anal. Appl.* **38**(2), 649–655 (2017)
- Daniel, J.W.: The conjugate gradient method for linear and nonlinear operator equations. *SIAM J. Numer. Anal.* **4**(1), 10–26 (1967)
- Ellacott, S.W.: Computation of faber series with application to numerical polynomial approximation in the complex plane. *Math. Comput.* **40**(162), 575–587 (1983)
- Forsythe, G.E., Wasow, W.R.: Finite-Difference Methods for Partial Differential Equations. Wiley, New York (1960)
- Gilles, M.A., Townsend, A.: Continuous analogues of Krylov subspace methods for differential operators. *SIAM J. Numer. Anal.* **57**(2), 899–924 (2019)
- Göckler, T.: Rational Krylov subspace methods for ' ϕ '-functions in exponential integrators. PhD thesis, Karlsruher Institut für Technologie (2014)
- Grimm, V.: Resolvent Krylov subspace approximation to operator functions. *BIT Numer. Math.* **52**, 639–659 (2012)
- Güttel, S.: Rational Krylov methods for operator functions. PhD thesis, Technische Universität Bergakademie Freiberg (2010)
- Hashimoto, Y., Nodera, T.: Inexact shift-invert Arnoldi method for evolution equations. *ANZIAM J.* **58**(E), E1–E27 (2016)
- Hashimoto, Y., Ishikawa, I., Ikeda, M., Matsuo, Y., Kawahara, Y.: Krylov subspace method for nonlinear dynamical systems with random noise. *J. Mach. Learn. Res.* **21**(172), 1–29 (2020)
- Hashimoto, Y., Nodera, T.: Shift-invert rational Krylov method for an operator ϕ -function of an unbounded linear operator. *Jpn. J. Ind. Appl. Math.* **36**, 421–433 (2019)
- Hochbruck, M., Lubich, C., Selhofer, H.: Exponential integrators for large systems of differential equations. *SIAM J. Sci. Comput.* **19**(5), 1552–1574 (1998)
- Ikeda, M., Ishikawa, I., Sawano, Y.: Composition operators on reproducing kernel Hilbert spaces with analytic positive definite functions. [arXiv:1911.11992](https://arxiv.org/abs/1911.11992) (2019)
- Ishikawa, I., Fujii, K., Ikeda, M., Hashimoto, Y., Kawahara, Y.: Metric on nonlinear dynamical systems with Perron–Frobenius operators. *Adv. Neural Inf. Process. Syst.* **31**, 2856–2866 (2018)
- Kadioğlu, M., Mudrick, S.: On the implementation of the GMRES(m) method to elliptic equations in meteorology. *J. Comput. Phys.* **102**(2), 348–359 (1992)
- Kawahara, Y.: Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis. *Adv. Neural Inf. Process. Syst.* **29**, 911–919 (2016)
- Klus, S., Nüske, F., Peitz, S., Niemann, J.-H., Clementi, C., Schütte, C.: Data-driven approximation of the Koopman generator: model reduction, system identification, and control. *Phys. D* **406**, 132416 (2020)
- Klus, S., Schuster, I., Muandet, K.: Eigen decompositions of transfer operators in reproducing kernel Hilbert spaces. *J. Nonlinear Sci.* **30**, 283–315 (2020)
- Koopman, B.O.: Hamiltonian systems and transformation in Hilbert space. *Proc. NatL. Acad. Sci.* **17**(5), 315–318 (1931)

24. Krylov, A.N.: On the numerical solution of the equation by which in technical questions frequencies of small oscillations of material systems are determined. *Izvestija AN SSSR.* **7**(4), 491–539 (1931). (in Russian)
25. Lusch, B., Nathan Kutz, J., Brunton, S.: Deep learning for universal linear embeddings of nonlinear dynamics. *Nat. Commun.* **9**, 4950 (2018)
26. Málek, J., Strakoš, Z.: Preconditioning and the conjugate gradient method in the context of solving PDEs. *SIAM Spotlights* **20**, 20 (2015)
27. Moret, I., Novati, P.: RD-rational approximations of the matrix exponential. *BIT Numer. Math.* **44**, 595–615 (2004)
28. Moret, I.: A note on the superlinear convergence of GMRES. *SIAM J. Numer. Anal.* **34**(2), 513–516 (1997)
29. Muandet, K., Fukumizu, K., Sriperumbudur, B.K., Schölkopf, B.: Kernel mean embedding of distributions: a review and beyond. *Found. Trends Mach. Learn.* **20**, 20 (2017)
30. Olver, S.: GMRES for the differentiation operator. *SIAM J. Numer. Anal.* **47**(5), 3359–3373 (2009)
31. Hestenes, M.R., Stiefel, E.: Methods of conjugate gradients for solving linear systems. *J. Res. Natl Bureau Std.* **49**(6), 409–436 (1952)
32. Ruhe, A.: Rational Krylov: a practical algorithm for large sparse nonsymmetric matrix pencils. *SIAM J. Sci. Comput.* **19**(5), 1535–1551 (1998)
33. Ruhe, A.: Rational Krylov for real pencils with complex eigenvalues. *Taiwan. J. Math.* **14**(3A), 795–803 (2010)
34. Saad, Y.: Analysis of some Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.* **29**(1), 209–228 (1992)
35. Saad, Y.: Iterative Methods for Sparse Linear Systems, 2nd edn. SIAM, Philadelphia (2003)
36. Saad, Y., Schultz, M.H.: GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.* **7**(3), 856–869 (1986)
37. Saitoh, S., Sawano, Y.: Theory of Reproducing Kernels and Applications. Springer, Berlin (2016)
38. Sriperumbudur, B.K., Gretton, A., Fukumizu, K., Schölkopf, B., Lanckriet, G.R.G.: Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **11**, 1517–1561 (2010)
39. Sriperumbudur, B.K., Fukumizu, K., Lanckriet, G.R.G.: Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12**, 2389–2410 (2011)
40. Susuki, Y., Mezic, I., Raak, F., Hikihara, T.: Applied Koopman operator theory for power systems technology. *Nonlinear Theory Appl. IEICE.* **7**(4), 430–459 (2016)
41. Takeishi, N., Kawahara, Y., Yairi, T.: Learning Koopman invariant subspaces for dynamic mode decomposition. *Adv. Neural Inf. Process. Syst.* **30**, 1131–1141 (2017)
42. Takeishi, N., Kawahara, Y., Yairi, T.: Subspace dynamic mode decomposition for stochastic Koopman analysis. *Phys. Rev. E* **96**, 033310 (2017)
43. Van der Vorst, H.A.: Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.* **13**(2), 631–644 (1992)
44. Van der Vorst, H.A., Vuik, C.: The superlinear convergence behaviour of GMRES. *J. Comput. Appl. Math.* **48**(3), 327–341 (1993)
45. Van der Vorst, H.A.: Iterative Krylov Methods for Large Linear Systems. Cambridge University Press, Cambridge (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.