ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Mathematics and Computers in Simulation ▮ (▮▮▮▮) ▮▮▮–▮▮▮

MATHEMATICS AND COMPUTERS IN SIMULATION

www.elsevier.com/locate/matcom

Original articles

Analysis of inexact Krylov subspace methods for approximating the matrix exponential

Khanh N. Dinh, Roger B. Sidje*

Department of Mathematics, The University of Alabama, Tuscaloosa, AL 35487, USA

Received 26 September 2015; received in revised form 24 May 2016; accepted 2 January 2017

Available online xxxx

Abstract

Krylov subspace methods have proved quite effective at approximating the action of a large sparse matrix exponential on a vector. Their numerical robustness and matrix-free nature have enabled them to make inroads into a variety of applications. A case in point is solving the chemical master equation (CME) in the context of modeling biochemical reactions in biological cells. This is a challenging problem that gives rise to an extremely large matrix due to the curse of dimensionality. Inexact Krylov subspace methods that build on truncation techniques have helped solve some CME models that were considered computationally out of reach as recently as a few years ago. However, as models grow, truncating them means using an even smaller fraction of their whole extent, thereby introducing more inexactness. But experimental evidence suggests an apparent success and the aim of this study is to give theoretical insights into the reasons why. Essentially, we show that the truncation can be put in the framework of inexact Krylov methods that relax matrix–vector products and compute them expediently by trading accuracy for speed. This allows us to analyze both the residual (or defect) and the error of the resulting approximations to the matrix exponential from the viewpoint of inexact Krylov methods.

Keywords: Matrix exponential; Inexact Krylov method; Chemical master equation

1. Introduction

Given a large sparse nonsymmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and vector $\boldsymbol{p}_0 \in \mathbb{R}^n$, letting $\boldsymbol{v} = \boldsymbol{p}_0$ and taking $m \ll n$ Arnoldi steps with a starting vector $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|$, where $\|\cdot\|$ means the 2-norm, we obtain an orthonormal basis $\boldsymbol{V}_m = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m] \in \mathbb{R}^{n\times m}$ of the Krylov subspace $\mathcal{K}_m(\boldsymbol{A}, \boldsymbol{v}) = \text{span}\{\boldsymbol{v}, \boldsymbol{A}\boldsymbol{v}, \ldots, \boldsymbol{A}^{m-1}\boldsymbol{v}\}$, and an upper Hessenberg matrix $\boldsymbol{H}_m \in \mathbb{R}^{m\times m}$ that satisfy

$$\boldsymbol{A}\boldsymbol{V}_m = \boldsymbol{V}_{m+1}\bar{\boldsymbol{H}}_m = \boldsymbol{V}_m\boldsymbol{H}_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T, \tag{1a}$$

$$\boldsymbol{H}_m = \boldsymbol{V}_m^T\boldsymbol{A}\boldsymbol{V}_m, \tag{1b}$$

* Corresponding author.
E-mail addresses: kdinh@crimson.ua.edu (K.N. Dinh), roger.b.sidje@ua.edu (R.B. Sidje).

http://dx.doi.org/10.1016/j.matcom.2017.01.002

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

where $\boldsymbol{e}_m = (0, \ldots, 0, 1)^T$, and $\bar{\boldsymbol{H}}_m \in \mathbb{R}^{(m+1)\times m}$ is $\boldsymbol{H}_m$ augmented with $h_{m+1,m}\boldsymbol{e}_m^T$ under its last row. The standard Krylov approximation to the matrix exponential takes the form

$$\exp(\tau \boldsymbol{A})\boldsymbol{v} \approx \boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1, \quad \boldsymbol{e}_1 = (1, 0, \ldots, 0)^T, \ \beta = \|\boldsymbol{v}\|. \tag{2}$$

It is well-known that (1) is also the cornerstone for building very efficient Krylov subspace solution techniques for other problems such as eigenvalue problems or linear systems. In the latter, there has been recent interest in transitioning from exact to inexact (or relaxed) matrix–vector products in the Arnoldi process [3,4,15], either out of necessity or deliberately, trading accuracy for speed. It is customary to model these inexact products as

$$\boldsymbol{A}\boldsymbol{v}_k \approx (\boldsymbol{A} + \boldsymbol{E}_k)\boldsymbol{v}_k, \tag{3}$$

where $\boldsymbol{E}_k$ is some error matrix that varies at each invocation, and note that setting $\boldsymbol{E}_k = \boldsymbol{0}$ recovers the exact evaluation. To make the difference clear, we refer to the classical method as the exact Arnoldi and it is not meant to imply exact arithmetic. The foremost implication of such a relaxation is that the classical Arnoldi relationship (1) does not hold anymore, but Simoncini and Szyld [15] made the key observation that we end up with

$$(\boldsymbol{A} + \boldsymbol{\mathcal{E}}_m)\boldsymbol{V}_m = \boldsymbol{V}_m \boldsymbol{H}_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T, \quad \boldsymbol{\mathcal{E}}_m = \sum_{k=1}^{m} \boldsymbol{E}_k \boldsymbol{v}_k \boldsymbol{v}_k^T, \tag{4}$$

which is similar to (1), except that $\boldsymbol{V}_m$, which still remains orthonormal, is now a basis of a Krylov subspace obtained by a perturbed $\boldsymbol{A}$. When we use the computed $\boldsymbol{V}_m$ and $\boldsymbol{H}_m$ from (4) in GMRES for instance, classical error bounds do not apply anymore. However, from theoretical and experimental evidence (such as [14]), the method can withstand cases where the norm of the perturbation $\boldsymbol{\mathcal{E}}_m$ grows quite large.

The analysis of Simoncini and Szyld [15] provided insights into inexact GMRES for solving a linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, but it has so far remained unclear how inexactness affects the Krylov approximation (2). Since we now have (4) instead of (1), we also lose classical error bounds on the matrix exponential (e.g., Gallopoulos and Saad [5], Saad [11], Hochbruck and Lubich [9]). Thus our study fills a gap in the literature by looking at the error in the inexact Krylov counterpart of (2). We additionally offer another related way of assessing the accuracy by investigating the defect or residual [2] from the fact that (2) arises when solving a system of linear ODEs of the form

$$\begin{cases} \boldsymbol{p}'(t) = \boldsymbol{A}\boldsymbol{p}, & t \in [0, \ t_f] \\ \boldsymbol{p}(0) = \boldsymbol{p}_0, & \text{initial condition.} \end{cases} \tag{5}$$

It is worth recalling that, in the exact case, the effectiveness of approximating $\exp(\boldsymbol{A})\boldsymbol{v}$ by projecting it onto $\mathcal{K}_m(\boldsymbol{A}, \boldsymbol{v})$ hinges on the fact that all polynomials of $\boldsymbol{A}$ of degree $\leq m - 1$ can be calculated exactly through $\boldsymbol{H}_m$, or more precisely,

$$q_{m-1}(\boldsymbol{A})\boldsymbol{v} = \boldsymbol{V}_m q_{m-1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1,$$

where q_{m-1} is any polynomial of degree $\leq m - 1$. By the same reasoning as in the exact case (e.g., Saad [11, Lemma 3.1]), it can be shown from (4) that, for the same polynomial q_{m-1},

$$q_{m-1}(\boldsymbol{A} + \boldsymbol{\mathcal{E}}_m)\boldsymbol{v} = \boldsymbol{V}_m q_{m-1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1.$$

The significance of all this is that the inexact Krylov subspace method for $\mathcal{K}_m(\boldsymbol{A}, \boldsymbol{v})$ with the relaxation matrices $\boldsymbol{E}_k$, $k = 1, \ldots, m$, can be seen as the exact Krylov subspace method for $\mathcal{K}_m(\tilde{\boldsymbol{A}}, \boldsymbol{v})$ with

$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{\mathcal{E}}_m = \boldsymbol{A} + \sum_{k=1}^{m} \boldsymbol{E}_k \boldsymbol{v}_k \boldsymbol{v}_k^T,$$

which is another simple way to understand the method.

The rest of the paper is organized as follows: Section 2 gives some background on modeling biochemical reactions and on the finite state projection (FSP) algorithm for solving the underlying chemical master equation (CME), which was the challenging application that initially motivated the research presented here. Section 3 analyzes the residual (or defect) of the inexact Krylov method both when the ODE problem is homogeneous or nonhomogeneous, with two different approaches considered for the latter. Section 4 analyzes the error and includes a special treatment that

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

exploits the structure of the matrix when it arises from stochastic processes such as that involved in the CME. Section 5 reports some numerical experiments. Section 6 finally wraps the presentation with some concluding remarks.

2. Inexact chemical master equation — a motivation

In a biological cell containing different molecular species undergoing various chemical reactions, the state is a vector of integers counting the different species of molecules. Such a discrete formulation is prompted by key regulatory molecules that exist in small numbers, making a continuous formulation (via concentrations) inappropriate. The counter of a species goes up or down when a chemical reaction occurs, depending on whether the reaction produces or consumes that species. Starting from a particular state, the cell will transition to different states as reactions happen. The CME has the form (5) and depicts the evolution of the system's probability distribution, which characterizes the probability of finding the cell in a given state at a given time. The challenge here is that even with simple biochemical models having 4 or 5 reaction channels and a relatively low count of each molecular species, there can be millions of possible states, and the variety of models means that calculations cannot rely on generic simplifications.

The finite state projection (FSP) algorithm of Munsky and Khammash [10] is a model reduction method to cope with the huge size of the CME, and we outline it here because it vividly illustrates how inexactness comes into play. With $J = \{1, \ldots, k\}$, and k being the cardinality of J, let

$$\boldsymbol{A} = \left(\begin{array}{c|c} \boldsymbol{A}_J & * \\ \hline * & * \end{array}\right) \in \mathbb{R}^{n\times n},$$

i.e., $\boldsymbol{A}_J$ is a $k \times k$ submatrix of the true operator $\boldsymbol{A}$. The FSP algorithm takes

$$\boldsymbol{p}(t_f) = \exp(t_f\boldsymbol{A})\boldsymbol{p}_0 \approx \begin{pmatrix} \exp(t_f\boldsymbol{A}_J) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{p}_J(0) \\ \boldsymbol{0} \end{pmatrix}. \tag{6}$$

Note that $\boldsymbol{p}_0 = \boldsymbol{p}(0)$ is also truncated according to J. Munsky and Khammash [10] assessed the loss of the probability mass and gave a theoretical justification of the merit of this approach from a probabilistic point of view. Our study recasts the analysis in the framework of inexact methods.

With J now an arbitrary subset of $\{1, \ldots, n\}$, and the corresponding submatrix $\boldsymbol{A}_J$ padded with zeros as necessary, the 'truncation' can be formalized as

$$\boldsymbol{A} = \boldsymbol{A}_J + \boldsymbol{R}_J,$$

where $\boldsymbol{R}_J$ is the error matrix from the truncation. Then (6) is in turn further approximated by an inexact Krylov method for the matrix exponential, which we saw previously means that the Arnoldi process for approximating the solution uses inexact (or relaxed) matrix–vector products of the form

$$\boldsymbol{A}_J\boldsymbol{v} \approx (\boldsymbol{A}_J + \boldsymbol{S}_J)\boldsymbol{v},$$

where $\boldsymbol{S}_J$ models some error in the product. Combining with the truncation, we get

$$\boldsymbol{A}\boldsymbol{v} \approx (\boldsymbol{A} - \boldsymbol{R}_J + \boldsymbol{S}_J)\boldsymbol{v},$$

so that $\boldsymbol{E} = -\boldsymbol{R}_J + \boldsymbol{S}_J$ captures both error terms. In particular, if $\boldsymbol{S}_J = 0$, only the truncation error is in effect, whereas if $\boldsymbol{R}_J = 0$, there will only be the error from the relaxed matrix–vector product. From now on, our theoretical analysis will simply assume that the true matrix $\boldsymbol{A}$ interacts through the inexact evaluation $\boldsymbol{A}\boldsymbol{v}_k \approx (\boldsymbol{A} + \boldsymbol{E}_k)\boldsymbol{v}_k$ without regard to the real source of the error.

3. Bounds on the residual

Given the differential equation (5), consider an approximation $\boldsymbol{p}_m(t) \approx \boldsymbol{p}(t)$, then in the terminology of ODEs the 'residual' or 'defect' is defined as

$$\boldsymbol{r}_m(t) = \boldsymbol{p}_m'(t) - \boldsymbol{A}\boldsymbol{p}_m(t). \tag{7}$$

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

This definition reminds what happens when approximating the solution to a linear system $\boldsymbol{Ax} = \boldsymbol{b}$. Take GMRES, which uses the approximation $\boldsymbol{x}_m = \boldsymbol{x}_0 + \boldsymbol{V}_m\boldsymbol{y}_m$, where $\boldsymbol{x}_0$ is an initial guess, $\boldsymbol{y}_m = \overline{\boldsymbol{H}}_m^{\dagger}\beta\boldsymbol{e}_1$, with $\boldsymbol{V}_m$ and $\overline{\boldsymbol{H}}_m$ arising from the Arnoldi process for $\mathcal{K}_m(\boldsymbol{A}, \boldsymbol{r}_0)$, and denoting $r_0 = \boldsymbol{b} - \boldsymbol{Ax}_0$, $\beta = \|\boldsymbol{r}_0\|$, and $\overline{\boldsymbol{H}}_m^{\dagger}$ the pseudo-inverse of $\overline{\boldsymbol{H}}_m$. Applying (4), the inexact GMRES method ends up with (the norm of) the *computed* residual

$$\tilde{\boldsymbol{r}}_m = \boldsymbol{r}_0 - \boldsymbol{V}_{m+1}\overline{\boldsymbol{H}}_m\boldsymbol{y}_m, \tag{8}$$

while the *true* residual satisfies

$$\boldsymbol{r}_m = \boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}_0 + \boldsymbol{V}_m\boldsymbol{y}_m) = \tilde{\boldsymbol{r}}_m + \boldsymbol{\mathcal{E}}_m\boldsymbol{V}_m\boldsymbol{y}_m. \tag{9}$$

From (8) and (9), and the fact that $\boldsymbol{\mathcal{E}}_m\boldsymbol{V}_m = [\boldsymbol{E}_1\boldsymbol{v}_1, \ldots, \boldsymbol{E}_m\boldsymbol{v}_m]$ because $\boldsymbol{V}_m$ is orthonormal, there is an unknown *residual gap* that satisfies

$$\delta_m = \|\boldsymbol{r}_m - \tilde{\boldsymbol{r}}_m\| = \|\boldsymbol{\mathcal{E}}_m\boldsymbol{V}_m\boldsymbol{y}_m\| = \|[\boldsymbol{E}_1\boldsymbol{v}_1, \ldots, \boldsymbol{E}_m\boldsymbol{v}_m]\boldsymbol{y}_m\|. \tag{10}$$

The inexact solver still reports $\|\tilde{\boldsymbol{r}}_m\|$ as the estimate of the residual, but the residual gap (10) is not obvious, and so the reliability of the final solution is not guaranteed. Simoncini and Szyld [15] obtained the bound (see also [6,14]):

$$\delta_m \le \sum_{k=1}^{m} |\eta_k^{(m)}| \cdot \|\boldsymbol{E}_k\boldsymbol{v}_k\| \le \sum_{k=1}^{m} |\eta_k^{(m)}| \cdot \|\boldsymbol{E}_k\|, \qquad \boldsymbol{y}_m = \overline{\boldsymbol{H}}_m^{\dagger}\beta\boldsymbol{e}_1 = (\eta_1^{(m)}, \eta_2^{(m)}, \ldots, \eta_m^{(m)})^T.$$

Returning to the differential equation (5) where the residual of an approximation is defined by (7), and using (2) based on the exact Arnoldi process, we get the Krylov approximation $\boldsymbol{p}_m(\tau) = \boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1$ that leads to

$$\begin{aligned}\boldsymbol{p}_m'(\tau) - \boldsymbol{A}\boldsymbol{p}_m(\tau) &= \boldsymbol{p}_m'(\tau) - \boldsymbol{A}\boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1 \\ &= \boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\boldsymbol{H}_m\beta\boldsymbol{e}_1 - [\boldsymbol{V}_m\boldsymbol{H}_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T] \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1 \\ &= -\beta h_{m+1,m}\left(\boldsymbol{e}_m^T \exp(\tau\boldsymbol{H}_m)e_1\right)\boldsymbol{v}_{m+1}.\end{aligned}$$

Expokit [12] uses variants of this estimate (with some scaling) to monitor the accuracy. We define the *computed* residual as

$$\tilde{\boldsymbol{r}}_m(\tau) = -\beta h_{m+1,m}\left(\boldsymbol{e}_m^T \exp(\tau\boldsymbol{H}_m)e_1\right)\boldsymbol{v}_{m+1},$$

because this would still be the economical quantity (or a related one from it) used to estimate the *true* residual. As we shall see in the following section, with the inexact Arnoldi process (4), the *true* residual in the ODE problem becomes

$$\boldsymbol{r}_m(\tau) = \tilde{\boldsymbol{r}}_m(\tau) + \boldsymbol{\mathcal{E}}_m\boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1,$$

which is reminiscent of (9), but with $\boldsymbol{y}_m(\tau) = \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1$ instead of the least squares solution $\boldsymbol{y}_m = \overline{\boldsymbol{H}}_m^{\dagger}\beta\boldsymbol{e}_1$. Simoncini and Szyld [15] refined their bounds by exploiting properties satisfied by the components of $\boldsymbol{y}_m$ through Givens rotations in the case of GMRES. Our analysis will derive bounds without assuming those properties since Givens rotations are not involved in the case of the matrix exponential.

3.1. Homogeneous case

The Krylov technique for solving (5) is typically done by using the integration scheme

$$\begin{cases}\boldsymbol{p}(0) = \boldsymbol{p}_0 \\ \boldsymbol{p}(t_{k+1}) = \exp(\tau_k\boldsymbol{A})\boldsymbol{p}(t_k),\end{cases} \tag{11}$$

with some strategy for choosing the stepsizes $\tau_k = t_{k+1} - t_k$. The problem remains how to effectively approximate $\exp(\tau\boldsymbol{A})\boldsymbol{v}$ given τ and $\boldsymbol{v}$.

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

Now, using (2) based on the inexact Arnoldi process (4), the *true* residual of the resulting Krylov approximation $\boldsymbol{p}_m(\tau) = \boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1$ satisfies

$$\begin{aligned}\boldsymbol{r}_m &= \boldsymbol{p}_m' - \boldsymbol{A}\boldsymbol{p}_m \\ &= (\boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1)' - \boldsymbol{A}\boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1 \\ &= \boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\boldsymbol{H}_m \beta \boldsymbol{e}_1 - (\boldsymbol{V}_m \boldsymbol{H}_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T - \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m) \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1 \\ &= -h_{m+1,m}(\boldsymbol{e}_m^T \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1)\boldsymbol{v}_{m+1} + \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1 \\ &= \tilde{\boldsymbol{r}}_m + \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1,\end{aligned}$$

where $\tilde{\boldsymbol{r}}_m$ is the computed residual defined earlier. The dependency on the time τ will not be made explicit unless there is a risk of ambiguities. The quantity

$$\delta_m^{\text{res}} = \|\boldsymbol{r}_m - \tilde{\boldsymbol{r}}_m\| = \|\boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1\| \tag{12}$$

is the *residual gap* between $\boldsymbol{r}_m$ and $\tilde{\boldsymbol{r}}_m$. It depends on the relaxation matrices $\boldsymbol{E}_k$ and therefore cannot be computed in a straightforward manner. From the derivation above, we can bound the norm of the true residual as

$$\|\boldsymbol{r}_m\| \le \|\tilde{\boldsymbol{r}}_m\| + \delta_m^{\text{res}}. \tag{13}$$

When all matrix–vector products are exact, i.e., all $\boldsymbol{E}_k = \boldsymbol{0}$, then $\boldsymbol{\mathcal{E}}_m = \boldsymbol{0}$ and $\delta_m^{res} = 0$ as expected. More generally with $\boldsymbol{y}_m = \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1$, the residual gap can be written as

$$\delta_m^{\text{res}} = \|\boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \boldsymbol{y}_m\| = \|[\boldsymbol{E}_1 \boldsymbol{v}_1, \ldots, \boldsymbol{E}_m \boldsymbol{v}_m]\boldsymbol{y}_m\|,$$

and similarly to [15, Prop. 4.1], if we write $\boldsymbol{y}_m = (\eta_1^{(m)}, \ldots, \eta_m^{(m)})^T$, the following upper bound on the residual gap holds:

$$\delta_m^{\text{res}} \le \sum_{k=1}^{m} |\eta_k^{(m)}| \cdot \|\boldsymbol{E}_k\|, \tag{14}$$

and furthermore we have the following result:

Proposition 3.1. *Given* $\epsilon^{res} > 0$, *if* $\|\boldsymbol{E}_k\| \le \frac{\epsilon^{res}}{m\beta \|\exp(\tau \boldsymbol{H}_m)\|}$, $k = 1, \ldots, m$, *then we have*

$$\delta_m^{res} \le \epsilon^{res} \tag{15}$$

and therefore $\|\boldsymbol{r}_m\| \le \|\tilde{\boldsymbol{r}}_m\| + \epsilon^{res}$.

Proof. We have $|\eta_k^{(m)}| \le \|\boldsymbol{y}_m\| = \|\exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1\| \le \beta \|\exp(\tau \boldsymbol{H}_m)\|$, and so combining the condition on $\|\boldsymbol{E}_k\|$ with the inequality (14) gives (15). The bound on $\|\boldsymbol{r}_m\|$ then follows naturally from the relation $\|\boldsymbol{r}_m\| \le \|\tilde{\boldsymbol{r}}_m\| + \delta_m^{\text{res}}$. □

Remark 3.2. As Giraud et al. [6] pointed out in the context of inexact GMRES, Proposition 3.1 does not allow us to anticipate the relaxation matrices $\boldsymbol{E}_k$ in advance because of the dependence on $\boldsymbol{H}_m$. It can however be used in a postmortem manner to check if the error condition is satisfied.

3.2. Nonhomogeneous case

3.2.1. Bounding the residual gap via the φ function

It is well known (see, e.g., Expokit [12]) that the numerical solution to the system of nonhomogeneous ODEs

$$\begin{cases}\boldsymbol{p}'(t) = \boldsymbol{A}\boldsymbol{p}(t) + \boldsymbol{b} \\ \boldsymbol{p}(0) = \boldsymbol{p}_0\end{cases} \tag{16}$$

with constant $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{b} \in \mathbb{R}^n$, can be found using the integration scheme

$$\begin{cases}\boldsymbol{p}(0) = \boldsymbol{p}_0 \\ \boldsymbol{p}(t_{k+1}) = \tau_k \varphi(\tau_k \boldsymbol{A})[\boldsymbol{A}\boldsymbol{p}(t_k) + \boldsymbol{b}] + \boldsymbol{p}(t_k)\end{cases} \tag{17}$$

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

where $\varphi(\tau\boldsymbol{A}) = \sum_{i=0}^{\infty} \frac{(\tau\boldsymbol{A})^i}{(i+1)!}$ [8, Chap. 2.1, Chap. 10.7]. This integration scheme circumvents using the representation of the analytical solution of (16), $\boldsymbol{p}(t) = \exp(t\boldsymbol{A})\boldsymbol{p}_0 + t\varphi(t\boldsymbol{A})\boldsymbol{b}$, which would need both $\mathcal{K}_m(\boldsymbol{A}, \boldsymbol{p}_0)$ and $\mathcal{K}_m(\boldsymbol{A}, \boldsymbol{b})$ instead of only one Krylov subspace as implied in (17). Using this scheme, we can derive a result similar to Proposition 3.1, but this can be avoided by the augmented approach shown below.

3.2.2. Bounding the residual gap via an augmented matrix exponential

An indirect way to solve (16) takes root in the analytical solution $\boldsymbol{p}(t) = \exp(t\boldsymbol{A})\boldsymbol{p}_0 + t\varphi(t\boldsymbol{A})\boldsymbol{b}$, and has proved convenient in other circumstances such as [1,13]. Define the augmented matrix

$$\boldsymbol{A}^+ = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{0} & 0 \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)},$$

then we have

$$\exp(t\boldsymbol{A}^+) = \begin{pmatrix} \exp(t\boldsymbol{A}) & t\varphi(t\boldsymbol{A})\boldsymbol{b} \\ \boldsymbol{0} & 1 \end{pmatrix},$$

so that the solution can be fetched as $\boldsymbol{p}(t) = \left[\exp(t\boldsymbol{A}^+)\boldsymbol{p}_0^+\right]_{1:n}$ with $\boldsymbol{p}_0^+ = \begin{pmatrix} p_0 \\ 1 \end{pmatrix}$.

The problem now amounts to getting $\exp(t\boldsymbol{A}^+)\boldsymbol{p}_0^+$. Transforming the problem back to the form $\exp(t\boldsymbol{A}^+)\boldsymbol{p}_0^+$ not only inherits the analysis done in the homogeneous case in an elegant way, but also enables seamless code re-use. Furthermore, the inexactness in the matrix–vector product with $\boldsymbol{A}^+$ is only triggered from $\boldsymbol{A}$ through

$$\boldsymbol{A}^+ + \boldsymbol{E}_k^+ = \begin{pmatrix} \boldsymbol{A} + \boldsymbol{E}_k & \boldsymbol{b} \\ \boldsymbol{0} & 0 \end{pmatrix},$$

and so results can nicely be recast in terms of $\boldsymbol{E}_k$. For this reason, we will not dwell any further on the nonhomogeneous case in the rest of our presentation.

4. Bounds on the error

4.1. General upper bound on the error

As pointed out in the introduction, the inexact method for $\mathcal{K}_m(\boldsymbol{A}, \boldsymbol{v})$ with the perturbation matrices $\boldsymbol{E}_k, k = 1, \ldots, m$, can be seen as the exact method for $\mathcal{K}_m(\tilde{\boldsymbol{A}}, \boldsymbol{v})$ with

$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{\mathcal{E}}_m = \boldsymbol{A} + \sum_{k=1}^{m} \boldsymbol{E}_k \boldsymbol{v}_k \boldsymbol{v}_k^T.$$

Therefore if we define

$$\begin{aligned} \tilde{\epsilon}_m &= \| \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v} - \boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1 \| \\ &= \| \exp(\tau(\boldsymbol{A} + \boldsymbol{\mathcal{E}}_m))\boldsymbol{v} - \boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1 \|, \end{aligned}$$

then bounds on $\tilde{\epsilon}_m$ have been given in the literature of exact Krylov methods [5,11,9]. However, our main focus is not so much on the bounds on $\tilde{\epsilon}_m$, but on the *true* error

$$\epsilon_m = \| \exp(\tau\boldsymbol{A})\boldsymbol{v} - \boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1 \|.$$

The relationship between ϵ_m and $\tilde{\epsilon}_m$ is straightforward from the triangle inequality

$$\epsilon_m \le \| \exp(\tau\boldsymbol{A})\boldsymbol{v} - \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v} \| + \| \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v} - \boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1 \| = \tilde{\epsilon}_m + \delta_m^{err},$$

where we define the *error gap*

$$\delta_m^{\mathrm{err}} = \| \exp(\tau\boldsymbol{A})\boldsymbol{v} - \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v} \|. \tag{18}$$

With upper bounds on $\tilde{\epsilon}_m$ ready in hand, it remains to get good bounds on δ_m^{err}, which turns out to be a matrix perturbation analysis that we discuss next.

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

4.2. Bounding the error gap

In [8, Chap. 10.2], Higham obtained

$$\exp(\tau\tilde{\boldsymbol{A}}) = \exp(\tau\boldsymbol{A}) + \int_0^\tau \exp((\tau - s)\boldsymbol{A})\boldsymbol{\mathcal{E}}_m \exp(s\boldsymbol{A})ds + \mathcal{O}\left(\|\tau\boldsymbol{\mathcal{E}}_m\|^2\right). \tag{19}$$

Post-multiplying each side by $\boldsymbol{v}$, we get

$$\exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v} = \exp(\tau\boldsymbol{A})\boldsymbol{v} + \int_0^\tau \exp((\tau - s)\boldsymbol{A})\boldsymbol{\mathcal{E}}_m \exp(s\boldsymbol{A})\boldsymbol{v}ds + \mathcal{O}\left(\|\tau\boldsymbol{\mathcal{E}}_m\|^2\right).$$

Hence

$$\delta_m^{\text{err}} \le \int_0^\tau \|\exp((\tau - s)\boldsymbol{A})\|\|\boldsymbol{\mathcal{E}}_m\|\|\exp(s\boldsymbol{A})\|\|\boldsymbol{v}\|ds + \mathcal{O}(\|\tau\boldsymbol{\mathcal{E}}_m\|^2),$$

which leads directly to the following statement.

Theorem 4.1. *For any arbitrary $\boldsymbol{A}$, $\boldsymbol{v}$ and $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{\mathcal{E}}_m$ from the inexact Krylov subspace method, we have*

$$\delta_m^{err} = \|\exp(\tau\boldsymbol{A})\boldsymbol{v} - \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v}\| \le \beta h_A^2\|\tau\boldsymbol{\mathcal{E}}_m\| + \mathcal{O}(\|\tau\boldsymbol{\mathcal{E}}_m\|^2),$$

where $\beta = \|\boldsymbol{v}\|$ and $h_A = \max_{s\in[0,\tau]}\|\exp(s\boldsymbol{A})\|$ is the so-called 'hump' on $[0, \tau]$.

When $\boldsymbol{A}$ originates from a Markov chain as is the case in the CME, (i.e., with nonnegative off-diagonal elements, negative diagonal elements and zero column sums) it is known that $\|\exp(s\boldsymbol{A})\|_1 = 1, \forall s \ge 0$ (see for example [8, section 2.3]). We can then draw from (19) that

$$\|\exp(\tau\boldsymbol{A})\boldsymbol{v} - \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v}\|_1 \le \int_0^\tau \|\exp((\tau - s)\boldsymbol{A})\|_1\|\boldsymbol{\mathcal{E}}_m\|_1\|\exp(s\boldsymbol{A})\|_1\|\boldsymbol{v}\|_1 ds + \mathcal{O}(\|\tau\boldsymbol{\mathcal{E}}_m\|_1^2).$$

and using in addition the fact that a probability vector has $\|\boldsymbol{v}\|_1 = 1$, we get the following simplified result.

Theorem 4.2. *Let $\boldsymbol{A}$ be the transition rate matrix of a Markov chain, then given a probability vector $\boldsymbol{v}$ and the perturbed $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{\mathcal{E}}_m$ from the inexact Krylov subspace method, we have*

$$\|\exp(\tau\boldsymbol{A})\boldsymbol{v} - \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v}\|_1 \le \|\tau\boldsymbol{\mathcal{E}}_m\|_1 + \mathcal{O}(\|\tau\boldsymbol{\mathcal{E}}_m\|_1^2).$$

4.3. Series expansion of the error

In [11], Saad derived the following series expansion of the error produced by the exact Krylov subspace method

$$\exp(\tau\boldsymbol{A})\boldsymbol{v} - \boldsymbol{V}_m \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1 = \tau h_{m+1,m}\sum_{k=1}^{\infty}\boldsymbol{e}_m^T\varphi_k(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1(\tau\boldsymbol{A})^{k-1}\boldsymbol{v}_{m+1}, \tag{20}$$

of which the first term in the series was argued to be a good estimate of the error when the stepsize τ is small enough. Here, the functions φ_k are defined as

$$\varphi_k(x) = \sum_{i=0}^{\infty}\frac{x^i}{(i+k)!}, \quad \varphi_k(0) = \frac{1}{k!},$$

which implies that

$$0 \le \varphi_k(x) \le \frac{\varphi_{k-1}(x)}{k} \le \cdots \le \frac{e^x}{k!}, \quad \text{if } x \ge 0.$$

In the context of the inexact Krylov subspace method, we can use (4) and substitute $\boldsymbol{A}$ for $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{\mathcal{E}}_m$ in (20), but that will bring up the unwieldy issue of the error gap again. Instead, we generalize the expansion in the following statement that reveals how the terms involving $\boldsymbol{\mathcal{E}}_m$ break out in a strikingly neat way.

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

Theorem 4.3. *The (true) error in the inexact Krylov subspace method for the matrix exponential has the series expansion*

$$\exp(\tau \boldsymbol{A})\boldsymbol{v} - \boldsymbol{V}_m \exp(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1 = \sum_{k=1}^{\infty} (\tau \boldsymbol{A})^{k-1} \left[\tau h_{m+1,m} \left(\boldsymbol{e}_m^T \varphi_k(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1 \right) \boldsymbol{v}_{m+1} - \tau \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \varphi_k(\tau \boldsymbol{H}_m)\beta \boldsymbol{e}_1 \right]. \tag{21}$$

Proof. As in the proof of [11, Theorem 5.1], define the error in approximating $\varphi_k(\boldsymbol{A})\boldsymbol{v}$ by projecting it onto $\boldsymbol{V}_m$ as $\boldsymbol{s}_m^k = \varphi_k(\boldsymbol{A})\boldsymbol{v} - \boldsymbol{V}_m \varphi_k(\boldsymbol{H}_m)\beta \boldsymbol{e}_1$.

From the definition of φ_k and the fact that $\varphi_k(\boldsymbol{0})\boldsymbol{v} = \boldsymbol{V}_m \varphi_k(\boldsymbol{0})\beta \boldsymbol{e}_1$, we directly have

$$\begin{aligned}
\varphi_k(\boldsymbol{A})\boldsymbol{v} &= \boldsymbol{A}\varphi_{k+1}(\boldsymbol{A})\boldsymbol{v} + \varphi_k(\boldsymbol{0})\boldsymbol{v} \\
&= \boldsymbol{A}[\boldsymbol{V}_m \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 + \boldsymbol{s}_m^{k+1}] + \varphi_k(\boldsymbol{0})\boldsymbol{v} \\
&= \boldsymbol{V}_m[\varphi_k(\boldsymbol{0})\beta \boldsymbol{e}_1 + \boldsymbol{H}_m \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1] + h_{m+1,m} \boldsymbol{v}_{m+1} \boldsymbol{e}_m^T \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 \\
&\quad - \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 + \boldsymbol{A}\boldsymbol{s}_m^{k+1} \\
&= \boldsymbol{V}_m \varphi_k(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 + h_{m+1,m} \boldsymbol{v}_{m+1} \boldsymbol{e}_m^T \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 - \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 + \boldsymbol{A}\boldsymbol{s}_m^{k+1},
\end{aligned}$$

resulting in another expression for $\boldsymbol{s}_m^k$ through a recurrence

$$\boldsymbol{s}_m^k = h_{m+1,m} \left(\boldsymbol{e}_m^T \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 \right) \boldsymbol{v}_{m+1} - \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \varphi_{k+1}(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 + \boldsymbol{A}\boldsymbol{s}_m^{k+1}.$$

Using these expressions of the error terms gives

$$\begin{aligned}
\exp(\boldsymbol{A})\boldsymbol{v} - \boldsymbol{V}_m \exp(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 &= \boldsymbol{s}_m^0 \\
&= h_{m+1,m} \left(\boldsymbol{e}_m^T \varphi_1(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 \right) \boldsymbol{v}_{m+1} - \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \varphi_1(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 + \boldsymbol{A}\boldsymbol{s}_m^1 = \cdots \\
&= h_{m+1,m} \sum_{k=1}^{j} \left(\boldsymbol{e}_m^T \varphi_k(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 \right) \boldsymbol{A}^{k-1} \boldsymbol{v}_{m+1} - \sum_{k=1}^{j} \boldsymbol{A}^{k-1} \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \varphi_k(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 + \boldsymbol{A}^j \boldsymbol{s}_m^j,
\end{aligned}$$

in which

$$\|\boldsymbol{A}^j \boldsymbol{s}_m^j\| \le \|\boldsymbol{A}\|^j \|\boldsymbol{s}_m^j\| \le \|\boldsymbol{A}\|^j \beta \left(\varphi_j(\|\boldsymbol{A}\|) + \varphi_j(\|\boldsymbol{H}_m\|) \right) \le \beta \left(e^{\|\boldsymbol{A}\|} + e^{\|\boldsymbol{H}_m\|} \right) \frac{\|\boldsymbol{A}\|^j}{j!}$$

converges to 0 as $j \to \infty$. Taking this into account in the sums above, we get

$$\exp(\boldsymbol{A})\boldsymbol{v} - \boldsymbol{V}_m \exp(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 = \sum_{k=1}^{\infty} \boldsymbol{A}^{k-1} [h_{m+1,m} \left(\boldsymbol{e}_m^T \varphi_k(\boldsymbol{H}_m)\beta \boldsymbol{e}_1 \right) \boldsymbol{v}_{m+1} - \boldsymbol{\mathcal{E}}_m \boldsymbol{V}_m \varphi_k(\boldsymbol{H}_m)\beta \boldsymbol{e}_1].$$

Finally, if we rescale the inexact Arnoldi relation in (4) with the stepsize τ, we get (21). □

4.4. Exactness in the case of truncated approximations

The analysis so far has not made any assumption on the structure of $\boldsymbol{A}$, $\boldsymbol{v}$ or $\boldsymbol{E}_k$. In this section, we show a counter-intuitive result that, when $\boldsymbol{E}_k$ arises from a truncated approximation of a special form of the matrix $\boldsymbol{A}$, the inexact scheme can be exact.

Consider a banded matrix, and assume that $\boldsymbol{v} = \boldsymbol{e}_1 = (1, 0, \ldots, 0)^T$. The multiplication of $\boldsymbol{A}$ with $\boldsymbol{v}$ will therefore not involve the contribution of elements located in the trailing submatrices of $\boldsymbol{A}$. Generalizing this observation, we get the following result that arises frequently in the CME discussed in Section 2 and is therefore of wide interest.

Theorem 4.4. *Let $l \ge 0$, $k - l \ge 2$, assume that*

$$\boldsymbol{A} = \left(\begin{array}{c|c} \boldsymbol{A}_k & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array} \right) \in \mathbb{R}^{n \times n},$$

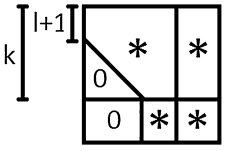
where

$$\boldsymbol{A}_k \in \mathbb{R}^{k\times k}, \quad (\boldsymbol{A}_k)_{ij} = 0 \text{ if } i > j + l,$$

and

$$\boldsymbol{C} \in \mathbb{R}^{(n-k)\times k}, \quad (\boldsymbol{C})_{ij} = 0 \text{ if } j \leq k - l - 1.$$

That is, $\boldsymbol{A}$ visually has the form:



Also assume that $\boldsymbol{v} = \boldsymbol{e}_1$ and the relaxation matrices are identical,

$$\boldsymbol{E} = \left(\begin{array}{c|c} \boldsymbol{0} & -\boldsymbol{B} \\ \hline -\boldsymbol{C} & -\boldsymbol{D} \end{array}\right) \in \mathbb{R}^{n\times n}.$$

Then if $m \leq M = \max\{j : (j-1)l + 1 \leq k - l - 1\}$, where m is the dimension of the basis built by the inexact Arnoldi algorithm based on matrix–vector products with $\boldsymbol{A} + \boldsymbol{E}$, we have

$$\boldsymbol{\mathcal{E}}_m = \boldsymbol{0},$$

so that

$$\exp(\tau\boldsymbol{A}) = \exp(\tau\tilde{\boldsymbol{A}}),$$

and

$$\delta_m^{err} = \boldsymbol{0}.$$

Proof. Observe first that, because of the special forms of $\boldsymbol{A}$ and $\boldsymbol{v}$, the $\boldsymbol{v}_j$ produced by the Arnoldi process is such that

$$(\boldsymbol{v}_j)_i = 0, \quad i > (j-1)l + 1, \ 1 \leq j \leq M.$$

This means that, up to $j = M$, the exact Arnoldi process (where multiplications are performed with $\boldsymbol{A}$) and the inexact Arnoldi process (where they are instead performed with $\boldsymbol{A} + \boldsymbol{E}$) coincide, due to the fact that the multiplications only depend on the first $k - l - 1$ columns of $\boldsymbol{A}$ and $\boldsymbol{A} + \boldsymbol{E}$, which are the same.

Because the first $k - l - 1$ columns of $\boldsymbol{E}$ are zero, and only the first $(j-1)l + 1 \leq k - l - 1$ elements of the vectors $\boldsymbol{v}_j$ are nonzero for $1 \leq j \leq m \leq M$, we have $\boldsymbol{E}\boldsymbol{v}_j = \boldsymbol{0}$, and therefore

$$\boldsymbol{\mathcal{E}}_m = \sum_{j=1}^{m} \boldsymbol{E}\boldsymbol{v}_j\boldsymbol{v}_j^T = \boldsymbol{0}.$$

Hence $\boldsymbol{A} = \tilde{\boldsymbol{A}}$ and naturally $\delta_m^{\text{err}} = \| \exp(\tau\boldsymbol{A})\boldsymbol{v} - \exp(\tau\tilde{\boldsymbol{A}})\boldsymbol{v}\| = 0$. □

Remark 4.5. The analysis in Theorem 4.4 will explain results apparently intriguing in the experiments, as the reader will soon discover in the following section. Note however that it only works for $\boldsymbol{v} = \boldsymbol{e}_1$. Since the step-by-step integration scheme (11) is typically used, $\boldsymbol{v}$ will not remain $\boldsymbol{e}_1$ past the first step, in which case the analysis of δ_m^{err} done in the previous section takes effect.

Remark 4.6. Even though this theorem shows that we have $\boldsymbol{\mathcal{E}}_m = 0$ and $\exp(\tau\boldsymbol{A}) = \exp(\tau\tilde{\boldsymbol{A}}), \forall\tau$, with a Krylov subspace of small dimension, this does not mean that the inexact Krylov approximation has no error. As shown in Theorem 4.3, even if $\boldsymbol{\mathcal{E}}_m = 0$, the expansion collapses to (20), which is generally not zero.

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

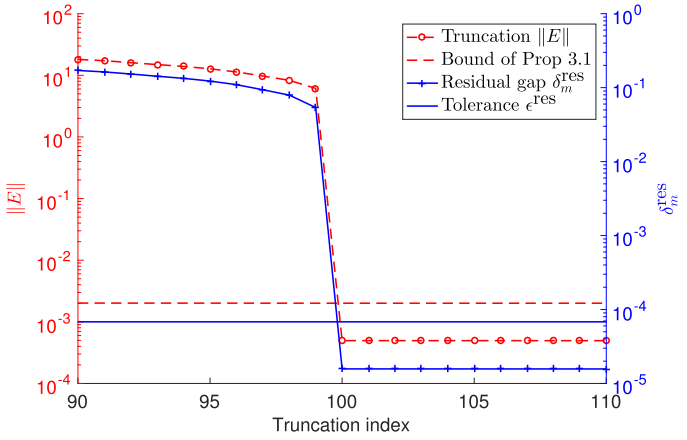


Fig. 1. Example 1: when the principal 100 × 100 submatrix becomes contained in the truncated matrix, the bound condition of Proposition 3.1 is satisfied, resulting in δ_m^{res} within the tolerance.

5. Numerical examples

We illustrate some of the theoretical results using three examples. The first two examples illustrate how Proposition 3.1 can be applied, by first showing how the relaxation scheme works according to the proposition, and secondly how the scheme fails when the condition of the proposition is not satisfied. The third example is a special one to demonstrate the peculiar behavior hinted in Remark 4.5: the scheme still works well even though the (sufficient) condition of Proposition 3.1 is not satisfied due to the reason given in Theorem 4.4. The examples were implemented in MATLAB.

We recall the following quantities that were given in the text to assess the inexact method:

$$\begin{aligned}
\text{true residual} &= \|\boldsymbol{r}_m\| = \|\boldsymbol{V}_m\boldsymbol{H}_m\boldsymbol{y}_m - \boldsymbol{A}\boldsymbol{V}_m\boldsymbol{y}_m\| \\
\text{computed residual} &= \|\tilde{\boldsymbol{r}}_m\| = |h_{m+1,m}\boldsymbol{e}_m^T\boldsymbol{y}_m| \\
\text{residual gap} &= \delta_m^{\text{res}} = \|\boldsymbol{r}_m - \tilde{\boldsymbol{r}}_m\| = \|\boldsymbol{\mathcal{E}}_m\boldsymbol{V}_m\boldsymbol{y}_m\| \\
\text{true error} &= \epsilon_m = \|\exp(\tau\boldsymbol{A})\boldsymbol{v} - \boldsymbol{V}_m\boldsymbol{y}_m\|,
\end{aligned}$$

where $\boldsymbol{y}_m = \exp(\tau\boldsymbol{H}_m)\beta\boldsymbol{e}_1$, with a given τ, and with $\boldsymbol{H}_m$ and $\boldsymbol{V}_m$ constructed by the inexact Krylov method using an initial given vector $\boldsymbol{v} = \boldsymbol{p}_0$ with $\beta = \|\boldsymbol{v}\|$.

5.1. Example 1 — illustration of the residual gap when Proposition 3.1 is satisfied

We take a 1000 × 1000 matrix, where the principal 100 × 100 submatrix is uniformly distributed in [0, 1], and entries outside this submatrix are uniformly distributed in $[0, 10^{-6}]$. The initial vector is $\boldsymbol{v} = (10^{-3}, \ldots, 10^{-3})^T$. We take the time point $\tau = 10^{-3}$, and tolerance on the residual gap $\epsilon^{\text{res}} = 10^{-3}$. The Krylov subspace is chosen to be of dimension $m = 15$.

For the inexact scheme, we define $\boldsymbol{A}_k$ to be a 1000 × 1000 matrix containing the principal $k \times k$ submatrix of $\boldsymbol{A}$ and 0 outside. Inexact multiplications with $\boldsymbol{A}$ are then performed with $\boldsymbol{A}_k$ instead. Observations are shown in Fig. 1. Since the key entries of $\boldsymbol{A}$ are in the principal 100 × 100 submatrix, it is clear that if $k < 100$, then $\boldsymbol{A}_k$ will leave out vital entries and inflate the error matrix $\boldsymbol{E} = \boldsymbol{A} - \boldsymbol{A}_k$, which in turn will cause the residual gap to be large. However, if $k \geq 100$, then $\|\boldsymbol{E}\| \leq \frac{\epsilon^{\text{res}}}{m\beta\|\exp(\tau\boldsymbol{H}_m)\|}$, which is the bound condition in Proposition 3.1, and therefore $\delta_m^{\text{res}} \leq \epsilon^{\text{res}}$ as guaranteed there. Fig. 1 illustrates this with the truncation index $90 \leq k \leq 110$.

5.2. Example 2 — illustration of the residual gap when Proposition 3.1 is not satisfied

We now consider another 1000 × 1000 matrix, where the main diagonal is uniformly distributed in [0, 1], and the off-diagonal entries are uniformly distributed in $[0, 10^{-6}]$. As in the first example, we keep $\boldsymbol{v} = (10^{-3}, \ldots,$

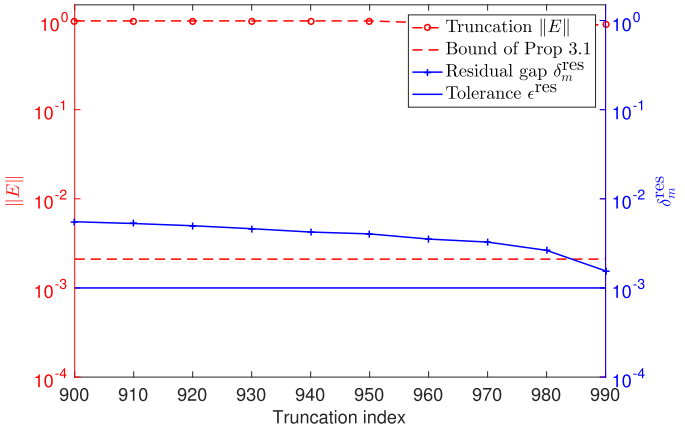


Fig. 2. Example 2: The bound condition of Proposition 3.1 is never satisfied, and δ_m^{res} is greater than the tolerance.

Table 1
Michaelis–Menten reactions and propensities.

	Reaction	Propensity	Rate constant (s^{-1})
1.	$\mathrm{E}+\mathrm{S} \xrightarrow{\kappa_1} \mathrm{ES}$	$\alpha_1 = \kappa_1\,[\mathrm{E}]\,[\mathrm{S}]$	$\kappa_1 = 1.0$
2.	$\mathrm{ES} \xrightarrow{\kappa_2} \mathrm{E}+\mathrm{S}$	$\alpha_2 = \kappa_2\,[\mathrm{ES}]$	$\kappa_2 = 1.0$
3.	$\mathrm{ES} \xrightarrow{\kappa_3} \mathrm{E}+\mathrm{P}$	$\alpha_3 = \kappa_3\,[\mathrm{ES}]$	$\kappa_3 = 0.1$

$10^{-3})^T$, $\tau = 10^{-3}$, $\epsilon^{\text{res}} = 10^{-3}$, and $m = 15$. The inexact setup is also the same, with inexact matrix–vector products against $\boldsymbol{A}$ performed using $\boldsymbol{A}_k$, where $\boldsymbol{A}_k$ contains the $k \times k$ principal submatrix of $\boldsymbol{A}$ and zeros elsewhere.

As Fig. 2 shows, in this example, the residual gap δ_m^{res} is consistently above the tolerance on the residual gap ϵ^{res} even when k is almost the size of $\boldsymbol{A}$. The reason for this is that since the diagonal elements are significant to be omitted, truncating even only one or two of them would make $\|\boldsymbol{E}\|$ large and not satisfy the bound condition of Proposition 3.1.

The conclusion to draw from these two examples is that when inexactness is achieved in a reasonable way, the inexact Krylov method works well. Depending on the particular problem at hand, one can decide how to relax the matrix–vector multiplications to satisfy Proposition 3.1, in which case the residual gap can be controlled by ϵ^{res}, ensuring that the computed residual $\|\tilde{\boldsymbol{r}}_m\|$ serves as a reliable approximation of the true residual $\|\boldsymbol{r}_m\|$.

5.3. *Example 3 — illustration of the error and residuals when Theorem 4.4 is satisfied*

We consider the CME arising from the Michaelis–Menten enzyme kinetics, which is a well known system of biochemical reactions in cell biology. There are four species: substrates (S), enzymes (E), enzyme–substrate complexes (ES) and products (P), interacting according to the three chemical reactions listed in Table 1, and we took reaction rates as in [7]. The state vector is $\boldsymbol{x} = ([\mathrm{P}], [\mathrm{E}], [\mathrm{S}], [\mathrm{ES}])^T$, where [X] is the current number of copies of species X. If we start with $\boldsymbol{x}_1 = (0, 50, 50, 0)^T$, i.e., a maximum of 50 substrates, the resulting matrix $\boldsymbol{A}$ of the underlying CME is of dimension $n = 1{,}326$. We use MATLAB's `expm` command to check for correctness. Fig. 3 shows the sparsity pattern of $\boldsymbol{A}$.

Since we know that the system starts in state $\boldsymbol{x}_1$, the initial probability vector is $\boldsymbol{p}_0 = \boldsymbol{e}_1$, and in the spirit of (6), we compute the approximation

$$\exp(\tau\boldsymbol{A})\boldsymbol{p}_0 \approx \exp(\tau\boldsymbol{A}_J)\boldsymbol{p}_0,$$

where $\boldsymbol{A}_J$ is a padded truncation of $\boldsymbol{A}$. We simply take a principal submatrix instead of the more general scheme where J can be an arbitrary subset of $\{1, \ldots, n\}$. There is no loss of generality because it can be assumed that there has been a reordering $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ where $\boldsymbol{P}$ is an appropriate permutation matrix. Our aim is to test the theory developed here and not really to craft efficient implementation details with elaborate sparse data structures that are best communicated

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

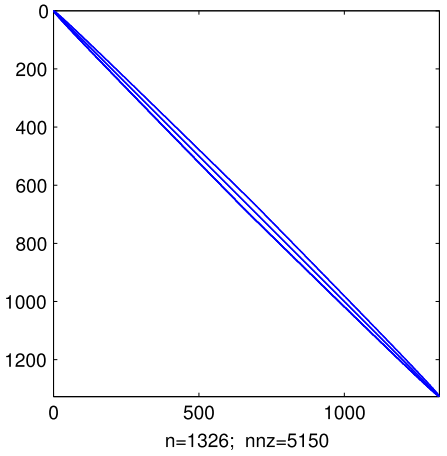


Fig. 3. Sparsity pattern of the matrix $\boldsymbol{A}$ from the CME of the Michaelis–Menten enzyme kinetics in Example 3.

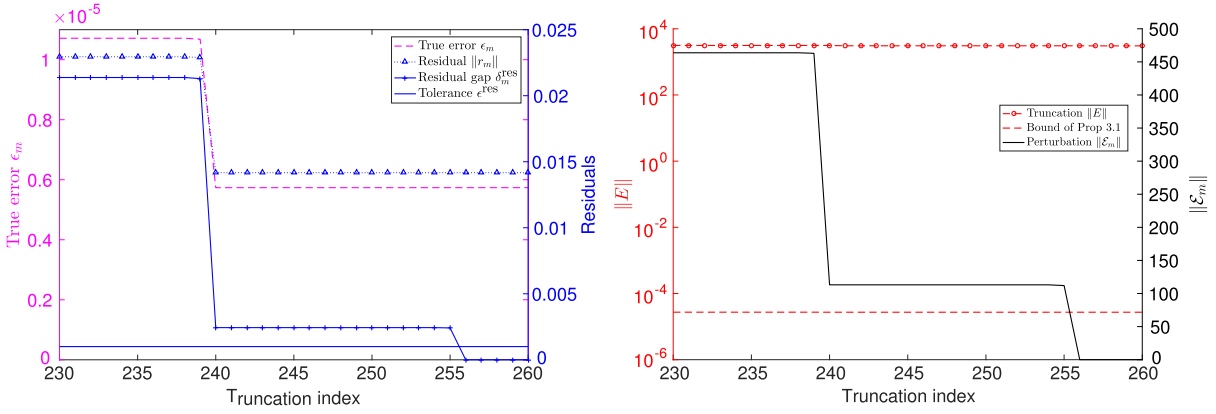


Fig. 4. Example 3. (Left) the true error (on the left y-axis), and the true residual and residual gap (on the right y-axis); (Right) $\|\boldsymbol{E}\|$ and $\|\boldsymbol{\mathcal{E}}_m\|$; computed as $|J|$ increases from 230 to 280; $\boldsymbol{v} = \boldsymbol{e}_1$, $\tau = 10^{-2}$, and $m = 30$.

elsewhere. We vary $|J|$, the cardinality of J that determines the size of the truncation from 230 to 280, so that $|J|$ ranges from about 17% to 21% of n in this example. We take $\boldsymbol{v} = \boldsymbol{e}_1$, $\tau = 10^{-2}$, and $m = 30$.

Fig. 4 shows the results. We see on the right plot a decrease of $\|\boldsymbol{\mathcal{E}}_m\|$ as the truncation size $|J|$ increases. This example has the particularity of illustrating the phenomenon described in Section 4.4, where $\boldsymbol{\mathcal{E}}_m$ becomes a zero matrix for a big enough truncation size $|J|$, even though $|J|$ is still only about 20% of n. Because of this phenomenon, the residual gap becomes 0, as the left plot shows. These results agree with the theory presented here. Note also on the left plot how the stair functions change in unison, illustrating the connection between the residuals and the error.

6. Conclusion

This work has analyzed inexact Krylov methods for approximating the action of the matrix exponential and provided insights into why they can be successful. We obtained results that in hindsight connect well with previous results, but it is worth recalling that it was unclear at the beginning exactly how such inexact methods related to previous works. The rigorous treatment presented here made the connection clear and established the details. This therefore fills a gap in the literature. We also brought into focus a particularly attractive aspect of inexact methods: they set a framework that encompasses model reduction methods in a generic way. We gave the important application of solving the chemical master equation (CME) as an example to motivate this viewpoint, with truncation methods such as the finite state projection (FSP) method that fit naturally in the framework. The study included numerical experiments to illustrate the theory.

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002

Acknowledgments

This work was supported in part by NSF grant DMS-1320849. We thank the anonymous reviewers for their comments that helped improve the paper.

References

[1] A.H. Al-Mohy, N.J. Higham, Computing the action of the matrix exponential, with an application to exponential integrators, SIAM J. Sci. Comput. 33 (2) (2011) 488–511.
[2] M.A. Botchev, V. Grimm, M. Hochbruck, Residual, restarting and Richardson iteration for the matrix exponential, SIAM J. Sci. Comput. 35 (3) (2013) A1376–A1397.
[3] A. Bouras, V. Fraysse, Inexact matrix–vector products in Krylov methods for solving linear systems: A relaxation strategy, SIAM J. Matrix Anal. Appl. 26 (3) (2005) 660–678.
[4] J.V.D. Eshof, G.L.G. Sleijpen, Inexact Krylov subspace methods for linear systems, SIAM J. Matrix Anal. Appl. 26 (1) (2004) 125–153.
[5] E. Gallopoulos, Y. Saad, Efficient solution of parabolic equations by Krylov approximation methods, SIAM J. Sci. Stat. Comput. 13 (5) (1992) 1236–1264.
[6] L. Giraud, S. Gratton, J. Langou, Convergence in backward error of relaxed GMRES, SIAM J. Sci. Comput. 29 (2) (2007) 710–728.
[7] J. Goutsias, Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems, J. Chem. Phys. 122 (18) (2005) 184102.
[8] N.J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia, PA, USA, 2008.
[9] M. Hochbruck, C. Lubich, On Krylov subspace approximations to the matrix exponential operator, SIAM J. Numer. Anal. 34 (1997) 1911–1925.
[10] B. Munsky, M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, J. Chem. Phys. 124 (4) (2006) 044104.
[11] Y. Saad, Analysis of some Krylov subspace appproximations to the matrix exponential operator, SIAM J. Numer. Anal. 29 (1) (1992) 209–228.
[12] R.B. Sidje, EXPOKIT: A software package for computing matrix exponentials, ACM Trans. Math. Software 24 (1) (1998) 130–156.
[13] R.B. Sidje, Inexact uniformization and GMRES methods for large Markov chains, Numer. Linear Algebra Appl. 18 (2011) 947–960.
[14] R.B. Sidje, N. Winkles, Evaluation of the performance of inexact GMRES, J. Comput. Appl. Math. 235 (2011) 1956–1975.
[15] V. Simoncini, D.B. Szyld, Theory of inexact Krylov subspace methods and applications to scientific computing, SIAM J. Sci. Comput. 25 (2) (2003) 454–477.

Please cite this article in press as: K.N. Dinh, R.B. Sidje, Analysis of inexact Krylov subspace methods for approximating the matrix exponential, Math. Comput. Simulation (2017), http://dx.doi.org/10.1016/j.matcom.2017.01.002