Available online at www.sciencedirect.com
ScienceDirect
Computers and Mathematics with Applications 53 (2007) 1293–1305
An International Journal
computers & mathematics with applications
www.elsevier.com/locate/camwa

An error analysis of the modified scaling and squaring method

S. Koikari

Department of Mechanical Engineering, Graduate School of Engineering, Kyoto University, Yoshida-Honmachi, Kyoto, Japan

Received 10 November 2005; received in revised form 27 March 2006; accepted 6 April 2006

Abstract

As a time discretization scheme for an ordinary differential equation with a stiff linear term, there is a class of methods that utilize the exponential or related functions of the coefficient matrix of the linear term. To implement these methods, we must compute a set of matrix functions called "φ-function", that includes the exponential itself, and it is important to compute these functions efficiently and accurately. In this paper, we consider the modified scaling and squaring method for the computation of φ-function. An algorithm based on Higham's method is defined, and the bounding parameter θ_m appropriate for φ-function is determined from an analysis of the truncation error under the assumption of the exact arithmetic. We also consider the propagation of the rounding error in the squaring process, and show that the error of φ-function is expected to be less than or roughly equal to that of the matrix exponential. Several evaluations are performed for famous test matrices, and the result shows that when the matrix exponential is computed accurately, the other φ-functions can also be obtained with the same level of accuracy.

Keywords: φ-function; Scaling and squaring methods; Truncation error analysis; Rounding error analysis; Exponential integrators

1. Introduction

Since the 1960's [1], exponential integrators have long been studied as time discretization methods especially suited for linearly stiff ODEs such as

$$y'(x) = \Lambda y(x) + f(y(x)), \tag{1.1}$$

where Λ is an $N \times N$ constant matrix. Various types of formulas including multistep type, Runge–Kutta type and Rosenbrock type have been proposed by many authors and have been investigated in terms of their theoretical foundation and practical efficiency (e.g., [2–6]). An extensive review work on this subject has recently been done by Minchev and Wright [7], and many relevant references are found therein.

For the most part of these various exponential integrators, a class of matrix functions, called "φ-function", are commonly used as the coefficients of time stepping formulas, and their definition is written as

$$\varphi_n(\Lambda h) := (\Lambda h)^{-n}\left(e^{\Lambda h} - \sum_{k=0}^{n-1}\frac{(\Lambda h)^k}{k!}\right) = \sum_{k=0}^{\infty}\frac{(\Lambda h)^k}{(k+n)!} = \frac{1}{n!}\,{}_1F_1(1, n+1; \Lambda h), \tag{1.2}$$

where ${}_1F_1$ is the confluent hypergeometric function [8].

E-mail address: koikari@mech.kyoto-u.ac.jp.

0898-1221/$ - see front matter
doi:10.1016/j.camwa.2006.04.032

It is well known that the computation of φ-function includes many difficult problems. In addition to the difficulties in the matrix exponentiation itself [9,10], the definition (1.2) cannot be used for a non-invertible matrix, while in many important applications the matrix does not have its inverse. Even when the inverse can be obtained, a serious cancellation occurs in the computation such as $e^x - 1$ for x nearly zero. To overcome these difficulties, various methods have been proposed, e.g., Krylov subspace approximation [11,12], the tridiagonal reduction [13], Cauchy integral method [14], and the modified (or corrected) scaling and squaring method based on Taylor approximation [3] or [6/6] Padé approximant [2,15]. Although these methods have their own advantages and disadvantages, we consider in this paper the modified scaling and squaring method. One of the advantages of this type of algorithm is that it can be applied to arbitrary square matrices without a priori knowledge of their structure. In addition, when a time integration is performed in a constant step width, an efficient exponential integrator is implemented with this method, because the matrix coefficients can be computed once and for all before starting the time stepping.

The purpose of this paper is to define an algorithm for computing φ-function as a natural extension of Higham's scaling and squaring method for the matrix exponential [16], and to provide an error analysis both for the truncation in the initial approximation and for the rounding in the squaring process. Based on the result of the truncation error analysis, a set of optimal parameters suited for φ-function is also determined. In Section 2, the definition of the algorithm studied in this paper is provided, and the truncation error analysis of the algorithm is performed in Section 3.1, where the effect of the truncation is considered under the assumption of the exact arithmetic. After introducing a class of auxiliary functions, a mixed forward–backward error result is obtained. By utilizing this result, we consider a set of optimal parameter in Section 3.2, and provide numerical evidence of their validity as much as possible with the aid of symbolic manipulation and multiple precision arithmetic. The propagation of the rounding error in the squaring process is also investigated in Section 3.3, and an expected tendency of the forward error of φ-function is confirmed through numerical experiments summarized in Section 4.

2. Description of the algorithm

Since the modified scaling and squaring method is an extension of the scaling and squaring algorithm for the matrix exponential, we briefly summarize the latter in the first step. The scaling and squaring method for $e^{\Lambda h}$ usually consists of the following three steps (e.g., [9,10]),

(1) Divide the matrix Λh by the sth power of two,
(2) Compute the diagonal Padé approximant $r_m(\Lambda h/2^s) \sim e^{\Lambda h/2^s}$,
(3) Square the matrix $r_m(\Lambda h/2^s)$ repeatedly to obtain $(r_m(\Lambda h/2^s))^{2^s} \sim e^{\Lambda h}$.

This algorithm is based only on the exponent law $e^{\Lambda h} = e^{\Lambda h/2}\, e^{\Lambda h/2}$, hence a similar algorithm is possible for a function other than the exponential, when the function has a "double-angle" relation. Although each φ-function does not have such a property except the exponential $e^{\Lambda h} =: \varphi_0(\Lambda h)$ itself, it is known that such a relation holds for a set of φ-functions $\{\varphi_n\}_{n=0}^{n_M}$ [2,3,15], which may be written in a general form as

$$\varphi_n(\Lambda h) = \frac{1}{2^n}\varphi_0(\Lambda h/2)\,\varphi_n(\Lambda h/2) + \frac{1}{2^n}\sum_{j=1}^{n}\frac{1}{(n-j)!}\,\varphi_j(\Lambda h/2), \tag{2.1}$$

or in the matrix-vector form as

$$\begin{bmatrix}\varphi_0(\Lambda h)\\ \varphi_1(\Lambda h)\\ \varphi_2(\Lambda h)\\ \varphi_3(\Lambda h)\end{bmatrix} = \begin{bmatrix}\varphi_0(\Lambda h/2) & 0 & 0 & 0\\ 0 & \frac{1}{2}(I+\varphi_0(\Lambda h/2)) & 0 & 0\\ 0 & \frac{1}{4}I & \frac{1}{4}(I+\varphi_0(\Lambda h/2)) & 0\\ 0 & \frac{1}{16}I & \frac{1}{8}I & \frac{1}{8}(I+\varphi_0(\Lambda h/2))\end{bmatrix}\begin{bmatrix}\varphi_0(\Lambda h/2)\\ \varphi_1(\Lambda h/2)\\ \varphi_2(\Lambda h/2)\\ \varphi_3(\Lambda h/2)\end{bmatrix}, \tag{2.2}$$