A numerical study of stiffness effects on some high order splitting methods

J. Salcedo-Ruíz
Instituto Politécnico Nacional, Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Culhuacán,
Av. Santa Ana No. 1000, Col. San Francisco Culhuacán, 04430 México, D.F. México.

F.J. Sánchez-Bernabe
Univ. Autónoma Metropolitana-Iztapalapa, Depto. de Matemáticas,
San Rafael Atlixco No. 186, Col. Vicentina, 09340 México, D.F. México.

Recibido el 18 de marzo de 2005; aceptado el 31 de enero de 2006

In this paper we compare operator splitting methods of first, second, third and fourth orders that are applied to problems with stiff matrices. In order to efficiently solve the resultant subproblems is necessary to use implicit Runge-Kutta methods. It is known that, in this context, the precision order of operator splitting schemes is reduced. Furthermore, we propose a fifth order operator splitting method that is obtained by applying Richardson extrapolation to a fourth order method. All methods are tested with a model problem with matrices such that its condition number is taken up to 20,000. Our conclusion is that order reduction is more severe for low order operator splitting methods.

Keywords: Operator splitting; stiff matrix; Richardson extrapolation; implicit Runge-Kutta methods.

En este trabajo se comparan métodos de descomposición de operadores de órdenes uno, dos, tres y cuatro, que se aplican a problemas cuyas matrices son de tipo rígido. A fin de poder resolver eficientemente los problemas intermedios que aparecen es necesario aplicar métodos de Runge-Kutta de tipo implícito. Se ha observado que en estas condiciones, el orden de precisión de los esquemas de descomposición de operadores se reduce. Se propone además un método de descomposición de operadores de orden cinco que se obtiene al aplicar extrapolación de Richardson a un esquema de orden cuatro. Todos los métodos se aplican a un problema modelo con matrices cuyo número de condición se incrementa hasta 20,000. Se concluye que el fenómeno de reducción de orden es más severa para los métodos de orden bajo.

Descriptores: Descomposición de operadores; matriz rígida; extrapolación de Richardson; métodos de Runge-Kutta implícitos.

PACS: 02.70.Bf; 02.90.+p; 02.60.-x

1. Introduction

High-order operator decomposition methods are important in the solution of many theoretical physics differential equations, in particular non-linear equations like *reaction-diffusion differential equations*. These equations describe the change in density of substances that spread in space and react chemically with other substances. This equations of this kind are often stiff [1].

We are interested in the analysis of the effect of the *order increase* of such methods on the relative errors of their results, especially in the case of *stiff differential equations*. In this paper we study the effect of both the order of the operator decomposition methods and the stiffness of the differential equations involved over the relative errors of results.

Verwer and Sportisse's work ([1], 1998), in which they achieved this analysis for first and second order operation decomposition methods, is a background reference for this paper. Previously, Goldman and Kaper ([2], 1996) studied third order methods.

In addition, we present and analyze in this paper a *fifth order method* which was obtained by applying Richardson's extrapolation (Stoer and Bulirsch, ([3], 2002) to a fourth order method developed by Sornborger and Stewart ([4], 1999).

Stiff differential equations are equations where implicit integration methods perform better, usually tremendously better, than explicit ones ([5], pp. 1 to 14). Taking that into account, we solved the initial value subproblems obtained by applying operator decomposition methods with Runge-Kutta implicit methods.

2. Operator decomposition methods

Given the following initial value problem

$$\frac{d\phi}{dt} = A\phi, \qquad t \in (0,T]\,, \tag{1}$$

with the initial condition

$$\phi(0) = \phi_0, \tag{2}$$

we are interested in solving (1) according to (2) in the case that A can be decomposed into a finite sum of simpler operators A_i, as follows:

$$A = A_1 + A_2 + \ldots + A_M. \tag{3}$$

Operator decomposition methods provide time discretization schemes. This is achieved by dividing the interval $(0,T]$, in which the problem is being solved, into s subintervals of equal length Δt (time discretization step), with moments t_i in which the values of the solution are needed (we will use the notations $t^n = n\Delta t$ and $\phi^n = \phi(n\Delta t), n \geq 1)$. Then, given the value of the solution at $t = 0$, for $n > 1$, ϕ^{n+1} is obtained from ϕ^n by solving a specific amount of simpler problems of the form

$$\frac{d\phi}{dt} = A_i\phi \tag{4}$$

For the rest of the paper, and in the numerical experiments, we shall assume that operators A_i are square matrices.

One of the *first order* operator decomposition methods with $M = 2$ approaches the solution of (1) on the interval $(t^n, t^{n+1})$by solving

$$\frac{dv}{dt} = A_1 v, \quad t \in \left(t^n, t^{n+1}\right], \tag{5}$$

with the initial condition ($\phi(t^n)$ is known):

$$v(t^n) = \phi(t^n), \tag{6}$$

Then it defines

$$\phi^{n+\frac{1}{2}} = v(t^{n+1}). \tag{7}$$

and solves

$$\frac{dv}{dt} = A_2 v, \quad t \in \left(t^n, t^{n+1}\right], \tag{8}$$

with the initial condition

$$v(t^n) = \phi^{n+\frac{1}{2}}. \tag{9}$$

Finally it defines

$$\phi(t^{n+1}) = v(t^{n+1}). \tag{10}$$

If matrices A_1 and A_2 are commutative, the following equality holds:

$$\phi(t^{n+1}) = e^{A_2\Delta t} e^{A_1\Delta t} \phi(t^n), \tag{11}$$

and the scheme (5)-(10) is exact. In the general case in which A_1 and A_2 are not commutative the scheme (5)-(10) is only of the first order with respect to time.

Likewise, assuming that A_1 and A_2 are commutative, the following equality holds:

$$\phi(t^{n+1}) = e^{A_2\Delta t/2} e^{A_1\Delta t} e^{A_2\Delta t/2} \phi(t^n). \tag{12}$$

This equality is the basis of Strang's scheme [7]. Further references can be found in Refs. 6 and 8. If matrices A_1 and A_2 are not commutative, Strang's scheme is of the *second order* with respect to time.

Using Sornborger and Stewart's notation [4],

$$(\Delta t) = e^{A_2\Delta t} e^{A_1\Delta t}, \tag{13}$$

and considering that

$$e^{A_2\frac{\Delta t}{2}} e^{A_1\Delta t} e^{A_2\frac{\Delta t}{2}} = e^{A_2\frac{\Delta t}{2}} e^{A_1\frac{\Delta t}{2}} e^{A_1\frac{\Delta t}{2}} e^{A_2\frac{\Delta t}{2}}, \tag{14}$$

Eq. (12) can be expressed as

$$\phi(t^{n+1}) = \left(\frac{\Delta t}{2}\right)\left(\frac{\Delta t}{2}\right)^T \phi(t^n). \tag{15}$$

Every method of the third or higher order with respect to time must include at least one backward time development problem [2] of the form

$$-\frac{dv}{dt} = A_j v \quad \text{en} \quad \left(t^n, t^{n+1}\right], \tag{16}$$

subject to a time condition:

$$v(t^n) = \phi(t^n), \tag{17}$$

Thus notation introduced in (13) becomes

$$(-\Delta t) = e^{-A_2\Delta t} e^{-A_1\Delta t}. \tag{18}$$

In the same paper, Sornborger and Stewart [4] present several *third* and *fourth order* methods. One of the third order schemes is named S_3 and defined by

$$\widetilde{S}_3(\Delta t) = (\Delta t)^T(\Delta t)(\Delta t)(\Delta t)(\Delta t)^T, \tag{19}$$

$$\widehat{S}_3(\Delta t) = (-2\Delta t)^T(\Delta t)(\Delta t)(\Delta t), \tag{20}$$

$$S_3(\Delta t) = \widetilde{S}_3(\Delta t)\widehat{S}_3(\Delta t). \tag{21}$$

One of the fourth order methods is

$$\widetilde{S}_4(\Delta t) = (\Delta t)^T(\Delta t)(\Delta t)^T(-2\Delta t)(\Delta t)^T(\Delta t)^T, \tag{22}$$

$$\widehat{S}_4(\Delta t) = (\Delta t)^T(\Delta t)^T(\Delta t)(\Delta t)^T(\Delta t)(\Delta t), \tag{23}$$

$$\overline{S}_4(\Delta t) = (\Delta t)(\Delta t)(-2\Delta t)^T(\Delta t)(\Delta t)^T(\Delta t), \tag{24}$$

$$S_4(\Delta t) = \widetilde{S}_4(\Delta t)\widehat{S}_4(\Delta t)\overline{S}_4(\Delta t). \tag{25}$$

Another way of obtaining high order time operator decomposition methods is to apply Richardson's extrapolation to low order methods. Thus the scheme

$$D_4(\Delta t) = \frac{4}{3}S_2(\Delta t/2)S_2(\Delta t/2) - \frac{1}{3}S_2(\Delta t), \tag{26}$$

where

$$S_2(\Delta t) = (\Delta t)(\Delta t)^T. \tag{27}$$

is obtained by applying Richardson's extrapolation to Strang's scheme (15). This scheme is studied in detail in Descombes [9].

An even higher order time operator decomposition method is obtained by applying Richardson's extrapolation to scheme (25). The result is the following *fifth order method*:

$$D_5(\Delta t) = \frac{16}{15}S_4(\Delta t/2)S_4(\Delta t/2) - \frac{1}{15}S_4(\Delta t). \tag{28}$$

Scheme (5)-(10) can be applied to decomposition in three operators by grouping two of them as follows:

$$A = A_1 + (A_2 + A_3). \tag{29}$$

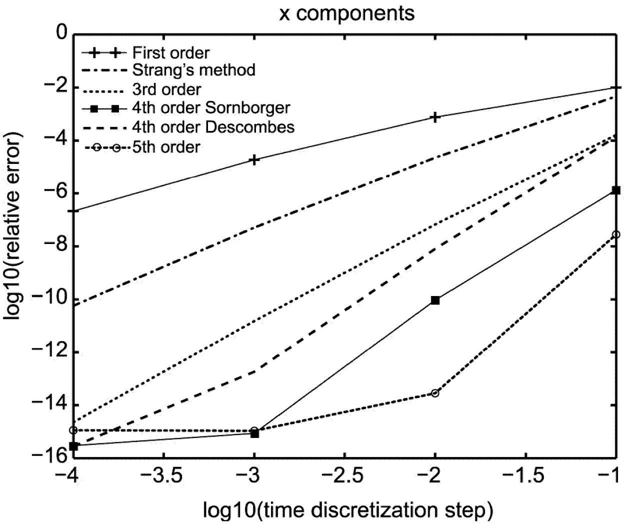

FIGURE 1. Relative errors of x-component of solutions to problem (32)-(37) for $n = 1$ and matriz A_1 applied before matrix A_2.

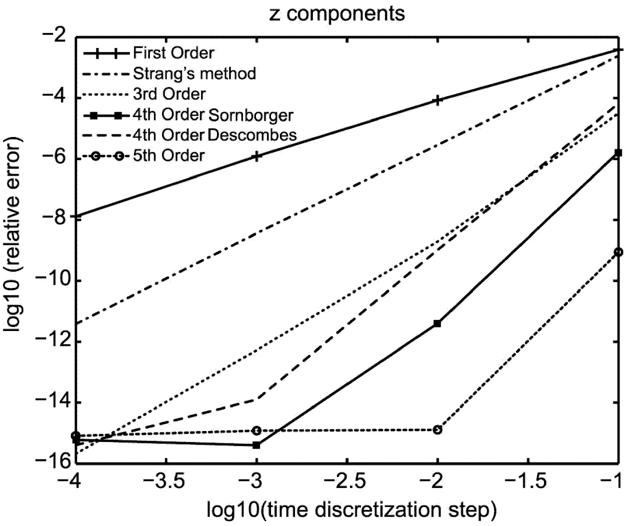

FIGURE 2. Relative errors of z-component of solutions to problem (32)-(37) for $n = 1$ and matriz A_1 applied before matrix A_2.

In some operator decomposition methods like the Θ-method [6], the operator A can only be decomposed into two simpler operators. The most important property of all the operator decomposition schemes analyzed in this paper is the possibility of applying them in cases when an operator must be decomposed into more than two operators. This is a great advantage when solving problems such as Bingham equations, in which it is convenient to decompose the problem into three operators, as shown in Sánchez [11].

3. Implicit Runge-Kutta methods

The application of high order operator decomposition methods requires the solving of problems of the form

$$\frac{dv}{dt} = A_i v, \quad t \in \left(t^n, t^{n+1}\right], \tag{30}$$

with the initial condition:

$$v(t^n) = \phi(t^n), \tag{31}$$

where $\phi(t^n)$ is the solution obtained on the interval $(t^{n-1}, t^n]$ by the applied method.

The derivatives from the differential Eqs. (30) must be discretized with a scheme at least as accurate as the corresponding operator decomposition methods, to avoid decreasing the accuracy of those methods. Besides, when the differential equations from problem (30)-(31) are stiff [3], the solutions of those problems obtained with an explicit method are not satisfactory unless an extremely short time step is used. Therefore, a high order implicit method should be used to solve such problems. In this paper, we chose the implicit Runge Kutta 5th order Radau IIA method [5,10].

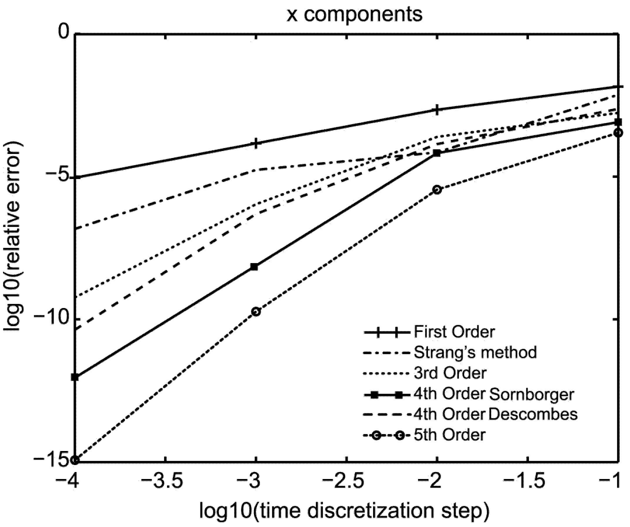

FIGURE 3. Relative errors of x-component of solutions to problem (32)-(37) for $n = 3$ and matriz A_1 applied before matrix A_2.

4. Stiffness and condition number

Stiff differential equations are equations where implicit methods perform better than explicit ones ([5], pp. 1 to 14).

The *condition number* of a matrix is equal to the product of its norm and the norm of its inverse. The condition number of a *symmetrical matrix* equals the quotient of its greatest eigenvalue over its lowest eigenvalue.

Given an stiff differential equation of the form

$$\frac{d\phi}{dt} = A\phi, \qquad t > 0,$$

if the operator A is discretized either by finite differences or by finite element method, the condition number of the resulting matrix A_d is generally a good measurement of the stiffness of that equation. A greater condition number of the matrix A_d usually implies a greater stiffness.

The following initial value problem:

$$\frac{dv}{dt} = Av, \tag{32}$$

with the initial condition

$$v(0) = v_0, \tag{33}$$

where

$$A = A_1 + A_2, \tag{34}$$

$$A_1 = \begin{pmatrix} -10^n & 10^n & 1 \\ 10^n & -10^n & 2 \\ 1 & 1 & -2 \end{pmatrix}, \tag{35}$$

$$A_2 = \begin{pmatrix} -1 & 0.5 & 0.25 \\ 0.1 & 0 & 0.1 \\ 0.2 & 0.4 & -1.0 \end{pmatrix}, \tag{36}$$

$$v_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \tag{37}$$

has the exact solution:

$$v(t) = v_0 \exp(At).$$

Increasing n increases the condition number of A_1, and hence the stiffness of the differential Eq. (32).

This problem has been solved with each of the operator decomposition methods studied, using in each case different values of n to determine how the stiffness of the differential equation affects the results. The approximate solutions obtained have been compared to the exact solution to determine their relative errors.

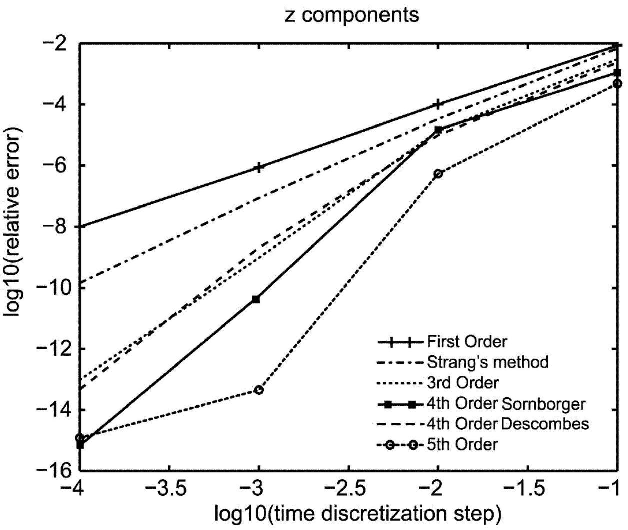

FIGURE 4. Relative errors of z-component of solutions to problem (32)-(37) for $n = 3$ and matriz A_1 applied before matrix A_2.

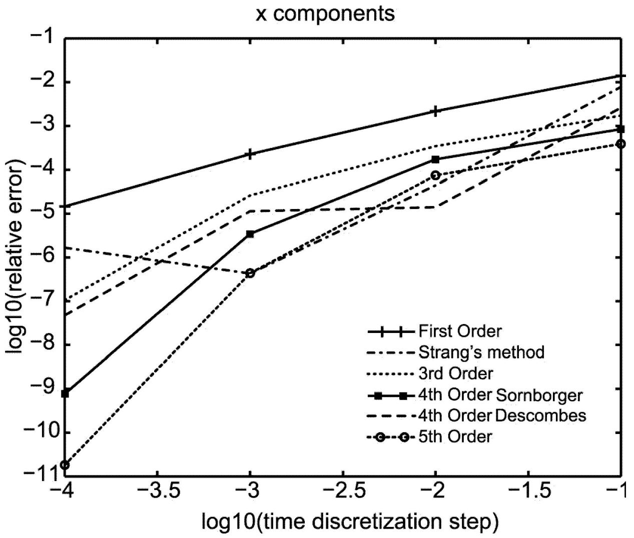

FIGURE 5. Relative errors of x-component of solutions to problem (32)-(37) for $n = 4$ and matriz A_1 applied before matrix A_2.

5. Numerical results

Results were obtained for the first order method (5)-(10) (represented graphically with the crossed line); Strang's method (15) (represented graphically with the line formed with dots and dashes); the third order method defined with (21) (represented graphically with the line formed with dots); the fourth order methods (25) (represented graphically with squares over a line) and (26) (represented graphically with the line formed with dashes); finally, results were obtained for the fifth order method (28) (represented graphically with circles over a line formed with dots). Relative errors were obtained for Δ t=0.1, 0.01, 0.001 and 0.0001. The condition number of the matrix A_2 is equal to 13.833 in all cases.

The relative errors of the x and z components of the solutions to the problem (32)-(37) with $n = 1$ (condition number of the matrix A_1 equal to 20.626) obtained with each analyzed method are graphed in Figs. 1 and 2, respectively. The relative errors of the y component behave similarly to those in Fig. 1. As expected, those figures show that the relative errors of the solutions decrease with the order of the method used. It also shows that method S_4 is more accurate than D_4.

The relative errors obtained with $n = 2$ (condition number of the matrix A_1 equal to 206.113) are similar to those obtained when using $n = 1$.

The relative errors of the x and z components of the solutions obtained with $n = 3$ (condition number of the matrix A_1 is equal to 2,061.113) are graphed in Figs. 3 and 4, respectively. The relative errors of the y component behave similar to those in Fig. 3. Those figures show that the relative errors are greater than those obtained when $n = 1$.

The relative errors of the x and z components of the solutions obtained with $n = 4$ (condition number of the matrix A_1 is equal to 20,611.132) are graphed in Figs. 5 and 6, respectively. The relative errors of the y component behave similarly to those in Fig. 5. In every case, the relative errors of the solutions are much greater than those obtained with $n = 3$, especially for the higher order methods. It is remarkable that, with $n = 4$ and time steps greater that Δ t=0.001, the Strang method behaves like the fifth order method. For lower time steps, the Strang method has the expected accuracy.

In all cases, we considered only the decomposition of operator A into two matrices. Thus it was possible to analyze the effect of inverting the order in which matrices A_1 and A_2 are applied. This is done in the first order scheme, for example, by substituting A_1 with A_2 in (5) and A_2 with A_1 in (8). There was no major difference in the accuracy obtained in this way for n equal to 1, 2 or 3. The relative errors of the x and z components of the solution obtained in this case with $n = 4$ are graphed in Figs. 7 and 8 respectively. Those figures show that the relative errors of the solutions obtained with the Strang S_2 and D_4 methods are greater than the corresponding relative error of the first order method for $\Delta t = 0.01$ or 0.001. For smaller time steps, S_2 and D_4 provide the expected results.

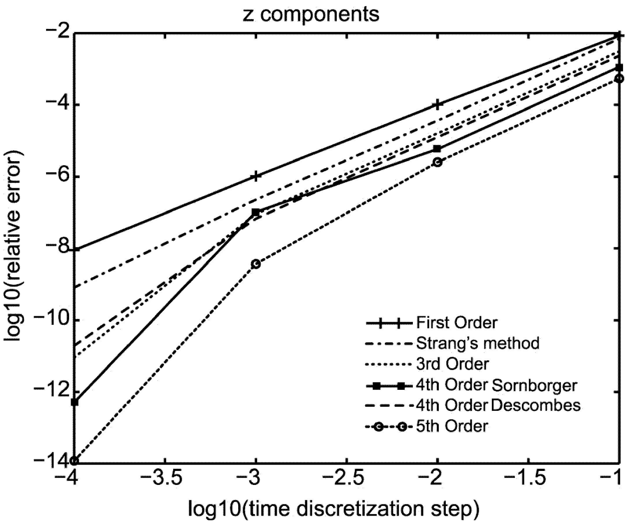

FIGURE 6. Relative errors of z-component of solutions to problem (32)-(37) for $n = 4$ and matriz A_1 applied before matrix A_2.

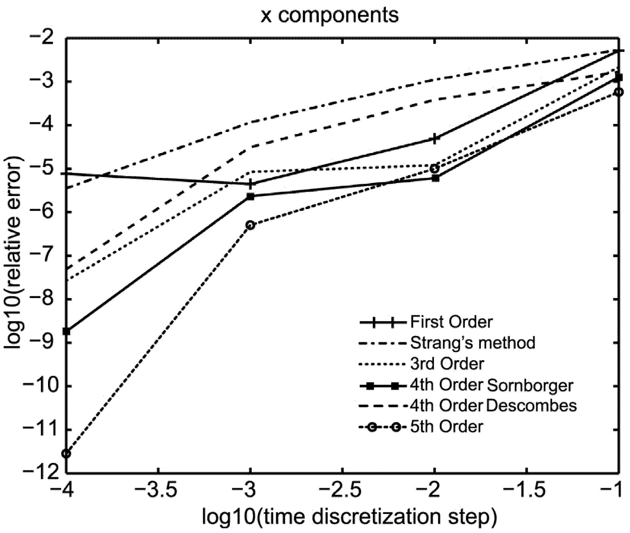

FIGURE 7. Relative errors of x-component of solutions to problem (32)-(37) for $n = 4$ and matriz A_2 applied before matrix A_1.

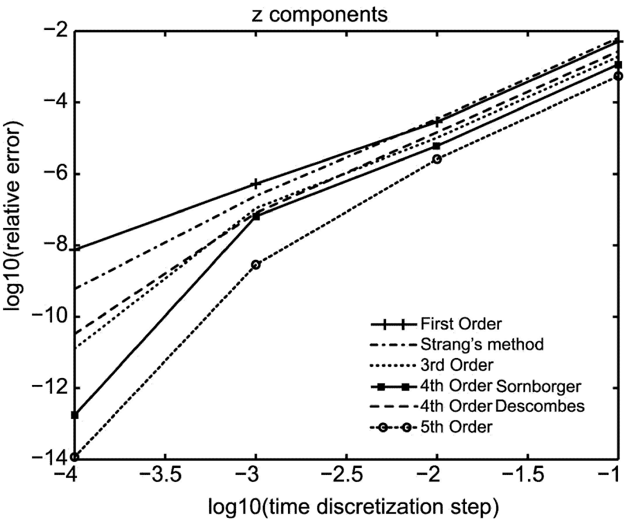

FIGURE 8. Relative errors of z-component of solutions to problem (32)-(37) for $n = 4$ and matriz A_2 applied before matrix A_1.

Two things are observed in all cases. The first one is that the relative errors of the y component and the relative errors of the x components behave similarly. And the second one is that the relative errors of the z component of the solutions are smaller than the relative errors of the x and y components.

6. Conclusions

1. Operator decomposition methods of the first, second, third, fourth, and fifth order have been studied in this paper. Results show that, as expected, the relative errors of the methods studied decrease with the order of the method used.

2. Results show that the relative errors of the methods studied increase with the stiffness of the differential equations involved especially for the higher order methods. But the increase in relative errors for high order methods is not as drastic as reported in Verwer y Sportisse [1], where only first and second order methods are discussed. As expected, the relative errors decrease monotonically with decreasing time steps.

3. In the case of stiff differential equations, Strang S_2 and D_4 methods behave suitable only when small time steps are used. Inverting the order in which matrices A_1 and A_2 are applied, the relative errors of these methods oscillate and are greater than relative errors obtained with lower order methods.

4. A fifth order operator decomposition method has been developed. This method produces good results. There was no major difference in the accuracy obtained by inverting the order in which matrices A_1 and A_2 are applied. In all cases considered, the relative errors obtained with this method were smaller than those obtained with the other methods analyzed, and does not oscillate with the change in the time step.

1. J.G. Verwer and B. Sportisse, *A Note on Operator Splitting in a Stiff Linear Case* (CWI Report: MAS-R9830, Amsterdam, 1998).
2. D. Goldman and T.J. Kaper, *SIAM J. Num. Analysis* **33** (1996) 349.
3. J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 3rd. ed. (Springer-Verlag, New York, 2002).
4. A. Sornborger and J. Stewart, *Physical Review A* **60** (1999) 1956.
5. E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems* (Springer-Verlag, Berlin, 1996).
6. R. Glowinski *Handbook of Numerical Analysis, Volume IX: Numerical Methods for Fluids (Part 3).* P.G. Ciarlet and J.L. Lions ed. (Elsevier Science, Amsterdam, 2003).
7. G. Strang, *SIAM J. Num. Analysis* **5** (1968) 506.
8. J. Salcedo y F. Sánchez, *Información Tecnológica, Revista Internacional*: **14** (2003) 69.
9. S. Descombes, *Mathematics of Computation* **70** (2000) 1481.
10. K. Dekker and J.G. Verwer, *Stability of Runge-Kutta methods for stiff nonlinear differential equations* (North-Holland, Amsterdam, 1984).
11. F.J. Sánchez, *Computers Math. Applic.* **36** (1998) 71.