

# On the linear stability of splitting methods

Sergio Blanes<sup>1\*</sup>    Fernando Casas<sup>2†</sup>    Ander Murua<sup>3‡</sup>

October 20, 2006

## Abstract

A comprehensive linear stability analysis of splitting methods is carried out by means of a  $2 \times 2$  matrix  $K(x)$  with polynomial entries (the stability matrix) and the stability polynomial  $p(x)$  (the trace of  $K(x)$ ). An algorithm is provided for determining the coefficients of a time-reversible splitting method directly from the stability polynomial. It is shown that  $p(x)$  carries, in fact, all the information needed to construct processed splitting methods for numerically approximating the evolution of linear systems. By selecting conveniently the stability polynomial, new integration methods with processing are built which are orders of magnitude more efficient than other algorithms previously available.

<sup>1</sup>Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, E-46022 Valencia, Spain.

<sup>2</sup>Departament de Matemàtiques, Universitat Jaume I, E-12071 Castellón, Spain.

<sup>3</sup>Konputazio Zientziak eta A.A. saila, Informatika Fakultatea, EHU/UPV, Donostia/San Sebastián, Spain.

*Key Words:* Splitting methods, Stability, Processing

## 1 Introduction

Splitting methods are frequently used in practice to integrate numerically differential equations. They constitute a natural choice when the vector field associated with the differential equation can be split into a sum of two or more parts that are each simpler to integrate than the original problem. Suppose we have an ordinary differential equation (ODE) of the form

$$z' = f(z) = f^{[A]}(z) + f^{[B]}(z) \quad (1.1)$$

such that the  $h$ -flows  $\varphi_h^{[A]}$  and  $\varphi_h^{[B]}$  corresponding to  $f^{[A]}$  and  $f^{[B]}$ , respectively, can be either exactly computed or accurately approximated. Then the exact flow  $\varphi_h$  of equation (1.1) can be approximated by a composition of the flows of the parts,

$$\psi_h = \varphi_{b_k h}^{[B]} \circ \varphi_{a_k h}^{[A]} \circ \cdots \circ \varphi_{b_2 h}^{[B]} \circ \varphi_{a_2 h}^{[A]} \circ \varphi_{b_1 h}^{[B]} \circ \varphi_{a_1 h}^{[A]}, \quad (1.2)$$

where the  $2k$  coefficients  $a_i, b_i$  are chosen as to ensure that  $\psi_h$  is a suitable approximation to the exact flow  $\varphi_h$ , typically in such a way that  $\psi_h = \varphi_h + \mathcal{O}(h^{p+1})$ : then the numerical integrator  $\psi_h$  is said to be accurate to order  $p$  in the time step  $h$ .

---

\*Email: [serblaza@imm.upv.es](mailto:serblaza@imm.upv.es)

†Email: [Fernando.Casas@uji.es](mailto:Fernando.Casas@uji.es)

‡Email: [Ander.Murua@ehu.es](mailto:Ander.Murua@ehu.es)

Perhaps the most frequently used splitting methods are

$$\psi_{h,1} = \varphi_h^{[B]} \circ \varphi_h^{[A]}, \quad \text{and} \quad \psi_{h,2} = \varphi_{h/2}^{[A]} \circ \varphi_h^{[B]} \circ \varphi_{h/2}^{[A]}, \quad (1.3)$$

corresponding to the first order Lie–Trotter method and the second order leapfrog (also called Störmer, Verlet, Strang splitting, etc.) method, respectively. Their straightforward implementation and low storage requirements have made them common tools for the numerical treatment of ODEs and partial differential equations (PDEs). In addition, they are very fast if the system is split in easy computable parts.

Splitting schemes have proved to be especially useful in the context of *geometric integration*, since they preserve structural features of the flow of  $f$  as long as the basic methods  $\varphi_h^{[A]}$  and  $\varphi_h^{[B]}$  also share these properties. Important examples include symplecticity, volume preservation, symmetries, etc. In this sense, schemes (1.3) can be considered as geometric integrators, and as such, they show smaller error growth than standard integrators. It is not surprising, then, that a systematic search for splitting methods of higher order of accuracy has taken place during the last two decades and a large number of them exist in the literature (see [11, 16, 23, 24, 27] and references therein) which have been specifically designed for different families of problems.

Another characteristic of a numerical integration method for differential equations is *stability*. Roughly speaking, the numerical solution provided by a stable numerical integrator does not tend to infinity when the exact solution is bounded. Although important, this feature has received considerably less attention in the specific case of splitting methods. In fact, in the process of building high order schemes, stability is not usually taken into account, ending with methods which are rather less stable than  $\psi_{h,1}$  and  $\psi_{h,2}$ , which, as is well known, attain the highest possible stability (at a given cost) among all methods given by (1.2) [8].

Usually, the  $k$ -stage scheme (1.2) is unstable for  $h > kL$  for some  $L > 0$ , so that stability imposes restrictions on the step size. In this sense, the length  $L$  of the stability interval should be as large as possible for a given order and a high order method possessing a tiny  $L$  could be useless in practice.

Typically, the stability analysis of (1.2) is carried out when the scheme is applied to the model problem [18]

$$y'' + \lambda^2 y = 0, \quad \lambda > 0, \quad (1.4)$$

with the standard  $((q, p) = (\lambda y, y'))$  splitting,

$$\begin{Bmatrix} q' \\ p' \end{Bmatrix} = \left[ \underbrace{\begin{pmatrix} 0 & \lambda \\ 0 & 0 \end{pmatrix}}_A + \underbrace{\begin{pmatrix} 0 & 0 \\ -\lambda & 0 \end{pmatrix}}_B \right] \begin{Bmatrix} q \\ p \end{Bmatrix}, \quad (1.5)$$

so that  $z = (q, p)^T$  and  $f^{[A]}(z) = Az$ ,  $f^{[B]}(z) = Bz$ .

As an illustration, the application of schemes (1.3) to (1.5) provides

$$\psi_{h,1} = \begin{pmatrix} 1 & x \\ -x & 1 - x^2 \end{pmatrix} \quad \text{and} \quad \psi_{h,2} = \begin{pmatrix} 1 - \frac{x^2}{2} & x - \frac{x^3}{4} \\ -x & 1 - \frac{x^2}{2} \end{pmatrix} \quad (1.6)$$

respectively, where  $x = h\lambda$ . Now, as both matrices have unit determinant, one concludes that  $\psi_{h,1}$  and  $\psi_{h,2}$  are stable if  $\text{tr}(\psi_{h,1}) = \text{tr}(\psi_{h,2}) = |2 - x^2| < 2$ , or  $|x| < 2$  and thus  $L = 2$  is the greatest value one can obtain for any splitting method, since  $k = 1$ . With respect to  $\psi_{h,2}$ , although  $k = 2$ , it requires in practice only one evaluation per step, as one  $\varphi_{h/2}^{[A]}$  can be concatenated with the next step.

Alternatively, notice that  $\psi_{h,2} = \varphi_{h/2}^{[A]} \circ \psi_{h,1} \circ (\varphi_{h/2}^{[A]})^{-1}$  and thus both possess the same value  $L = 2$ .

One of the aims of this paper is precisely to analyze the composition (1.2) to get high order schemes with relatively large stability intervals (as close as possible to those of (1.3)) with the ultimate goal of improving the performance of the existing splitting methods.

The analysis carried out here also allows us, in a natural way, to construct new methods of any order and with large stability intervals for the harmonic oscillator. They have the special structure

$$\hat{\psi}_h = \pi_h \circ \psi_h \circ \pi_h^{-1} \quad (1.7)$$

and, as thus, could be considered as generalizations of  $\psi_{h,2}$ . In (1.7)  $\psi_h$  is referred to as the kernel and  $\pi_h^{-1}$ ,  $\pi_h$  are the pre- and postprocessors or correctors. Application of  $\hat{\psi}_h$  over  $n$  integration steps with constant step size  $h$  gives

$$\hat{\psi}_h^n = \pi_h \circ \psi_h^n \circ \pi_h^{-1}.$$

The cost of  $\hat{\psi}_h$  is measured by the cost of the kernel because  $\pi_h^{-1}$  is computed only once to start the integration and  $\pi_h$ , in most cases, is only required occasionally. One says that  $\psi_h$  has effective order  $p$  if there exist maps  $\pi_h^{-1}$ ,  $\pi_h$  such that  $\hat{\psi}_h$  has order  $p$ . This processing technique was first introduced by Butcher [7] and later received a renewed attention in the context of geometric integration [26, 29, 21, 18, 19, 4, 5, 6, 1, 2]. The stability analysis of such class of methods is mainly concerned with the kernel, since the processor does not affect the stability (it is just a change of variables). Here we consider a composition (1.2) for the kernel.

Obviously, there is not much point in designing new and somewhat sophisticated numerical methods for the harmonic oscillator (1.5). It turns out, however, that splitting methods especially tailored for this system can be of great interest for the numerical treatment of non-trivial problems appearing, for instance, in quantum mechanics, electrodynamics and structural dynamics [9, 14, 15, 30]. We should also remark that many PDEs, once spatially discretized, give rise to large systems of coupled harmonic oscillators where the techniques and methods proposed in this work directly apply.

The paper is organized as follows. In section 2 we introduce and characterize the stability matrix of a splitting method as applied to the one-dimensional harmonic oscillator. In particular, we show that any splitting method is uniquely determined by its stability matrix. In section 3 the analysis is generalized to certain classes of linear systems relevant in applications. In section 4, by proposing different families of stability polynomials and stability matrices, we get particular splitting methods of the form (1.7) with high accuracy and enlarged stability domain for linear Hamiltonian systems. Finally, the new methods are illustrated in section 5 on some numerical examples.

## 2 Analysis of splitting methods applied to the harmonic oscillator

When applying the splitting method (1.2) to the one-dimensional harmonic oscillator (1.5) we approximate the exact  $2 \times 2$  solution matrix

$$O(x) = \begin{pmatrix} \cos(x) & \sin(x) \\ -\sin(x) & \cos(x) \end{pmatrix}, \quad x = h\lambda \quad (2.1)$$

by  $K(x)$ , where

$$K(x) = \begin{pmatrix} 1 & 0 \\ -b_k x & 1 \end{pmatrix} \begin{pmatrix} 1 & a_k x \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ -b_1 x & 1 \end{pmatrix} \begin{pmatrix} 1 & a_1 x \\ 0 & 1 \end{pmatrix}, \quad (2.2)$$

or alternatively, in terms of matrices  $A$  and  $B$  introduced in (1.5),

$$K(x) = e^{hb_k B} e^{ha_k A} \cdots e^{hb_1 B} e^{ha_1 A}, \quad (2.3)$$

with appropriate coefficients  $a_i, b_i \in \mathbb{R}$ . The matrix  $K(x)$  (the *stability matrix* of the splitting method) has the form

$$K(x) = \begin{pmatrix} K_1(x) & K_2(x) \\ K_3(x) & K_4(x) \end{pmatrix}, \quad (2.4)$$

with elements

$$\begin{aligned} K_1(x) &= 1 + \sum_{i=1}^{k-1} k_{1,i} x^{2i}, & K_2(x) &= \sum_{i=1}^{k-1} k_{2,i} x^{2i-1} \\ K_3(x) &= \sum_{i=1}^{k-1} k_{3,i} x^{2i-1}, & K_4(x) &= 1 + \sum_{i=1}^k k_{4,i} x^{2i}. \end{aligned} \quad (2.5)$$

Here  $k_{i,j}$  are homogeneous polynomials in the parameters  $a_i, b_i$ . In particular,  $k_{2,1} = \sum_{j=1}^k a_j$ ,  $k_{3,1} = -\sum_{j=1}^k b_j$  and  $k_{1,1} + k_{4,1} = k_{2,1} k_{3,1}$ . Since each individual matrix in the composition (2.2) has unit determinant, it must hold that  $\det K(x) \equiv 1$ . It is not difficult to check that

$$|d(K_1) - d(K_4)| \leq 2, \quad |d(K_2) - d(K_3)| \leq 2, \quad (2.6)$$

where we denote by  $d(K_i)$  the degree of each polynomial  $K_i(x)$  ( $i = 1, \dots, 4$ ).

As a matter of fact, the linear stability analysis of splitting methods is made easier if one considers a generic  $2 \times 2$  matrix (2.4) such that  $K_1(x)$  and  $K_4(x)$  (respectively,  $K_2(x)$  and  $K_3(x)$ ) are even (respectively, odd) polynomials in  $x$  and

$$\det K(x) = K_1(x)K_4(x) - K_2(x)K_3(x) \equiv 1, \quad (2.7)$$

$$K_1(0) = K_4(0) = 1. \quad (2.8)$$

In addition, since one is interested in consistent methods, we also require that

$$K_2'(0) = -K_3'(0) = 1. \quad (2.9)$$

In the application of an splitting method to the harmonic oscillator, an essential role is played by the so-called *stability polynomial*, defined as

$$p(x) = \frac{1}{2} \operatorname{tr} K(x) = \frac{1}{2} (K_1(x) + K_4(x)).$$

Clearly, the stability polynomial of a consistent splitting method is an even polynomial  $p(x)$  satisfying

$$p(x) = 1 - x^2/2 + \mathcal{O}(x^4) \quad \text{as } x \rightarrow 0. \quad (2.10)$$

In particular, for schemes (1.3) one has, from (1.6),  $p(x) = 1 - x^2/2$ .

## 2.1 From the stability matrix to the splitting method

Obviously, analyzing splitting methods through a generic stability matrix is useful as long as one is able to factorize  $K(x)$  as (2.2) and determine uniquely the coefficients  $a_i, b_i$  of the splitting method from a particular  $K(x)$  with polynomial entries. Only in such circumstances one could say that any splitting method of the form (1.2) is completely characterized by the result of applying one step of the method to the harmonic oscillator. What the following result shows is that, under certain assumptions, each matrix (2.4) can be decomposed uniquely in certain form, that includes (2.2) as a particular case.

**Proposition 2.1** *Given a matrix (2.4) such that  $K_1(x), K_4(x)$  are even polynomials and  $K_2(x), K_3(x)$  are odd polynomials verifying conditions (2.7)-(2.9), there exists a unique decomposition of  $K(x)$  of the form*

$$\begin{pmatrix} 1 & 0 \\ -B_m(x) & 1 \end{pmatrix} \begin{pmatrix} 1 & A_m(x) \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ -B_1(x) & 1 \end{pmatrix} \begin{pmatrix} 1 & A_1(x) \\ 0 & 1 \end{pmatrix}, \quad (2.11)$$

where  $A_j(x), B_j(x)$  ( $j = 1, \dots, m$ ) are odd polynomials in  $x$ .

**Proof:** We will prove the result by induction on the sum of the degrees of the two polynomials  $K_1(x)$  and  $K_4(x)$ . In the trivial case, where the sum of their degrees is 0, it holds by assumption that  $K_1(x) \equiv K_4(x) \equiv 1$ , and since  $\det K(x) \equiv 1$ , either  $K_2(x) \equiv 0$  or  $K_3(x) \equiv 0$ .

If  $d(K_1) + d(K_4) > 0$ , then two possibilities occur:

- $d(K_1) < d(K_2)$ . In that case  $d(K_3) < d(K_4)$ , since  $\det K(x) \equiv 1$ . Application of polynomial division uniquely determines the odd polynomials  $A_1(x)$  and  $\hat{K}_2(x)$  such that

$$K_2(x) = K_1(x)A_1(x) + \hat{K}_2(x), \quad (2.12)$$

where  $d(K_2) = d(K_1) + d(A_1)$  and  $d(\hat{K}_2) < d(K_1)$ . Now, we define the even polynomial  $\hat{K}_4(x) = K_4(x) - K_3(x)A_1(x)$ , so that

$$\begin{pmatrix} K_1(x) & K_2(x) \\ K_3(x) & K_4(x) \end{pmatrix} = \begin{pmatrix} K_1(x) & \hat{K}_2(x) \\ K_3(x) & \hat{K}_4(x) \end{pmatrix} \begin{pmatrix} 1 & A_1(x) \\ 0 & 1 \end{pmatrix}.$$

Clearly,  $\hat{K}_4(0) = 1$  and  $K_1(x)\hat{K}_4(x) - \hat{K}_2(x)K_3(x) = 1$ , which together with  $d(\hat{K}_2) < d(K_1)$  implies that  $d(\hat{K}_4) < d(K_3) < d(K_4)$ , and the required result follows by induction.

- $d(K_1) > d(K_2)$ . Then  $d(K_3) > d(K_4)$  because  $\det K(x) \equiv 1$ , and one similarly obtains the decomposition

$$\begin{pmatrix} K_1(x) & K_2(x) \\ K_3(x) & K_4(x) \end{pmatrix} = \begin{pmatrix} \hat{K}_1(x) & K_2(x) \\ \hat{K}_3(x) & K_4(x) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ B_1(x) & 1 \end{pmatrix},$$

where the odd polynomials  $B_1(x)$  and  $\hat{K}_3(x)$  are determined from the polynomial division of  $K_3(x)$  by  $K_4(x)$ , so that  $K_3(x) = K_4(x)B_1(x) + \hat{K}_3(x)$ , and then  $\hat{K}_1(x)$  is determined as  $\hat{K}_1(x) = K_1(x) - K_2(x)B_1(x)$ . Clearly,  $\hat{K}_1(0) = 1$  and  $\hat{K}_1(x)K_4(x) - K_2(x)\hat{K}_3(x) = 1$ . Since now  $d(\hat{K}_3) < d(K_4)$ , then  $d(\hat{K}_1) < d(K_2) < d(K_1)$  and the required result follows by induction.

This completes the proof.  $\square$

**Remarks:**

1. Notice from the proof of Proposition 2.1 that  $\hat{K}_4(x)$  (respectively  $\hat{K}_1(x)$ ) can also be obtained as the remainder of the polynomial division of  $K_4(x)$  by  $K_3(x)$  (respectively  $K_1(x)$  by  $K_2(x)$ ), which (in exact arithmetic) must give the same quotient  $A_1(x)$  (resp.  $B_1(x)$ ) due to the fact that  $\det K(x) \equiv 1$ .
2. Obviously, the decomposition (2.2) corresponds to (2.11) with  $A_j(x) = a_j x$  and  $B_j(x) = b_j x$  for  $j = 1, \dots, m$ , and  $m = k$ .

**Example:** Let us illustrate this result with two different stability matrices, leading to different types of decomposition. Consider first the matrix

$$K(x) = \begin{pmatrix} 1 - \frac{1}{2}x^2 + \frac{1}{32}x^4 & x - \frac{3}{16}x^3 + \frac{1}{128}x^5 \\ -x + \frac{1}{8}x^3 & 1 - \frac{1}{2}x^2 + \frac{1}{32}x^4 \end{pmatrix}. \quad (2.13)$$

which satisfies conditions (2.6)-(2.9). By applying the constructive proof of Proposition 2.1 it is straightforward to check that  $K(x)$  can be decomposed as

$$K(x) = \begin{pmatrix} 1 & \frac{x}{4} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{x}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{x}{2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{x}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{x}{4} \\ 0 & 1 \end{pmatrix}, \quad (2.14)$$

and this is in fact the only decomposition of the form (2.2) for the matrix (2.13). As a second example, consider now

$$\bar{K}(x) = \begin{pmatrix} 1 - \frac{1}{2}x^2 + \frac{1}{32}x^4 & x - \frac{1}{4}x^3 + \frac{1}{64}x^5 \\ -x + \frac{1}{16}x^3 & 1 - \frac{1}{2}x^2 + \frac{1}{32}x^4 \end{pmatrix} \quad (2.15)$$

with the same stability polynomial as  $K(x)$ . In this case, although the degrees of the entries coincide with those of (2.13) (and thus condition (2.6) holds),  $\bar{K}(x)$  can not be decomposed in the form (2.2). Instead it admits the unique decomposition

$$\bar{K}(x) = \begin{pmatrix} 1 & \frac{1}{2}x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -x + \frac{1}{16}x^3 & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2}x \\ 0 & 1 \end{pmatrix}, \quad (2.16)$$

which is not of the form (2.2).  $\square$

## 2.2 From the stability polynomial to the splitting method

We have seen that, under certain circumstances, the stability matrix  $K(x)$  allows us to get the composition (2.2) in a unique way and therefore the values of the coefficients of the splitting method. The question we analyze now is whether something similar can be done starting from the stability polynomial. We show that, given such a  $p(x)$ , there exists a finite number of different time-reversible splitting schemes having  $p(x)$  as their stability polynomial.

To begin with, suppose one has an even polynomial  $p(x)$  satisfying (2.10). Obviously, there exist infinite splitting methods having  $p(x)$  as their stability polynomial. Nevertheless, for each arbitrary even polynomial  $r(x) \neq p(x)$  with  $r(0) = 0$  there exist a *finite* number of consistent splitting methods with stability matrix (2.4) verifying  $K_1(x) = p(x) + r(x)$  and  $K_4(x) = p(x) - r(x)$ , or equivalently

$$p(x) = \frac{K_1(x) + K_4(x)}{2}, \quad r(x) = \frac{K_1(x) - K_4(x)}{2}.$$

The remaining entries of (2.4) are obtained by considering all possible decompositions of the polynomial  $p(x)^2 - r(x)^2 - 1 (= K_1(x)K_4(x) - 1)$  as the product of two odd polynomials  $K_2(x)$  and  $K_3(x)$  satisfying (2.9).

Proposition 2.1 gives then a decomposition of the form (2.11) for each choice of the stability matrix  $K(x)$ . Finally, all possible consistent splitting methods corresponding to the polynomials  $p(x)$  and  $r(x)$  are obtained by selecting, among the finite number of different decompositions of the form (2.11) obtained in that way, those that are actually of the form (2.2).

With the simplest choice  $r(x) \equiv 0$  one has  $K_1(x) \equiv K_4(x)$ . In that case the stability matrix verifies the identity  $K(x)^{-1} \equiv K(-x)$ , and this is precisely the characterization of a time-reversible method [11]. In terms of the composition (2.2) it corresponds to taking either  $a_{k+1-i} = a_i$ ,  $b_k = 0$ ,  $b_{k-i} = b_i$  or  $a_1 = 0$ ,  $a_{k+1-i} = a_{i+1}$ ,  $b_{k+1-i} = b_i$ ,  $i = 1, 2, \dots$ , which results in a left-right palindromic composition. In this way it is possible to construct explicitly all consistent time-reversible splitting methods with a prescribed stability polynomial  $p(x)$  satisfying (2.10).

**Example:** Let us consider the stability polynomial of  $K(x)$  and  $\bar{K}(x)$  in (2.13) and (2.15), respectively,

$$p(x) = 1 - \frac{1}{2}x^2 + \frac{1}{32}x^4, \quad \text{so that} \quad p(x)^2 - 1 = -x^2 \left(1 - \frac{x^2}{16}\right) \left(1 - \frac{x^2}{8}\right)^2 \quad (2.17)$$

There are six different ways of factorizing  $p(x)^2 - 1$  as a product of two odd polynomials  $K_2(x)$  and  $K_3(x)$  with  $K_2'(0) = 1 = -K_3'(0)$ . The three choices satisfying  $d(K_3) < d(K_4)$  are: (i)  $K_3(x) = -x \left(1 - \frac{x^2}{8}\right)$  which gives the stability matrix (2.13), with unique decomposition (2.14); (ii)  $K_3(x) = -x \left(1 - \frac{x^2}{16}\right)$  which gives the stability matrix (2.15) (with unique decomposition (2.16), and thus not corresponding to an splitting method of the form (1.2)); and (iii)  $K_3(x) = -x$  which leads to the stability matrix

$$\tilde{K}(x) = \begin{pmatrix} 1 - \frac{1}{2}x^2 + \frac{1}{32}x^4 & x - \frac{5}{16}x^3 + \frac{x^5}{32} - \frac{x^7}{1024} \\ -x & 1 - \frac{1}{2}x^2 + \frac{1}{32}x^4 \end{pmatrix}. \quad (2.18)$$

Application of the algorithm used in the proof of Proposition 2.1 allows us to factorize the matrix (2.18) as

$$\tilde{K}(x) = \begin{pmatrix} 1 & \frac{1}{2}x - \frac{1}{32}x^3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -x & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2}x - \frac{1}{32}x^3 \\ 0 & 1 \end{pmatrix}, \quad (2.19)$$

which does not correspond to an splitting method of the form (1.2), as expected because  $\tilde{K}(x)$  does not satisfy conditions (2.6). The remaining three stability matrices are obtained by interchanging the roles of the polynomials  $K_2(x)$  and  $-K_3(x)$ .  $\square$

### 2.3 Stability

According to the notion of stability given in the introduction, a splitting method is stable when applied to the harmonic oscillator if  $K(x)^n$  can be bounded independently of  $n \geq 1$ . As is well known, if the method is stable for a given  $x \in \mathbb{R}$ , then  $|p(x)| \leq 1$ . The converse is not in general true, as shown by the following simple example:

$$K(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \text{so that} \quad K(x)^n = \begin{pmatrix} 1 & nx \\ 0 & 1 \end{pmatrix},$$

and thus  $K(x)$  is linearly unstable. The following proposition gives an useful characterization of the stability of  $K(x)$ .

**Proposition 2.2** *Let  $K(x)$  be a  $2 \times 2$  matrix with  $\det K(x) = 1$ , and  $p(x) = \frac{1}{2}\text{tr} K(x)$ . Then  $K(x)$  is stable for a given  $x \in \mathbb{R}$  if and only if any of the following two conditions hold:*

1. *The matrix  $K(x)$  is diagonalizable with eigenvalues of modulus one,*
2.  *$|p(x)| \leq 1$ , and  $K(x)$  is similar to the matrix*

$$S(x) = \begin{pmatrix} \cos(\Phi(x)) & \sin(\Phi(x)) \\ -\sin(\Phi(x)) & \cos(\Phi(x)) \end{pmatrix}, \text{ where } \Phi(x) = \arccos(p(x)). \quad (2.20)$$

**Proof:** This is in fact an elementary consequence of the symplecticity of the matrix  $K(x)$ . In more detail, let  $\lambda_1(x)$  and  $\lambda_2(x)$  be the eigenvalues of  $K(x)$ . The assumption  $\det K(x) = 1$  implies that  $\lambda_1(x)\lambda_2(x) = 1$ . Thus, if  $K(x)$  is stable, then  $|\lambda_1(x)| = |\lambda_2(x)| = 1$ , and  $K(x)$  is necessarily diagonalizable, unless  $+1$  or  $-1$  is an eigenvalue with multiplicity 2. In both cases, if  $K(x)$  is not diagonalizable, then it is linearly unstable. Hence if  $K(x)$  is stable it is diagonalizable with eigenvalues of modulus 1.

The eigenvalues of  $K(x)$  are the zeroes of  $\lambda^2 - (\text{tr} K(x))\lambda + 1$ . Thus, if  $K(x)$  is diagonalizable with eigenvalues of modulus 1, then  $|p(x)| \leq 1$ , and

$$\lambda_1(x) = e^{i\Phi(x)}, \quad \lambda_2(x) = e^{-i\Phi(x)}, \quad (2.21)$$

where  $\Phi(x)$  is given by (2.20). In consequence,  $K(x)$  is similar to the matrix  $S(x)$  given in (2.20), as  $S(x)$  is also diagonalizable with eigenvalues (2.21).

Finally, the second condition clearly implies that the matrices  $S(x)$  and  $K(x)$  are both stable.  $\square$

Observe that  $S(x)$  can be considered as the stability matrix of a time-reversible method. In consequence, the above proposition allows us to conclude that for sufficiently small values of  $x$  each matrix  $K(x)$  with  $\det K(x) = 1$  is similar to a time-reversible matrix, and this provides an additional justification to taking only into account time-reversible splitting schemes.

In the stability analysis, the real parameters  $x_*$  and  $x^*$  defined next play a crucial role.

**Definition 2.3** *Given a  $2 \times 2$  matrix  $K(x)$  depending on a real parameter  $x$  such that  $\det K(x) \equiv 1$ , we denote by  $x_*$  the largest non-negative real number such that  $K(x)$  is stable for all  $x \in [-x_*, x_*]$ . We say that  $x_*$  is the stability threshold of  $K(x)$ , and that  $[-x_*, x_*]$  is the stability interval of  $K(x)$ .*

**Definition 2.4** *Let  $p(x)$  be an even polynomial depending on a real parameter  $x$  with  $p(0) = 1$ . We denote by  $x^*$  the largest real positive number such that  $|p(x)| \leq 1$  for all  $x \in [0, x^*]$ .*

**Remarks:**

1. From the proof of Proposition 2.2, it is clear that  $(-x^*, x^*)$  is the largest interval including 0 such that  $K(x)$  has eigenvalues of modulus 1.
2. If  $p''(0) < 0$ , then  $p(x)^2 - 1 \leq 0$  for sufficiently small  $|x|$ , and in that case  $x^*$  is the smallest real positive zero with odd multiplicity of the polynomial  $(p(x)^2 - 1)$  (for this  $x^*$ , the sign of the polynomial  $p(x)^2 - 1$  does not change in the interval  $x \in (-x^*, x^*)$ , but it does when crossing  $x = \pm x^*$ ).

If  $p(x)$  is the stability polynomial of  $K(x)$  then it is clear that  $x_* \leq x^*$ . Next we analyze which conditions have to be imposed on  $K(x)$  to get the optimal stability threshold, i.e., to ensure that  $x_* = x^*$ .



**Proposition 2.5** *Assume that a  $2 \times 2$  matrix  $K(x)$  depending on a real parameter  $x$  is such that  $\det K(x) \equiv 1$ ,  $K_1(x)$  and  $K_4(x)$  are even polynomials,  $K_2(x)$  and  $K_3(x)$  are odd polynomials,  $K_1(0) = K_4(0) = 1$ , and  $K_2'(0)K_3'(0) < 0$ . Let  $p(x) = \frac{1}{2} \operatorname{tr} K(x)$  be the corresponding stability polynomial and let  $0 = x_0 < x_1 < \dots < x_l$  be the real zeros with even multiplicity of the polynomial  $(p(x)^2 - 1)$  in the interval  $[0, x^*]$ . Then,  $x_* = x^*$  if*

$$K_2(x_j) = K_3(x_j) = 0 \quad (2.22)$$

for each  $j = 1, \dots, l$ . Otherwise,  $x_*$  is the smallest  $x_j$  that violates condition (2.22).

**Proof:** On the one hand, by differentiating  $\det K(x) \equiv 1$  twice, replacing  $x$  by 0 and taking into account that  $K_1(0) = K_4(0) = 1$  and  $K_2'(0)K_3'(0) < 0$ , we conclude that  $p''(0) = K_1''(0) + K_4''(0) = 2K_2'(0)K_3'(0) < 0$ , which guarantees that the eigenvalues of  $K(x)$  for  $x \in (-x^*, x^*)$  have modulus one, and thus  $K(x)$  is stable if and only if it is diagonalizable. When  $|p(x)| < 1$  the eigenvalues of  $K(x)$  are distinct, and thus  $K(x)$  is diagonalizable. If  $x \in (-x^*, x^*)$  and  $|p(x)| = 1$ , that is, if  $x = x_j$  for some  $j = 1, \dots, l$ , then  $K(x)$  has the double eigenvalue 1 or  $-1$ , and in that case,  $K(x)$  is diagonalizable if and only if  $K(x)$  or  $-K(x)$  are the identity matrix  $I$ , that is, if and only if (2.22) holds.  $\square$

Notice that the assumptions in Proposition 2.5 hold for the matrix  $K(x)$  (2.2) of any consistent splitting method (c.f., conditions (2.7)-(2.9)).

**Example:** Consider the polynomial  $p(x)$  given in (2.17). Then  $l = 1$ ,  $x_1 = 2\sqrt{2}$ , and  $x^* = 4$ . For the stability matrix (2.13) we have  $K_2(x_1) = K_3(x_1) = 0$ , and thus the corresponding stability threshold is  $x_* = x^* = 4$ . As for the stability matrices  $\tilde{K}(x)$  and  $\tilde{K}(x)$  in (2.15) and (2.18), respectively, one has  $\tilde{K}_3(x_1) = -\sqrt{2} \neq 0$  and  $\tilde{K}_3(x_1) = -2\sqrt{2} \neq 0$ , and thus  $x_* = 2\sqrt{2} < 4 = x^*$  in both cases.  $\square$

The above proposition allows us to build easily a stability matrix corresponding to a time-reversible splitting method with the optimal stability threshold:

**Proposition 2.6** *Let  $p(x)$  be an even polynomial satisfying (2.10). Then there exists a stability matrix of the form*

$$K(x) = \begin{pmatrix} p(x) & K_2(x) \\ K_3(x) & p(x) \end{pmatrix} \quad (2.23)$$

for which  $x_* = x^*$ .

**Proof:** Let  $0 = x_0 < x_1 < \dots < x_l$  be the real zeros with even multiplicity of the polynomial  $(p(x)^2 - 1)$  in the interval  $[0, x^*]$ . Then the polynomial  $p(x)^2 - 1$  can be decomposed as

$$p(x)^2 - 1 = -x^2 Q(x) \prod_{j=1}^l ((x/x_j)^2 - 1)^2,$$

where  $Q(x)$  is an even polynomial verifying  $Q(0) = 1$ . We then choose a decomposition  $Q(x) = Q_2(x)Q_3(x)$  of even polynomials such that  $Q_2(0) = Q_3(0) = 1$ , and determine  $K_2(x), K_3(x)$  as

$$K_2(x) = xQ_2(x) \prod_{j=1}^l ((x/x_j)^2 - 1), \quad K_3(x) = -xQ_3(x) \prod_{j=1}^l ((x/x_j)^2 - 1). \quad (2.24)$$

This completes the proof.  $\square$

## 2.4 Accuracy

As mentioned in the introduction, an accurate high order method can be useless in practice if it possesses a tiny stability domain. Similarly, a very stable but poorly accurate method is also of no interest in practical applications. In consequence, when designing new integration schemes of the form (2.2) the goal is to achieve the right balance between accuracy and stability. This, of course, is not an easy task in general, although a partial analysis has been done for the harmonic oscillator [9, 22].

Obviously, the accuracy of a splitting method depends on the difference between the matrix  $K(x)$  and the exact solution,  $O(x)$ . Roughly speaking, if  $\|K(x) - O(x)\|$  is small, then the eigenvalues of  $K(x)$  must be close to the eigenvalues of  $O(x)$ . According to Proposition 2.2, the stability polynomial  $p(x)$  must be an approximation to  $\cos(x)$ . In particular, for a splitting method of order  $2n$  it necessarily holds that

$$p(x) = \cos(x) + \mathcal{O}(x^{2n+2}) \quad \text{as} \quad x \rightarrow 0,$$

i.e.  $\Phi(x) = x + \mathcal{O}(x^{2n+1})$  in (2.20). In addition, under the assumptions of Proposition 2.5, if (2.2) is a good approximation to the solution matrix in a large subinterval of the stability interval  $[-x_*, x_*]$ , say  $[-x_k, x_k]$  for some  $k \leq l$ , then  $x_j \approx j\pi$  and  $p(x_j) = (-1)^j$  for  $j = 1, \dots, k$ .

It is worth stressing that if one is interested in processed splitting methods of the form (1.7), then, according to Proposition 2.2, their accuracy when applied to the harmonic oscillator only depends on the quality of the approximation  $p(x) \approx \cos(x)$  of their stability polynomial  $p(x)$  in the stability interval  $[-x_*, x_*]$ .

## 2.5 Geometric properties

Consider the application of a consistent splitting method (with stability matrix  $K(x)$  and stability polynomial  $p(x)$ ) to the harmonic oscillator (1.4) split as (1.5). According to Proposition 2.2, if  $h\lambda \in (-x_*, x_*)$ , one step of the method is conjugate to the exact  $h$ -flow of the modified harmonic oscillator

$$y'' + \tilde{\lambda}(h)^2 y = 0, \tag{2.25}$$

with  $\tilde{\lambda}(h) = \frac{1}{h}\Phi(h\lambda)$ . In other words, there exists a well defined matrix

$$P(x) = \begin{pmatrix} P_1(x) & P_2(x) \\ P_3(x) & P_4(x) \end{pmatrix} \tag{2.26}$$

(with  $\det P(x) = 1$ ) such that

$$S(x) = P(x)K(x)P^{-1}(x) \tag{2.27}$$

is given in (2.20). In the particular case of time-reversible methods (i.e., satisfying that  $K(-x) = K^{-1}(x)$  or equivalently  $K_4(x) = K_1(x)$ ), one can choose

$$P(x) = \begin{pmatrix} P_1(x) & 0 \\ 0 & P_4(x) \end{pmatrix}, \tag{2.28}$$

where

$$P_1(x) = \sqrt[4]{\frac{-K_3(x)}{K_2(x)}}, \quad P_4(x) = \sqrt[4]{\frac{-K_2(x)}{K_3(x)}}. \tag{2.29}$$

Obviously, these expressions are only valid when  $K_2(x) \neq 0 \neq K_3(x)$ . Otherwise,  $K_2(x) = 0 = K_3(x)$  for  $x \in (-x_*, x_*)$ , and then  $P_1(x) = 1 = P_4(x)$ .

The exact solution  $O(x)$  is an orthogonal matrix. Alternatively, the complex vector  $u = q + ip$  evolves through a unitary operator, i.e.  $u(h) = U(x)u(0)$  with  $U(x)$  a unitary matrix (in fact,  $u$  verifies the differential equation  $iu' = \lambda u$ ). Although a splitting method of the form (2.2) does not preserve the unitarity of  $U(x)$  or, equivalently, the orthogonality of  $O(x)$ , the previous considerations show that the average relative errors due to the lack of preservation of unitarity or orthogonality do not grow with time, since the scheme is conjugate (when  $|h\lambda| < x_*$ ) to orthogonal or unitary methods.

**Definition 2.7** We denote by  $r^*$  the radius of convergence of the expansion in powers of  $x$  of  $\Phi(x) = \arccos(p(x)) = \arcsin \sqrt{1 - p(x)^2}$ , that is, the maximum of the modulus of the (non-necessarily real) zeroes of  $1 - p(x)^2$  with odd multiplicity.

Notice that  $r^* \leq x^*$ , but not necessarily  $r^* \leq x_*$ . Whence,  $\Phi(x) = x + \phi_3 x^3 + \phi_5 x^5 + \dots$  when  $|x| < r^*$ . Then, one step of the method is conjugate to the exact  $h$ -flow of the modified harmonic oscillator (2.25) with

$$\tilde{\lambda}(h) = \lambda + h^2 \phi_3 \lambda^3 + h^4 \phi_5 \lambda^5 + \dots \quad (2.30)$$

whenever  $|h\lambda| < \min(x_*, r^*)$ .

### 3 Application of splitting methods to linear systems

One could reasonably argue that there is not much interest in designing new splitting methods with high accuracy and enlarged stability for the numerical integration of the simple harmonic oscillator. There are, however, at least two different issues to be taken into account in regarding this assertion. First, it is unlikely that a splitting method applied to an arbitrary nonlinear system provides good efficiency if it performs poorly when applied to the harmonic oscillator. In particular, good stability and accuracy for the harmonic oscillator is a necessary condition for a good performance when applied to systems that can be considered as perturbations of harmonic oscillators. Second, there are several PDEs modeling highly relevant physical phenomena that, once spatially discretized, give rise to systems of coupled harmonic oscillators where the previous analysis can be used to build accurate and stable algorithms for their numerical treatment. In the sequel we briefly consider four classes of linear systems for which the results in Section 2 are of interest.

(i) As a first instance, we consider the time-dependent Schrödinger equation

$$i \frac{\partial}{\partial t} \psi(x, t) = \left( -\frac{1}{2\mu} \nabla^2 + V(x) \right) \psi(x, t), \quad \psi(x, 0) = \psi_0(x), \quad (3.1)$$

where  $\psi(x, t) : \mathbb{R}^D \times \mathbb{R} \rightarrow \mathbb{C}$  is the wave function associated with the system. A common procedure for numerically solving this problem consists in taking first a discrete spatial representation of the wave function. For simplicity, let us consider the one-dimensional case and a given interval  $x \in [x_0, x_d]$  ( $\psi(x_0, t) = \psi(x_d, t) = 0$  or it has periodic boundary conditions). The interval is split in  $d$  parts of length  $\Delta x = (x_d - x_0)/d$  and the vector  $u = (u_0, \dots, u_{d-1})^T \in \mathbb{C}^d$  is formed, with  $u_j = \psi(x_j, t)$  and  $x_j = x_0 + j\Delta x$ ,  $j = 0, 1, \dots, d-1$ . The partial differential equation (3.1) is then replaced by the  $d$ -dimensional linear ODE

$$iu'(t) = H u(t), \quad (3.2)$$

where  $H \in \mathbb{R}^{d \times d}$  represents the (real symmetric) matrix associated with the Hamiltonian. Complex vectors can be avoided by writing  $u = q + ip$ , with  $q, p \in \mathbb{R}^d$ . Equation (3.2) is then equivalent to [9, 10, 22, 32, 3, 17]

$$q' = Hp, \quad p' = -Hq. \quad (3.3)$$

Let  $H$  be factorized as  $H = U^{-1}\Lambda U$ , where  $\Lambda$  is the diagonal matrix containing the eigenvalues of  $H$ . The system (3.3) can be decoupled, after the change of variables  $Q = Uq$ ,  $P = Up$ , into a system of  $d$  one-dimensional harmonic oscillators

$$Q' = \Lambda P, \quad P' = -\Lambda Q. \quad (3.4)$$

However, since in practice  $d \gg 1$ , it turns prohibitively expensive to carry out the diagonalization of  $H$ . On the other hand, due to the nature of this problem, Fourier techniques can be used and the products  $Hq, Hp$  can be evaluated with  $\mathcal{O}(d \log(d))$  operations with the Fast Fourier Transform (FFT) algorithm. Thus, numerical methods requiring only the computation of matrix-vector products of the form  $Hq$  and  $Hp$  may lead to very efficient integration algorithms. This is precisely the case when the splitting method (1.2) is applied to (3.3) with

$$f^{[A]}(z) = (Hp, 0)^T, \quad f^{[B]}(z) = (0, -Hq)^T, \quad z = (q, p)^T,$$

as the corresponding  $h$ -flows  $\varphi_h^{[A]}$  and  $\varphi_h^{[B]}$  are

$$\varphi^{[A]}(z) = (q + hHp, p), \quad \varphi^{[B]}(z) = (q, p - hHq).$$

Since by assumption  $H$  is a symmetric matrix, the solution operator of (3.3) is orthogonal. Equivalently, the vector  $u$  associated with the wave function evolves through a unitary operator, and the norm of  $u$  is preserved. Then, as we have seen in section 2.5, when a splitting method is applied, the averaged errors in the preservation of unitarity do not grow with time (a fact already noticed numerically in [28]).

Let us analyze this issue in more detail. Clearly, applying a splitting method to (3.3) is equivalent, after the change of variables  $Q = Uq$ ,  $P = Up$ , to applying the same method to the system (3.4) of decoupled harmonic oscillators. Hence, one step of the method is conjugate to the exact  $h$ -flow of harmonic oscillators of the form (2.25) with  $\tilde{\lambda}(h) = \Phi(h\lambda)/h$ . In consequence, by reversing the change of variables, one step of the method is conjugate to the exact  $h$ -flow of the modified system

$$q' = \tilde{H}(h)p, \quad p' = -\tilde{H}(h)q, \quad (3.5)$$

(or equivalently,  $iu' = \tilde{H}(h)u$ ) for values of  $h$  such that  $|h|\rho(H) < x_*$ , where  $\rho(H)$  is the spectral radius of  $H$ . Here  $\tilde{H}(h) = \frac{1}{h}U^{-1}\Phi(h\Lambda)U$  (and  $\Phi$  is applied componentwise to each entry of the diagonal of the diagonal matrix  $h\Lambda$ ). If  $|h|\rho(H) < \min(r^*, x_*)$ , we have that

$$\tilde{H}(h) = H + h^2\phi_3H^3 + h^4\phi_5H^5 + \dots \quad (3.6)$$

Notice that  $\tilde{H}(h)$  is obviously symmetric, provided that  $H$  is symmetric.

(ii) Another system which can also be decoupled into a number of one dimensional harmonic oscillators is

$$y'' + Ky = 0, \quad (3.7)$$

where  $y \in \mathbb{R}^d$  and the matrix  $K \in \mathbb{R}^{d \times d}$  is diagonalizable with real positive eigenvalues ( $K$  can be, in particular, a real symmetric positive definite matrix). Systems of this form arise after a semidiscretization of some parabolic PDEs whose linear part has to be efficiently computed (see Chapter XIII of [11] and references therein).

In this case, one can use a change of variables of the form  $P = \Lambda^{-1}Uy'$ ,  $Q = Uy$  (where  $\Lambda$  is the diagonal matrix such that  $K = U^{-1}\Lambda^2U$ ) to show that the application of one step of size  $h$  of a splitting method to the system (3.7) is, provided that  $|h|\sqrt{\rho(K)} < x_*$ , conjugate to the exact  $h$ -flow of the modified system

$$y'' + \tilde{K}(h)y = 0,$$

where  $\tilde{K}(h)$  is a perturbation of  $K$  defined in terms of  $\Phi$ . In particular,  $\tilde{K}(h) = K + h^2\phi_3K^2 + h^4\phi_5K^3 + \dots$  whenever  $|h|\sqrt{\rho(K)} < \min(r^*, x_*)$ .

(iii) A more general class of problems which can be reduced to a system of decoupled harmonic oscillators is the following:

$$q' = Mp, \quad p' = -Nq. \quad (3.8)$$

Here  $q \in \mathbb{R}^{d_1}$ ,  $p \in \mathbb{R}^{d_2}$ ,  $M \in \mathbb{R}^{d_1 \times d_2}$  and  $N \in \mathbb{R}^{d_2 \times d_1}$ , with the additional constraint that either  $MN$  or  $NM$  is diagonalizable with real positive eigenvalues. Notice that this class includes both (3.3) and (3.7) as particular instances, with  $M = N = H$  in the first case and  $M = I$ ,  $N = K$  in the second one.

Systems of the form (3.8) arise, for example, after spacial discretization of the Maxwell equations, of relevant interest in physics and engineering [30, 25, 12].

The relation of (3.8) with (3.7) becomes clear by observing that the solution of (3.8) can be provided by solving either

$$q'' + MNq = 0, \quad p' = -Nq,$$

or

$$p'' + NMp = 0, \quad q' = Mp.$$

As a matter of fact, a splitting method applied to (3.8) is conjugate, for sufficiently small  $h$ , to the solution of the modified system

$$q' = \tilde{M}(h)p, \quad p' = -\tilde{N}(h)q, \quad (3.9)$$

with

$$\tilde{M}(h) = M(I + \phi_3h^2(NM) + \phi_5h^4(NM)^2 + \dots) \quad (3.10)$$

$$= (I + \phi_3h^2(MN) + \phi_5h^4(MN)^2 + \dots)M,$$

$$\tilde{N}(h) = N(I + \phi_3h^2(MN) + \phi_5h^4(MN)^2 + \dots) \quad (3.11)$$

$$= (I + \phi_3h^2(NM) + \phi_5h^4(NM)^2 + \dots)N.$$

This can be established, as in the cases (i) and (ii) before, in two particular situations:

- (a) The matrix  $MN$  is diagonalizable with positive eigenvalues, and in addition,  $M$  and  $N$  are square matrices and  $M$  is invertible. In that case, let  $MN$  be decomposed as  $MN = U^{-1}\Lambda^2U$ , where  $\Lambda$  is a diagonal matrix. Then the system (3.8) is decoupled as (3.4) with the transformation  $P = \Lambda^{-1}UMp$ ,  $Q = Uq$ . A similar argument to that used in (i) leads to the required result whenever  $|h|\sqrt{\rho(MN)} < \min(r^*, x_*)$ .
- (b) The matrix  $NM$  is diagonalizable with positive eigenvalues, and in addition,  $M$  and  $N$  are square matrices and  $N$  is invertible. Then, a similar argument leads to the required result whenever  $|h|\sqrt{\rho(NM)} < \min(r^*, x_*)$ .

In general, we will next show that there exists  $r_* > 0$  (satisfying that  $r_* \leq \min(r^*, x_*)$ ), depending on the parameters  $a_j, b_j$  of the splitting method such that the approximate solution of (3.8) obtained is conjugate to the solution of the modified system (3.9)–(3.11) whenever  $|h|\sqrt{\rho_{M,N}} < r_*$ , where

$$\rho_{M,N} = \min(\rho(NM), \rho(MN)). \quad (3.12)$$

As this assertion can be proved for arbitrary matrices  $N$  and  $M$  (without any constraint about the square matrices  $NM$  and  $MN$ ), it is worth considering a more general class of systems:

(iv) Let us examine now a linear system of the form (3.8) with arbitrary matrices  $N$  and  $M$ . One step of the splitting method (1.2) applied to that system with

$$f^{[A]}(z) = (Mp, 0)^T, \quad f^{[B]}(z) = (0, -Nq)^T, \quad z = (q, p)^T,$$

gives  $\psi_h(z) = \mathbf{K}(h)z$ , where

$$\mathbf{K}(h) = \begin{pmatrix} I & 0 \\ -b_k N & I \end{pmatrix} \begin{pmatrix} I & a_k M \\ 0 & I \end{pmatrix} \cdots \begin{pmatrix} I & 0 \\ -b_1 N & I \end{pmatrix} \begin{pmatrix} I & a_1 M \\ 0 & I \end{pmatrix}. \quad (3.13)$$

Studying the application of splitting methods to the system (3.8) is of interest, for instance, when considering the linearization around a stationary point of a Hamiltonian system with Hamiltonian function of the form  $H(q, p) = T(p) + V(q)$ .

Clearly,  $\mathbf{K}(h)$  can be rewritten in the form

$$\mathbf{K}(h) = \begin{pmatrix} I_{d_1} + \sum_{j \geq 1} k_{1,j} h^{2j} (MN)^j, & \sum_{j \geq 1} k_{2,j} h^{2j-1} M (NM)^{j-1} \\ \sum_{j \geq 1} k_{3,j} h^{2j-1} N (MN)^{j-1} & I_{d_2} + \sum_{j \geq 1} k_{4,j} h^{2j} (NM)^j \end{pmatrix}. \quad (3.14)$$

Here the coefficients  $k_{i,j}$  are those of the polynomial entries (2.5) of the stability matrix (2.4) of the splitting method. It is worth noting that one step  $\psi_h(z)$  of an arbitrary partitioned Runge-Kutta (PRK) method applied to the partitioned system (3.8) is also of the form  $\psi_h(z) = \mathbf{K}(h)z$ , where the matrix  $\mathbf{K}(h)$  is of the form (3.14). For implicit PRK methods there is an infinite number of non-zero coefficients  $k_{i,j}$ , and the expressions

$$\begin{aligned} K_1(x) &= 1 + \sum_{j \geq 1} k_{1,j} x^{2j}, & K_2(x) &= \sum_{j \geq 1} k_{2,j} x^{2j-1} \\ K_3(x) &= \sum_{j \geq 1} k_{3,j} x^{2j-1}, & K_4(x) &= 1 + \sum_{j \geq 1} k_{4,j} x^{2j} \end{aligned} \quad (3.15)$$

are obtained by expanding in series of powers of  $x$  certain rational functions (the entries of the stability matrix  $K(x)$  of the method). For explicit PRK methods,  $K_1(x), K_2(x), K_3(x), K_4(x)$  are polynomial functions.

The next result shows that any partitioned method (splitting methods, partitioned Runge-Kutta methods) applied to the system (3.8) is, for a sufficiently small step size  $h$ , conjugate to a non-partitioned method (defined as a power series expansion).

**Proposition 3.1** *Consider a  $2 \times 2$  matrix (2.4) depending on the variable  $x$  whose entries are defined as power series (3.15) with non-zero radius of convergence. For arbitrary matrices  $M \in \mathbb{R}^{d_1 \times d_2}$ ,  $N \in \mathbb{R}^{d_2 \times d_1}$ , consider the square matrix of dimension  $(d_1 + d_2)$  given by (3.14).*

Then, there exists  $r > 0$  and a power series  $\sum_{j \geq 1} s_j x^j$  such that (3.14) is, provided that  $|h| \sqrt{\rho_{M,N}} < r$ , similar to

$$\mathbf{S}(h) = I_{d_1+d_2} + \sum_{j \geq 1} s_j h^j \begin{pmatrix} 0 & M \\ -N & 0 \end{pmatrix}^j. \quad (3.16)$$

In addition, if  $\det K(x) \equiv 1$  (in particular, if  $K(x)$  is the stability matrix of a splitting method), then  $\mathbf{S}(h)$  is the exponential of the matrix

$$h \begin{pmatrix} 0 & \tilde{M}(h) \\ -\tilde{N}(h) & 0 \end{pmatrix},$$

where  $\tilde{M}(h)$  and  $\tilde{N}(h)$  are given by (3.10)–(3.11), and  $x + \phi_3 x^3 + \phi_5 x^5 + \dots$  is the power series expansion of

$$\Phi(x) = \arccos \frac{K_1(x) + K_4(x)}{2} = \arcsin \sqrt{1 - \left( \frac{K_1(x) + K_4(x)}{2} \right)^2}.$$

**Proof:** We first show that there exist  $r > 0$  and a  $2 \times 2$  matrix (2.26) whose entries are defined as power series

$$\begin{aligned} P_1(x) &= 1 + \sum_{j \geq 1} p_{1,j} x^{2j}, & P_2(x) &= \sum_{j \geq 1} p_{2,j} x^{2j-1}, \\ P_3(x) &= \sum_{j \geq 1} p_{3,j} x^{2j-1}, & P_4(x) &= 1 + \sum_{j \geq 1} p_{4,j} x^{2j}, \end{aligned}$$

with non-zero radius of convergence, such that, for  $|x| < r$ ,  $P(x)K(x)P(x)^{-1}$  is well defined and has the form

$$S(x) = \begin{pmatrix} S_1(x) & S_2(x) \\ -S_2(x) & S_1(x) \end{pmatrix}, \quad (3.17)$$

where

$$S_1(x) = 1 + \sum_{j \geq 1} (-1)^j s_{2j} x^{2j}, \quad S_2(x) = \sum_{j \geq 1} (-1)^{j-1} s_{2j-1} x^{2j-1}. \quad (3.18)$$

Once this is proven, it is straightforward to check that, for the square matrix  $\mathbf{P}(h)$  of dimension  $(d_1 + d_2)$

$$\mathbf{P}(h) = \begin{pmatrix} I_{d_1} + \sum_{j \geq 1} p_{1,j} h^{2j} (MN)^j, & \sum_{j \geq 1} p_{2,j} h^{2j-1} M(NM)^{j-1} \\ \sum_{j \geq 1} p_{3,j} h^{2j-1} N(MN)^{j-1} & I_{d_2} + \sum_{j \geq 1} p_{4,j} h^{2j} (NM)^j \end{pmatrix}, \quad (3.19)$$

the matrix  $\mathbf{P}(h)\mathbf{K}(h)\mathbf{P}(h)^{-1}$  coincides, whenever  $|h| \sqrt{\rho_{M,N}} < r$ , with

$$\begin{pmatrix} I_{d_1} + \sum_{j \geq 1} (-1)^j s_{2j} h^{2j} (MN)^j, & \sum_{j \geq 1} (-1)^{j-1} s_{2j-1} h^{2j-1} M(NM)^{j-1} \\ \sum_{j \geq 1} (-1)^j s_{2j-1} h^{2j-1} N(MN)^{j-1} & I_{d_2} + \sum_{j \geq 1} (-1)^j s_{2j} h^{2j} (NM)^j \end{pmatrix},$$

which is precisely (3.16).

Indeed, one can directly check that  $P(x)K(x)P(x)^{-1} = S(x)$  with

$$\begin{aligned} S_1(x) &= \frac{K_1(x) + K_4(x)}{2}, & S_2(x) &= \sqrt{\det(K(x)) - S_1(x)^2}, \\ P_1(x) &= \sqrt{\frac{-K_3(x)}{S_2(x)}}, & P_2(x) &= \frac{K_1(x) - K_4(x)}{2S_2(x)P_1(x)}, & P_3(x) &= 0, & P_4(x) &= \frac{1}{P_1(x)}, \end{aligned} \quad (3.20)$$

and  $r$  can be taken as the minimum among the radii of convergence of the series  $K_j(x), S_j(x), P_j(x)$ . According to (3.18),  $S_1(x) + iS_2(x) = \sum s_j(ix)^j$ , and thus  $\sum s_j x^j$  is the series expansion of  $S_1(-ix) + iS_2(-ix)$  where  $S_1(x)$  and  $S_2(x)$  are given by (3.20). If  $\det K(x) \equiv 1$  (in particular, if  $K(x)$  is the stability matrix of a splitting method), we have that

$$S_1(x) = \frac{K_1(x) + K_4(x)}{2} = \cos(\Phi(x)), \quad S_2(x) = \sqrt{1 - S_1(x)^2} = \sin(\Phi(x)),$$

where

$$\Phi(x) = \arccos S_1(x) = \arcsin \sqrt{1 - S_1(x)^2},$$

and thus  $\sum s_j x^j$  is the series expansion of  $\exp(i\Phi(-ix))$ . If  $x + \phi_3 x^3 + \phi_5 x^5 + \dots$  is the power series expansion of  $\Phi(x)$ , it is clear that  $i\Phi(-ix) = x - \phi_3 x^3 + \phi_5 x^5 - \dots$  and a simple calculation shows that  $\mathbf{S}(h)$  is the exponential of

$$i\Phi \left( -ih \begin{pmatrix} 0 & M \\ -N & 0 \end{pmatrix} \right) = h \begin{pmatrix} 0 & \tilde{M}(h) \\ -\tilde{N}(h) & 0 \end{pmatrix},$$

thus completing the proof.  $\square$

Notice that, in the proof above, there are other choices for  $P(x)$ . For instance, we could have required that  $P_2(x) = 0, P_4(x) = 1/P_1(x)$ . Different choices for  $P(x)$  in general will give a different value of  $r$ . Let us denote as  $r_*$  the maximal  $r$  for which the statement of Proposition 3.1 holds.

For splitting methods, we always have that  $r_* \leq r^*$  and  $r_* \leq x_*$ . The following generalization of Proposition 2.6 will be useful in the next section.

**Proposition 3.2** *Suppose we have an even polynomial  $p(x)$  satisfying (2.10), so that  $0 < r^* \leq x^*$ . Then, there exists a time-reversible stability matrix of the form*

$$K(x) = \begin{pmatrix} p(x) & K_2(x) \\ K_3(x) & p(x) \end{pmatrix} \quad (3.21)$$

for which  $x_* = x^*$  and  $r_* = r^*$ .

**Proof:** According to Definition 2.7  $r^* > 0$  is the maximum of the modulus of the zeroes of  $1 - p(x)^2$  with odd multiplicity. Now let  $0, \pm x_1, \dots, \pm x_l$  be all the zeros with even multiplicity of the polynomial  $1 - p(x)^2$  with modulus  $|x_j| < r^*$ . Then the polynomial  $1 - p(x)^2$  can be decomposed as

$$1 - p(x)^2 = x^{2m_0} Q(x) \prod_{j=1}^l ((x/x_j)^2 - 1)^{2m_j},$$

where each  $m_0, m_1, \dots, m_l$  is the multiplicity of the zeroes  $0, \pm x_1, \dots, \pm x_l$ , respectively, and  $Q(x)$  is an even polynomial satisfying that  $Q(0) = 1$  and has no zeroes with even multiplicity. We thus have that

$$\begin{aligned} \Phi(x) &= \arccos(p(x)) \\ &= \arcsin(\sqrt{1 - p(x)^2}) \\ &= \arcsin(\sqrt{Q(x)} x^{m_0} \prod_{j=1}^l ((x/x_j)^2 - 1)^{m_j}), \end{aligned}$$



so that  $r^*$  is the maximum of the modulus of the zeroes of  $Q(x)$ . We then choose a decomposition  $Q(x) = Q_2(x)Q_3(x)$  of even polynomials such that  $Q_2(0) = Q_3(0) = 1$ , and determine  $K_2(x), K_3(x)$  as

$$K_2(x) = x^{m_0} Q_2(x) \prod_{j=1}^l ((x/x_j)^2 - 1)^{m_j}, \quad K_3(x) = -x^{m_0} Q_3(x) \prod_{j=1}^l ((x/x_j)^2 - 1)^{m_j}. \quad (3.22)$$

This completes the proof, as, according to (2.29),  $P_1(x) = \sqrt[4]{Q_3(x)/Q_2(x)}$  and  $P_4(x) = \sqrt[4]{Q_2(x)/Q_3(x)}$ , and the radius of convergence of their powers series expansions is precisely the maximum of the modulus of the zeroes of  $Q(x) = Q_2(x)Q_3(x)$ , that is,  $r^*$ .  $\square$

## 4 Construction of processed splitting methods with high accuracy and enlarged stability domain

The theoretical analysis done in the previous sections shows in particular that the stability polynomial  $p(x)$  carries all the information needed to construct processed splitting methods for numerically approximating the evolution of the harmonic oscillator. Our aim in this section is precisely to use this analysis to obtain efficient processed splitting methods to solve numerically the linear system (3.8). In other words, our goal is to approximate the exponential

$$\exp \left[ t \begin{pmatrix} 0 & M \\ -N & 0 \end{pmatrix} \right] \quad (4.1)$$

by means of

$$\mathbf{S}(h)^m = \mathbf{P}(h) [\mathbf{K}(h)]^m \mathbf{P}(h)^{-1},$$

where  $h = t/m$  is sufficiently small, the kernel  $\mathbf{K}(h)$  is given by (3.13) (which can be rewritten as (3.14)), and  $\mathbf{P}(h)$  is defined in terms of power series expansions as in the proof of Proposition 3.1 provided that  $h\sqrt{\rho_{M,N}} < r_*$  (where  $\rho_{M,N}$  is given by (3.12)).

We will concentrate ourselves in the case where either  $NM$  or  $MN$  is diagonalizable with positive eigenvalues, so that the performance of the method will be directly related to the accuracy and stability of the method when applied to the harmonic oscillator.

In practice,  $\mathbf{P}(h)$  and  $\mathbf{P}(h)^{-1}$  will be approximated by polynomials, and one will be able to approximate the action of the exponential (4.1) on the vector  $(q, p)^T$  by means of matrix-vector products of the form  $Mp$  and  $Nq$ .

Taking previous considerations into account, we consider a composition of the form (3.13) whose stability polynomial  $p(x)$  fulfills the following requirements:

- C.1  $p(x) = \cos(x) + \mathcal{O}(x^{2n+2})$  as  $x \rightarrow 0$  for certain  $n \geq 1$  (thus achieving effective order  $q = 2n$ ).
- C.2 There exist  $l \geq 1$  and  $x_j \in \mathbb{R}$ ,  $j = 1, \dots, l$ , such that  $0 < x_1 < \dots < x_l$ , each  $x_j$  is a double zero of the polynomial  $p(x) - (-1)^j$ , and all the (non-necessarily real) zeroes of odd multiplicity of the polynomial  $(p(x)^2 - 1)$  have modulus greater than  $x_l$  (so that  $x^* \geq r^* > x_l$ ).
- C.3 For each  $j = 1, \dots, l$ ,  $x_j = j\pi$ .

Once the polynomial  $p(x)$  is determined, it still remains to determine a stability matrix  $K(x)$ . In general, we have that  $x_* \leq x^*$  and  $r_* \leq r^*$ . The optimal case is achieved when  $x_* = x^*$  and  $r_* = r^*$  hold, and any processed method for which these two equalities hold are equivalent. We thus construct symmetric stability matrices  $K(x)$  for which  $x_* = x^*$  and  $r_* = r^*$  (this is always possible according to Proposition 3.2). Once this matrix  $K(x)$  is determined, the algorithm described in the proof of Proposition 2.1 will give the coefficients  $a_i, b_i$  for the kernel leading to the composition (3.13). Finally, we provide simple procedures to obtain pre- and post-processors from the polynomials  $K_i(x)$ .

It is important to keep in mind that the cost of a composition method with polynomial stability  $p(x)$  is proportional to the degree of  $p(x)$ . We can always improve both the stability and the accuracy of  $p(x)$  by increasing its degree, but this makes only sense if the improvement compensates the extra computational cost.

#### 4.1 Construction of stability polynomials

According to the previous requirements, we take as candidates for  $p(x)$  polynomials in the following family. For each  $n, l \geq 0$  we consider the even polynomial  $p^{n,l}(x)$  with minimal degree among those satisfying (i)  $p^{n,l}(x) = \cos(x) + \mathcal{O}(x^{2n+2})$  as  $x \rightarrow 0$ , and (ii) for each  $1 \leq j \leq l$ ,  $x_j = j\pi$  is a double zero of  $(p^{n,l}(x) - (-1)^j)$ . For arbitrary  $n, l \geq 1$  we take  $p^{n,l}(x)$  as

$$p^{n,l}(x) = 1 + \sum_{j=1}^n (-1)^j \frac{x^{2j}}{(2j)!} + x^{2n} \sum_{j=1}^{2l} d_j x^{2j}, \quad (4.2)$$

where the coefficients  $d_j$  are uniquely determined by the requirement that

$$p(j\pi) = (-1)^j, \quad p'(j\pi) = 0, \quad j = 1, \dots, l \quad (4.3)$$

holds for  $p(x) = p^{n,l}(x)$ . Notice the interpolatory nature of  $p^{n,l}(x)$  (as  $\cos(j\pi) = (-1)^j$  and  $\cos'(j\pi) = -\sin(j\pi) = 0$ ). By using a symbolic algebra package one gets the numerical values of  $d_j$ ,  $j = 1, \dots, 2l$  with the desired accuracy and for any value of  $l$  of practical interest.

With the polynomial  $p^{n,l}(x)$  it is possible to get an estimate of the relative performance of the splitting methods which can be obtained from it. Since a  $k$ -stage method has a stability polynomial of degree  $2k$  we can take the degree of  $p^{n,l}(x)$  ( $d(p^{n,l}) = 2(n + 2l)$ ) as twice the number of stages needed for the composition methods. The accuracy can be measured by evaluating the error function  $\mathcal{E}_{n,l}(x) \equiv |\arccos(p^{n,l}(x)) - x|$  for different values of  $x \in [0, x_*]$ . Figure 1(a) shows  $\mathcal{E}_{n,l}(x)$  versus  $\text{COST} = (n + 2l)/x$  (to measure the accuracy at a given cost, where  $x$  plays the role of the time step) for  $n = 5, l = 3, 5, 7$  and  $n = 10, l = 6, 10, 14$  corresponding to approximations of order 10 and 20 (each choice for  $p^{n,l}(x)$  is denoted by  $(n, l, 0)$ ). The curves which stay below in the same picture show the highest performance, i.e. they give the smallest error at a given cost. From the figures we observe the surprising fact that the performance seems to improve with  $l$  for each fixed value of  $n$ , and for most values of  $\text{COST}$ . This occurs for nearly all values of  $n$  and  $l$  checked, up to  $(n + 2l) = 50$ , corresponding to methods with up to 50 stages. Figure 1(b) shows the results for  $(n, l) = (5, 7), (8, 12), (10, 14)$  which would correspond to methods with 19, 32 and 38 stages.

On the other hand, if one is interested in approximating the solution matrix  $O(x)$  accurately for  $x \in (-x_*, x_*)$  with  $|x|$  as large as possible, we may require in addition to C.1–C.3 above,

C.4 for a relatively low effective order  $q = 2n$ , the stability polynomial  $p(x)$  approximates  $\cos(x)$  with acceptable precision for all  $x \in (-x_l, x_l)$ .

Figure 1: Error versus  $\text{COST} = (n + 2l)/x$  in logarithmic scale: (a) for stability polynomials  $p^{n,l}(x)$  with  $n = 5, 10$  (corresponding to approximations of order 10 and 20) and different values of  $l$ , which are denoted by  $(n, l, 0)$ ; (b) for  $(n, l) = (5, 7), (8, 12), (10, 14)$ ; (c) the same for  $p^{n,l,m}(x)$  with  $(n, l, m) = (1, 7, 4), (1, 12, 7), (1, 14, 9)$  corresponding to polynomials of the same degree as in (b); (d) comparison of the most efficient stability polynomials.

With the purpose of fulfilling this goal, we propose considering the following polynomials as candidates for the stability polynomial  $p(x)$ . For each  $n, l, m \geq 0$ , we take

$$p^{n,l,m}(x) = \sum_{j=1}^n (-1)^j \frac{x^{2j}}{(2j)!} + x^{2n} \sum_{j=1}^{2l} d_j x^{2j} + x^{2n} \prod_{j=1}^l (x^2 - (j\pi)^2)^2 \sum_{i=1}^m e_i x^{2i}, \quad (4.4)$$

satisfying (4.3) for  $p(x) = p^{n,l,m}(x)$ , so that the coefficients  $d_j$  ( $1 \leq j \leq 2l$ ) verify the same conditions as before (and thus they are uniquely determined). We have now the free parameters  $e_i$  ( $1 \leq i \leq m$ ), which we propose to determine in such a way that

$$\int_{-l\pi}^{l\pi} \left(1 - \left(\frac{x}{l\pi}\right)^2\right)^{-1/2} \left(\frac{p^{n,l,m}(x) - \cos(x)}{x^{2n+2}}\right)^2 dx \quad (4.5)$$

is minimized. Notice that this is equivalent to minimizing in the least square sense the coefficients of the Chebyshev series expansion of  $(p^{n,l,m}(x) - \cos(x))/(x^{2n+2})$ , which depend linearly on  $e_i$  ( $1 \leq i \leq m$ ), and thus the values of  $e_i$  can be exactly determined. In practice, this can be conveniently done with the help of some symbolic algebra program.

By following this approach we have analyzed stability polynomials with up to  $(n + 2l + m) = 50$  (corresponding to methods up to 50 stages) for different values of

$n, l, m$ . Analogously, to measure their relative performances, we compare the error  $\mathcal{E}_{n,l,m}(x) \equiv |\arccos(p^{n,l,m}(x)) - x|$  versus  $\text{COST} = (n + 2l + m)/x$  and for different real values of  $x$ . In Figure 1(c) we show the results obtained for some representative choices of  $p^{n,l,m}(x)$ , denoted by  $(n, l, m)$ , corresponding to polynomials of the same degree as those shown in Figure 1(b). In both cases, it seems that the performance improves with the number of stages. Figure 1(d) compares the performance of the best stability polynomial of each family (both correspond to polynomials of degree 76 and associated to 38-stage methods). It seems that up to a very high accuracy, the best performance corresponds to the second order method (1, 14, 9).

## 4.2 Construction of the stability matrix

Once a polynomial  $p(x)$  satisfying conditions C.1–C.3 (and possibly C.4) is determined, the next step is choosing the remaining entries of the stability  $K(x)$ . As previously stated, two different  $k$ -stage left-right palindromic compositions are considered:

$$K(h) = e^{ha_{k+1}A} e^{hb_kB} e^{ha_kA} \dots e^{hb_1B} e^{ha_1A}, \quad (4.6)$$

with  $a_{k+2-i} = a_i$ ,  $b_{k+1} = 0$ ,  $b_{k+1-i} = b_i$ ,  $i = 1, 2, \dots$  and

$$K(h) = e^{hb_{k+1}B} e^{ha_{k+1}A} e^{hb_kB} e^{ha_kA} \dots e^{hb_1B} \quad (4.7)$$

with  $a_1 = 0$ ,  $a_{k+2-i} = a_{i+1}$ ,  $b_{k+2-i} = b_i$ ,  $i = 1, 2, \dots$

In both cases  $p(x)$  has degree  $2k$ ,  $K_1(x) = K_4(x) = p(x)$ , and  $(K_2(x), K_3(x))$  have degree  $(2k - 1, 2k + 1)$  for (4.6) and degree  $(2k + 1, 2k - 1)$  for (4.7). The polynomials  $(K_2(x), K_3(x))$  are such that

$$\begin{aligned} p(x)^2 - 1 &= K_2(x)K_3(x), \\ K_2'(0) &= -K_3'(0) = 1, \\ K_2(x_j) &= K_3(x_j) = 0, \quad j = 1, \dots, l, \end{aligned}$$

and satisfying (2.24), which guarantee that  $\det K(x) = 1$ ,  $K(x)$  is a consistent approximation to (2.1),  $x_* = x^*$  and  $r_* = r^*$ . Clearly, there is a finite number of different choices of such pairs  $(K_2(x), K_3(x))$ , and among them, we choose a pair such that the corresponding matrix  $K(x)$  admits a decomposition of the form (2.2).

As representative of the methods which can be obtained, in Table 1 we show the coefficients  $a_i, b_i$  for the kernel of a pair of processed methods. Denoting a  $k$ -stage processed method of order  $q$  by  $P_{kq}$ , the first set of coefficients corresponds to  $P_{19}10$ , given by the composition (4.6) with  $k = 19$ , whereas the second belongs to  $P_{32}16$ , given by the composition (4.7) with  $k = 32$ . Coefficients for a method  $P_{38}2$  can be found in [3]. It is worth noticing the small size of all the coefficients.

## 4.3 Construction of the processor

As for the processor, since the kernel is time-reversible, it can be chosen as (2.28). Recall that, for their practical implementation,  $P_1(x)$  and  $P_4(x)$  must be replaced by polynomial approximations, say

$$P_1(x) = \sum_{i=0}^s c_i x^{2i}, \quad P_4(x) = \sum_{i=0}^s d_i x^{2i} \quad (4.8)$$

for a given  $s$ . In this way the constraint  $P_4 = P_1^{-1}$  is relaxed to  $P_4 = P_1^{-1} + \mathcal{O}(x^{2s+2})$ .

If we assume that  $S(x) = P(x)K(x)P(x)^{-1}$  with  $S(x)$  given by (2.20), then the coefficients  $c_i, d_i$  in (4.8) can be obtained, for instance, by truncating the Taylor

Table 1: Coefficients for a 19-stage and a 32-stage symmetric kernels of orders 10,  $P_{19}10$ , and 16,  $P_{32}16$ , respectively.

| $P_{19}10$                                       |   |
|--|---|
| $a_1 = 0.0432386502874358427757883618871$        | $b_1 = 0.0874171140239240929444597874709$     |
| $a_2 = 0.0891872116514875241139576575882$        | $b_2 = 0.0895405507537538756041132269850$     |
| $a_3 = 0.0874015611733434678704032626168$        | $b_3 = 0.0864066075260518454826592764125$     |
| $a_4 = 0.0954273508490522988798690279811$        | $b_4 = 0.140834736382004911175445238602$      |
| $a_5 = -0.0753249126916028783286798309378$       | $b_5 = -0.0137118117308991304396120981534$    |
| $a_6 = 0.202523451531452141504790651968$         | $b_6 = 0.541807462991626392685440183001$      |
| $a_7 = -0.000603437796174370985636258252420$     | $b_7 = -0.461545568134225404224525737926$     |
| $a_8 = 0.141029942275295351245992767342$         | $b_8 = 0.414574847635699390317333308406$      |
| $a_9 = 0.0000764516092828444432144561097509$     | $b_9 = -0.417468813318454485878866802863$     |
| $a_{10} = \frac{1}{2} - (a_1 + \dots + a_9)$     | $b_{10} = 1 - 2(b_1 + \dots + b_9)$           |
| $a_{21-i} = a_i, \quad i = 1, \dots, 10$         | $b_{20-i} = b_i, \quad i = 1, \dots, 9$       |
| $P_{32}16$                                       |   |
| $a_1 = 0$  | $b_1 = 0.0246666504515374580138379933112$     |
| $a_2 = 0.0503626559561541491851284108304$        | $b_2 = 0.0526269985834362938158150887511$     |
| $a_3 = 0.0546948611952386879984253468680$        | $b_3 = 0.0557559872576229997353176147790$     |
| $a_4 = 0.0554620390434566637065911933769$        | $b_4 = 0.0537116878888677727588921080438$     |
| $a_5 = 0.0516143924380795892137585965956$        | $b_5 = 0.0519896869988046163617507304275$     |
| $a_6 = 0.0568363649879098885339104529672$        | $b_6 = 0.0666959676117604242374885628805$     |
| $a_7 = 0.0939589227273508162683355424334$        | $b_7 = -0.102796651142514055780607785308$     |
| $a_8 = -0.00445692008047188584894138698734$      | $b_8 = 0.182323867085459132242253779621$      |
| $a_9 = 0.0817426743654653601759083129289$        | $b_9 = -0.00542617878109449520635361125714$   |
| $a_{10} = -0.0366714030328452540070009347543$    | $b_{10} = 0.0593919899010186971711928695894$  |
| $a_{11} = 0.0620267535945808302363559446459$     | $b_{11} = 0.0462313377171662707918171716453$  |
| $a_{12} = -0.0316075550822111219959097903622$    | $b_{12} = -0.0137171722415664093079656810822$ |
| $a_{13} = 0.0518562640986284507641256284631$     | $b_{13} = 0.582408428792399942617750550408$   |
| $a_{14} = -0.0000737830036206379685982463916033$ | $b_{14} = -0.562094520697629270991481101437$  |
| $a_{15} = 0.0536217552433463298408750165913$     | $b_{15} = -0.0180034629218910159228722367539$ |
| $a_{16} = 0.0150674488859324181502166600981$     | $b_{16} = 0.00990593102843635080330651455161$ |
| $a_{17} = \frac{1}{2} - (a_1 + \dots + a_{16})$  | $b_{17} = 1 - 2(b_1 + \dots + b_{16})$        |
| $a_{34-i} = a_{i+1}, \quad i = 1, \dots, 16$     | $b_{34-i} = b_i, \quad i = 1, \dots, 16$      |

expansion of the expressions (2.29). Note that by construction, the radius of convergence of the series  $P_1(x)$  and  $P_4(x)$  is  $r_* = r^* > x_l$ . Coefficients  $c_i, d_i$  for  $P_{38}2$  can be found in [3].

There are many other ways to approximate the processor which might be more convenient for a given problem. For instance, one may consider even polynomials  $\tilde{P}_j(x)$  of degree  $2(n+m)$  such that  $P_j(x) - \tilde{P}_j(x) = \mathcal{O}(x^{2n+2})$  as  $x \rightarrow 0$ , which minimize

$$\int_{-l\pi}^{l\pi} \left(1 - \left(\frac{x}{l\pi}\right)^2\right)^{-1/2} \left(\frac{\tilde{P}_j(x) - P_j(x)}{x^{2n+2}}\right)^2 dx.$$

There is still another procedure which may be suitable in case the output is frequently required. The post-processor can be virtually cost free if approximated using the intermediate stages obtained during the computation of the kernel (see [1] for more details).

## 5 Numerical examples

We have analyzed the relative performance of the stability polynomials when approximating the function  $\cos(x)$ . This study has allowed us to choose some representative stability polynomials among those showing the best performance and subsequently we have built processed splitting methods from them. It is then im-

portant to check if this relative performance still takes place in practical applications for the splitting methods obtained.

For the numerical tests carried out here we have selected the following representative  $k$ -stage processed methods of order  $q$ ,  $P_kq$ , built in this paper: (i) the high order processed methods  $P_{19}10$ ,  $P_{32}16$  and  $P_{38}20$  obtained using the stability polynomial  $p^{n,l}(x)$  in (4.2) with  $(n, l) = (5, 7), (8, 12), (10, 14)$ , respectively, where  $k = n + 2l$  and  $q = 2n$  and (ii) the second order processed schemes  $P_{19}2$ ,  $P_{32}2$  and  $P_{38}2$ , obtained using the stability polynomial  $p^{n,l,m}(x)$  in (4.4) with  $(n, l, m) = (1, 7, 4), (1, 12, 7), (1, 14, 9)$ , respectively, where  $k = n + 2l + m$ , and optimized by minimizing (4.5).

The following standard non-processed splitting methods from the literature are chosen for comparison:

- The 1-stage second order leapfrog method,  $\psi_{h,2}$ , given in (1.3) and denoted by  $LF_12$ .
- The well known 3-stage fourth-order time-reversible method (YS<sub>3</sub>4) [31], and the 17-stage eighth-order time-reversible method (M<sub>17</sub>8) [20, 23] (very similar performances are attained with the eighth-order method given in [11, 13]). Both methods are used with  $\psi_{h,2}$  as the basic scheme.
- The  $q$ -stage  $q$ th-order non-symmetric methods (GM <sub>$q$</sub>  $q$ ) with  $q = 4, 6, 8, 10, 12$  given in [9], and specifically designed for the harmonic oscillator.

We have also considered as a reference the standard 4-stage fourth order non-symplectic Runge–Kutta method,  $RK_44$ .

## 5.1 The harmonic oscillator

As a first example we consider again the one dimensional harmonic oscillator

$$\begin{Bmatrix} q' \\ p' \end{Bmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{Bmatrix} q \\ p \end{Bmatrix}.$$

This trivial example is well suited as a test bench for the following purposes: (i) to check that all coefficients of the kernel and postprocessor are correct with sufficient accuracy, (ii) to see if the relative performance shown by the stability polynomials is still valid for the processed splitting methods obtained from them, and (iii) to compare the performance of the new processed methods with other well established methods from the literature.

We take as initial conditions  $(q, p) = (1, 1)$  and integrate for  $t \in [0, 2000\pi]$  using different (constant) time steps. The larger time step corresponds to the stability limit (we repeated the experiment by increasing the time step until an overflow appeared). We measure the average error in the Euclidean norm of  $(q, p)$  versus the number of evaluations (for processed methods, this number corresponds to the kernel). Figure 2(a) shows the results obtained for the standard non-processed methods from the literature ( $LF_12$ , YS<sub>3</sub>4, GM<sub>4</sub>4, M<sub>17</sub>8, GM<sub>10</sub>10, GM<sub>12</sub>12). We clearly see that  $LF_12$  is the most stable and GM<sub>12</sub>12 is the most efficient if accurate results are desired (recall that this is a method designed for the harmonic oscillator), so that they are chosen to compare with the new processed schemes. Figure 2(b) shows the results obtained with the high order processed methods  $P_{19}10$ ,  $P_{32}16$  and  $P_{38}20$ , while Figure 2(c) repeats the experiments for the second order processed schemes  $P_{19}2$ ,  $P_{32}2$  and  $P_{38}2$ . Finally, Figure 2(d) illustrates the performance achieved by the most efficient methods in each case. It is clear the superiority of  $P_{38}2$  when accurate results are desired.

For the processor we have considered (2.28)-(2.29) where  $P_1, P_2$  are approximated using (4.8) and the coefficients  $c_i, d_i$ ,  $i = 1, \dots, s$  are obtained from the Taylor series expansion. The error introduced by the processor is of local character and does not propagate with time. For most problems it is enough to take for the pre-processor  $s = k$  for  $2k$ -stage methods. With respect to the post-processor, we can take either the same value of  $s$  or a smaller one (depending on the accuracy required at intermediate outputs or the length of the integration interval) because this error does not propagate. If the output is required frequently we can always approximate the post-processor using the intermediate stages obtained during the computation of the kernel [1].

Finally, it is important to notice the agreement between the relative performance shown by the processed methods given in Figure 2 and the results obtained for the stability polynomials in Figure 1.

Figure 2: Error versus the number of evaluations in logarithmic scale: (a) for the 2nd- to 12th-order non-processed methods, LF<sub>12</sub>, YS<sub>34</sub>, GM<sub>44</sub>, M<sub>178</sub>, GM<sub>1010</sub>, GM<sub>1212</sub>; (b) comparison of processed methods P<sub>1910</sub>, P<sub>3216</sub> and P<sub>3820</sub>, with LF<sub>12</sub> and GM<sub>1212</sub>; (c) the same for P<sub>192</sub>, P<sub>322</sub> and P<sub>382</sub>; (d) comparison of the most efficient processed methods with the most efficient non-processed ones.

## 5.2 The Schrödinger equation

As a second example we now consider the one-dimensional time-dependent Schrödinger equation (3.1) with the Morse potential  $V(x) = D(1 - e^{-\alpha x})^2$ . We fix the parameters to the following values:  $\mu = 1745 \text{ a.u.}$ ,  $D = 0.2251 \text{ a.u.}$  and  $\alpha = 1.1741 \text{ a.u.}$ , which are frequently used for modelling the HF molecule. As initial conditions we take the Gaussian wave function  $\psi(x, t) = \rho \exp(-\beta(x - \bar{x})^2)$ , with

$\beta = \sqrt{k\mu}/2$ ,  $\mu = 2D\alpha^2$ ,  $\bar{x} = -0.1$  and  $\rho$  is a normalizing constant. Assuming that the system is defined in the interval  $x \in [-0.8, 4.32]$ , we split it into  $d = 128$  parts of length  $\Delta x = 0.04$ , take periodic boundary conditions and integrate along the interval  $t \in [0, 20 \cdot 2\pi/w_0]$  with  $w_0 = \alpha\sqrt{2D/\mu}$  (see [3] for more details on the implementation of the splitting methods to this problem).

As we have seen, the splitting methods considered in this work preserve symplecticity but not unitarity. In Figure 3 we show the error in the preservation of unitarity,  $|q^T(t)q(t) + p^T(t)p(t) - 1|$ , and the relative error in energy  $(|E(t) - E(0)|/E(0))$ , where  $E(t) = u^T(t)Hu(t) = q^T(t)Hq(t) + p^T(t)Hp(t)$  for the 4-stage fourth order methods RK<sub>44</sub> and GM<sub>44</sub>. Both methods require the same number of FFTs calls per step (GM<sub>44</sub> has less storage requirements) and they are used with the same time step  $h = (2\pi/w_0)/250$ . The scheme GM<sub>44</sub> is orders of magnitude more accurate and, as expected for a symplectic integrator, the error in energy does not grow secularly in time. A similar behavior is observed for the error in unitarity, in agreement with the results shown in this work.

Figure 3: Error in the preservation of unitarity and energy as a function of time in double logarithmic scale for the non symplectic RK<sub>44</sub> and symplectic splitting GM<sub>44</sub> methods applied to the Schrödinger equation. Both 4-stage explicit fourth order methods are used with the same time step.

Figure 4 shows the error in the Euclidean norm of the vector solution at the end of the integration versus the number of FFT calls in double logarithmic scale. The integrations are done starting from a sufficiently small time step and repeating the computation by slightly increasing the time step until an overflow occurs, which we identify with the stability limit. We present the results for the 2nd- and 12th-order non-processed splitting methods chosen from the previous example (the results corresponding to the remaining methods also stay between them). For the processed methods we take the 19 and 38-stage methods both of order two (P<sub>19</sub>2 and P<sub>38</sub>2) and of order ten (P<sub>19</sub>10) and twenty (P<sub>38</sub>20), respectively. We observe that these methods have a relative performance similar to such obtained from the study of the harmonic oscillator or the corresponding stability polynomials. For reference, we have also included the results for the RK<sub>44</sub> method.



Figure 4: Error of the vector solution for the Schrödinger equation versus the number of FFT calls in logarithmic scale for the 2nd- and 12th-order non-processed methods, LF<sub>12</sub>, GM<sub>12</sub>12, the standard RK<sub>44</sub> and for the processed methods P<sub>19</sub>2, P<sub>19</sub>10, P<sub>38</sub>2 and P<sub>38</sub>20.

## Acknowledgements

This work has been partially supported by Ministerio de Educación y Ciencia (Spain) under project MTM2004-00535 (co-financed by the ERDF of the European Union) and Generalitat Valenciana (ACOMP06/239). SB has also been supported by a contract in the Programme Ramón y Cajal 2001. FC and SB would like especially to thank Arieh Iserles for his hospitality during their stay at the University of Cambridge in the summer of 2006.

## References

- [1] S. Blanes, F. Casas, and A. Murua, On the numerical integration of ordinary differential equations by processed methods, *SIAM J. Numer. Anal.*, **42** (2004), pp. 531–552.
- [2] S. Blanes, F. Casas, and A. Murua, Composition methods for differential equations with processing, *SIAM J. Sci. Comput.*, **27** (2006), pp. 1817–1843.
- [3] S. Blanes, F. Casas, and A. Murua, Symplectic splitting operator methods tailored for the time-dependent Schrödinger equation, *J. Chem. Phys.*, **124** (2006), 234105.
- [4] S. Blanes, F. Casas and J. Ros, Symplectic integrators with processing: a general study, *SIAM J. Sci. Comput.*, **21** (1999), 711–727.
- [5] S. Blanes, F. Casas, and J. Ros, Processing symplectic methods for near-integrable Hamiltonian systems, *Celest. Mech. & Dyn. Astr.*, **77** (2000), pp. 17–35.

- [6] S. Blanes, F. Casas, and J. Ros, High-order Runge–Kutta–Nyström geometric integrators with processing, *Appl. Numer. Math.*, **39** (2001), pp. 245–259.
- [7] J. Butcher, The effective order of Runge-Kutta methods. In: Conference on the numerical solution of differential equations, Lecture Notes in Math., Vol. **109**, Springer, Berlin (1969), pp. 133–139.
- [8] M.M. Chawla and S.R. Sharma, Intervals of periodicity and absolute stability of explicit Nyström methods for  $y'' = F(x, y)$ , *BIT*, **21** (1981), pp. 455–464.
- [9] S. Gray and D.E. Manolopoulos, Symplectic integrators tailored to the time-dependent Schrödinger equation, *J. Chem. Phys.*, **104** (1996), pp. 7099–7112.
- [10] S. Gray and J.M. Verosky, Classical Hamiltonian structures in wave packet dynamics, *J. Chem. Phys.*, **100** (1994), pp. 5011–5022.
- [11] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Ser. Comput. Math. 31, Springer-Verlag, Berlin 2002.
- [12] T. Hirono, W. Lui, and S. Seki, A tree-dimensional fourth-order finite difference time-domain scheme using a symplectic integrator propagator, *IEEE Trans. Microwave Theory Tech.*, **49** (2001), pp. 1640–1648.
- [13] W. Kahan and R.C. Li, Composition constants for raising the order of unconventional schemes for ordinary differential equations, *Math. Comp.*, **66** (1997), 1089–1099.
- [14] R. Kosloff, Time-dependent quantum-mechanical methods for molecular dynamics, *J. Phys. Chem.*, **92** (1988), pp. 2087–2100.
- [15] C. Leforestier, R.H. Bisseling, C. Cerjam., M.D. Feit, R. Friesner, A. Guldberg, A. Hammerich, G. Jolicard, W. Karrlein, H.-D. Meyer, N. Lipkin, O. Roncero, and R. Kosloff, A comparison of different propagation schemes for the time dependent Schroedinger equation, *J. Comput. Phys.*, **94** (1991), pp. 59–80.
- [16] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge University Press, Cambridge 2004.
- [17] X. Liu, P. Ding, J. Hong, and L. Wang, Optimization of symplectic schemes for time-dependent Schrödinger equations, *Comput. Math. Appl.*, **50** (2005), pp. 637–644.
- [18] M.A. López-Marcos, J.M. Sanz-Serna, and R.D. Skeel, An explicit symplectic integrators with maximal stability interval, in: D.F. Griffiths and G.A. Watson, eds., Numerical Analysis, A.R. Mitchel 75th Birthday Volume (World Scietific, Singapore, 1996), pp. 163–176.
- [19] M.A. López-Marcos, J.M. Sanz-Serna, and R.D. Skeel, Explicit symplectic integrators using Hessian-vector products, *SIAM J. Sci. Comput.*, **18** (1997), pp. 223–238.
- [20] R.I. McLachlan, On the numerical integration of ordinary differential equations by symmetric composition methods, *SIAM J. Sci. Comput.*, **16** (1995), pp. 151–168.
- [21] R.I. McLachlan, More on symplectic correctors, in *Integration Algorithms and Classical Mechanics*, Vol. **10**, J.E. Marsden, G.W. Patrick, and W.F. Shadwick, eds., American Mathematical Society, Providence, R.I. (1996), pp. 141–149.

- [22] R.I. McLachlan and S.K. Gray, Optimal stability polynomials for splitting methods, with applications to the time-dependent Schrödinger equation, *Appl. Numer. Math.*, **25** (1997), pp. 275–286.
- [23] R.I. McLachlan and R.G.W. Quispel, Splitting methods, *Acta Numerica*, **11** (2002), pp. 341–434.
- [24] R.I. McLachlan and R.G.W. Quispel, Geometric integrators for ODEs, *J. Phys. A: Math. Gen.*, **39** (2006), pp. 5251–5285.
- [25] R. Rieben, D. White, and G. Rodrigue, High-order symplectic integration methods for finite element solutions to time dependent Maxwell equations, *IEEE Trans. Antennas Propagat.*, **52** (2004), pp. 2190–2195.
- [26] G. Rowlands, A numerical algorithm for Hamiltonian systems, *J. Comput. Phys.*, **97** (1991), pp. 235–239.
- [27] J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian Problems*, Chapman & Hall, London, 1994.
- [28] J.M. Sanz-Serna and A. Portillo, Classical numerical integrators for wave-packet dynamics, *J. Chem. Phys.*, **104** (1996), pp. 2349–2355.
- [29] J. Wisdom, M. Holman, and J. Touma, Symplectic correctors. In: *Integration Algorithms and Classical Mechanics*, Vol. **10**, J.E. Marsden, G.W. Patrick, and W.F. Shadwick, eds., American Mathematical Society, Providence, R.I. (1996), pp. 217–244.
- [30] K. Yee, Numerical solution of initial boundary value problems involving Maxwell’s equations in isotropic media, *IEEE Trans. Antennas Propagat.*, **14** (1966), pp. 302–307.
- [31] H. Yoshida, Construction of higher order symplectic integrators, *Phys. Lett. A*, **150** (1990), pp. 262–268.
- [32] W. Zhu, X. Zhao, and Y. Tang, Numerical methods with high order of accuracy applied in the quantum system, *J. Chem. Phys.*, **104** (1996), pp. 2275–2286.