

On the existence of positive coefficients for operator splitting schemes of order higher than two^{*}

Sergio Blanes and sblanes@mat.uji.es Fernando Casas
casas@mat.uji.es

Departament de Matemàtiques, Universitat Jaume I, 12071-Castellón, Spain

Abstract

In this paper we consider numerical integration methods applied to differential equations which are separable in solvable parts. These methods are compositions of flows associated with each part of the system. We propose an elementary proof of the necessary existence of negative coefficients if the schemes are of order, or effective order, $p \geq 3$ and provide additional information about the distribution of these negative coefficients. It is shown that if the methods involve flows associated with more general terms this result does not necessary apply and in some cases it is possible to build higher order schemes with positive coefficients.

Key words: Splitting methods, Composition methods, Effective order, Numerical integrators

1991 MSC: 65L05, 65L70

1 Introduction

Operator splitting schemes are numerical methods which are particularly useful to approximate the evolution of differential equations when they are separable in solvable parts [8]. Specifically, let us consider the system

$$\dot{x} = A(x) + B(x), \tag{1}$$

^{*} This work has been partially supported by Fundació Caixa Castelló–Bancaixa. SB has also been supported by the Ministerio de Ciencia y Tecnología (Spain) through a contract in the Programe Ramón y Cajal 2001.

where the dot indicates time derivative, $x \in \mathbb{R}^d$ and A, B are non commuting vector fields on \mathbb{R}^d whose flows are exactly computable or, equivalently, the equations $\dot{x} = A(x)$ and $\dot{x} = B(x)$ are solvable. Then, by writing the flow (i.e., the exact solution) of (1) as $x(t) = \exp(t(A + B))(x(0))$, the composition

$$\chi(h) = \exp(hA) \exp(hB) \quad (2)$$

approximates $x(h)$ with error of order h^2 , whereas

$$S(h) = \exp\left(\frac{h}{2}A\right) \exp(hB) \exp\left(\frac{h}{2}A\right) \quad (3)$$

is such that $x(h) - S(h) = \mathcal{O}(h^3)$. The order of the composition can be increased by including more exponentials in a time step h . In general, the composition method

$$\psi_h = e^{ha_1A} e^{hb_1B} \dots e^{ha_mA} e^{hb_mB}, \quad (4)$$

has order p if $\psi_h = \exp(h(A + B)) + \mathcal{O}(h^{p+1})$ for a proper choice of m and coefficients a_i, b_i . It can be assumed without loss of generality that in (4) none of the coefficients b_1, b_2, \dots, b_{m-1} as well as none of a_2, a_3, \dots, a_m are vanishing, i.e., only a_1 and/or b_m can be zero, since otherwise the corresponding exponentials could be removed and the rest of the exponentials would be concatenated.

High order numerical schemes based on the composition (4) have been successfully used in the literature for a large number of problems [5,8]. It has been noticed that some of the coefficients in (4) are negative for $p \geq 3$. In other words, the methods always involve stepping backwards in time. This constitutes a problem when equation (1) is defined in a semigroup, as arises, for instance, in dimensional splitting of diffusion equations, since then the method can only be conditionally stable [8].

This feature of the composition scheme (4) is in fact unavoidable, and can be established as the following two theorems:

Theorem 1 [10,11]. *If p is a positive integer such that $p \geq 3$, then there are no composition methods of the form (4) and finite m with all the coefficients a_i, b_i being positive.*

Theorem 2 [4]. *If p is a positive integer such that $p \geq 3$, then, for every p th-order method (4) with m any finite positive integer,*

$$\min_{1 \leq i \leq m} a_i < 0 \quad \text{and} \quad \min_{1 \leq j \leq m} b_j < 0.$$

Theorem 2 is stronger than Theorem 1 and their proofs are based on the (non-trivial) observation that a scheme of the form (4) with m any finite positive

integer and all the coefficients a_i, b_i being positive, cannot satisfy the order conditions up to $p = 3$, which can be written as

$$\begin{aligned}
\text{order 1} \quad & \sum_{i=1}^m a_i = 1; & \sum_{i=1}^m b_i = 1 \\
\text{order 2} \quad & \sum_{i=1}^{m-1} b_i \left(\sum_{j=1}^i a_j \right) = \frac{1}{2} \\
\text{order 3} \quad & \sum_{i=1}^{m-1} b_i \left(\sum_{j=i+1}^i a_j \right)^2 = \frac{1}{3}; & \sum_{i=1}^m a_i \left(\sum_{j=i}^m b_j \right)^2 = \frac{1}{3}.
\end{aligned}$$

One of the goals of this paper is to provide an alternative, elementary proof of Theorem 2, giving in addition more insight into the distribution of the coefficients in the composition.

On the other hand, during the last few years the processing technique has been used to find better composition methods requiring less evaluations than conventional schemes of order p [6]. The idea is to consider a method of the form

$$\hat{\psi}_h \equiv \pi_h \psi_h \pi_h^{-1}. \quad (5)$$

Then, to evaluate n time steps, it is clear that $\hat{\psi}_h^n = \pi_h \psi_h^n \pi_h^{-1}$ and thus only the cost of ψ_h is relevant. The composition (4) is said to be a method of *effective order* p if there exists a map π_h such that $\hat{\psi}_h = \exp(h(A+B)) + \mathcal{O}(h^{p+1})$, i.e., if it is conjugate to an integrator $\hat{\psi}_h$ of order p , which is called a processing method. The maps ψ_h and π_h are usually referred as the kernel and the post-processor of the method, respectively. If the kernel is given by the composition (4) then the number of conditions to be fulfilled by the coefficients a_i, b_i is considerably reduced (see [2] and references therein), whereas the post-processor is taken as a flow close to the identity, $\pi_h = I + \mathcal{O}(h)$, and can be chosen also as a composition (4) or, more generally, as an element of the universal enveloping algebra of the free Lie algebra generated by A and B . Obviously, a method of order p has also effective order p (taking $\pi_h = I$) or higher, but the converse is not true in general.

The simplest example of a processing method is the following:

$$S(h) = e^{\frac{h}{2}A} e^{hB} e^{\frac{h}{2}A} = e^{\frac{h}{2}A} e^{-hA} e^{hA} e^{hB} e^{\frac{h}{2}A} = e^{-\frac{h}{2}A} \chi(h) e^{\frac{h}{2}A}$$

showing that the first order method (2) has, in fact, effective order 2 because it is conjugate to the second order leapfrog/Strang splitting method $S(h)$.

In this paper we also address the following question: do Theorems 1 and 2 also hold for a composition ψ_h of effective order $p \geq 3$? Observe that, in principle, Theorem 2 applies to the whole composition $\hat{\psi}_h$, but if the negative coefficients are contained only in π_h (or π_h^{-1}) then in practice the processed method only involves stepping forward in time (recall that π_h^{-1} is computed at the initial

step and π_h only for output). Thus the answer to this problem could be useful in the search of effective methods of order higher than 2 for systems that evolve in a semigroup, such as the heat equation [8].

2 An elementary proof

Next we propose an elementary proof of the necessary existence of negative coefficients in the composition (4) for $p \geq 3$. We consider the first order method $\chi(h) = e^{hA} e^{hB}$ and its adjoint $\chi^*(h) = e^{hB} e^{hA}$. The Baker–Campbell–Hausdorff (BCH) theorem establishes that $\chi(h)$ is itself the flow of a vector field X which lies in the same Lie algebra as A and B :

$$\chi(h) = \exp(X) = \exp(hX_1 + h^2X_2 + h^3X_3 + \dots), \quad (6)$$

with $X_1 = A+B$, $X_2 = \frac{1}{2}[A, B]$, $X_3 = \frac{1}{12}[A-B, [A, B]]$, etc. Here $[A, B]$ stands for the commutator $AB - BA$. In general, X_k , $k > 1$, is a linear combination of nested commutators involving k operators A and B .

With respect to the adjoint method, the corresponding flow can be expressed as

$$\chi^*(h) = \exp(hX_1 - h^2X_2 + h^3X_3 - \dots). \quad (7)$$

It is easy to check that if the composition (4) satisfies the condition $\sum_{i=1}^m a_i = \sum_{i=1}^m b_i$ then it can be expressed as [7,5,8]

$$\begin{aligned} \psi_h &= e^{ha_1A} e^{hb_1B} \dots e^{ha_mA} e^{hb_mB} \\ &\quad \uparrow \nearrow \uparrow \nearrow \nearrow \uparrow \nearrow \\ &= \chi^*(\beta_0h) \chi(\beta_1h) \chi^*(\beta_2h) \dots \chi(\beta_{2m-1}h) \chi^*(\beta_{2m}h) \end{aligned} \quad (8)$$

with $\beta_0 = \beta_{2m} = 0$ and $\sum_{i=1}^m (a_i + b_i) = 2 \sum_{i=0}^{2m} \beta_i$. The consistency conditions are $\sum_{i=1}^m a_i = \sum_{i=1}^m b_i = \sum_{i=0}^{2m} \beta_i = 1$, and the order conditions for the coefficients a_i, b_i to get a method of order $p > 1$ are equivalent to the order conditions for the β_i [7]. From (8) the following relation between the coefficients is obtained:

$$a_i = \beta_{2i-1} + \beta_{2i-2}, \quad b_i = \beta_{2i} + \beta_{2i-1}, \quad i = 1, \dots, m. \quad (9)$$

Applying the BCH formula repeatedly to (8) and taking into account (6) and (7) gives

$$\psi_h = \exp \left(hf_{1,1}X_1 + h^2f_{2,1}X_2 + h^3(f_{3,1}X_3 + f_{3,2}[X_1, X_2]) + \mathcal{O}(h^4) \right),$$

where the coefficients $f_{k,j}$ are homogeneous polynomials of degree k in the variables β_i . In particular we have

$$\begin{aligned}
f_{1,1} &= \sum_{i=0}^{2m} \beta_i, & f_{2,1} &= \sum_{i=0}^{2m} (-1)^{i+1} \beta_i^2 \\
f_{3,1} &= \sum_{i=0}^{2m} \beta_i^3, & f_{3,2} &= \sum_{i=0}^{2m} (-1)^{i+1} \beta_i^2 \left(\sum_{j=i+1}^{2m} \beta_j - \sum_{j=0}^{i-1} \beta_j \right).
\end{aligned} \tag{10}$$

Conditions $f_{1,1} = 1$ and $f_{n,j} = 0$ for all $n \leq p$ are then sufficient for the method to have order p . Thus, equation

$$f_{3,1} = \sum_{i=0}^{2m} \beta_i^3 = 0 \tag{11}$$

is a necessary condition to be satisfied by any method of order $p \geq 3$. We suppose that more than two β_i are different from zero because $\beta_1^3 + \beta_2^3 = 0$ together with the consistency condition $\beta_1 + \beta_2 = 1$ have no real solution. Now (11) can be written as

$$\sum_{i=1}^m (\beta_{2i-1}^3 + \beta_{2i-2}^3) + \beta_{2m}^3 = \sum_{i=1}^m (\beta_{2i-1}^3 + \beta_{2i-2}^3) = 0$$

for any positive integer m . In consequence $\beta_{2j-1}^3 + \beta_{2j-2}^3$ has to be negative for some $1 \leq j \leq m$. But it is easy to verify that $\text{sign}(x^3 + y^3) = \text{sign}(x + y)$ for any $x, y \in \mathbb{R}$, so that

$$0 > \beta_{2j-1} + \beta_{2j-2} = a_j \tag{12}$$

for some j such that $1 \leq j \leq m$. Similarly, we can write (11) as

$$\beta_0^3 + \sum_{i=1}^m (\beta_{2i}^3 + \beta_{2i-1}^3) = \sum_{i=1}^m (\beta_{2i}^3 + \beta_{2i-1}^3) = 0$$

so that $\beta_{2k}^3 + \beta_{2k-1}^3 < 0$ for some $1 \leq k \leq m$, and again

$$0 > \beta_{2k} + \beta_{2k-1} = b_k. \tag{13}$$

In this way the proof is complete.

Distribution of the coefficients

We can get more information about the distribution of the negative coefficients in the composition (4) by applying a slightly more involved argument which, in fact, also provides another demonstration of Theorem 2.

If we denote $\alpha_{2i-1} = a_i$, $\alpha_{2i} = b_i$, $i = 1, \dots, m$, in the composition (4),

$$\psi_h = e^{h\alpha_1 A} e^{h\alpha_2 B} \dots e^{h\alpha_{2m-1} A} e^{h\alpha_{2m} B}$$

then it is a simple exercise to check that

$$\alpha_i = \beta_i + \beta_{i-1}, \quad i = 1, \dots, 2m \quad (14)$$

where β_j are the coefficients appearing in (8) ($\beta_0 = \beta_{2m} = 0$). Now all we need to prove Theorem 2 is to analyse how equations (11) and (14) imply that at least one odd as well as one even α_i coefficients are negative.

As before, we assume that there are more than two non-vanishing coefficients β_i and at least one of them negative.

- (a) Let us suppose first that only one coefficient is actually negative, say β_j , for some $0 < j < 2m$. Then, from eq. (11),

$$\beta_j = -\left(\sum_{i \neq j} \beta_i^3\right)^{1/3}$$

so that $|\beta_j| > \beta_i$ for all $i \neq j$. Therefore

$$\alpha_j = \beta_j + \beta_{j-1} < 0 \quad \text{and} \quad \alpha_{j+1} = \beta_{j+1} + \beta_j < 0$$

i.e., two consecutive α_k coefficients are negative, and thus at least one a_j and one b_j are negative.

- (b) Suppose now that the negative coefficients are $\beta_{j_1}, \beta_{j_2}, \dots, \beta_{j_k}$ with $j_1 < j_2 < \dots < j_k$.

- (b.1) If

$$\beta_{j_i-1} < |\beta_{j_i}| > \beta_{j_i+1} \quad (15)$$

for some $j_i \in \mathcal{I} \equiv \{j_1, j_2, \dots, j_k\}$ then also two consecutive coefficients α_k are negative, namely α_{j_i} and α_{j_i+1} .

- (b.2) On the other hand, when (15) does not hold for any $j_i \in \mathcal{I}$, then the following situations are possible:

- (i) if $\beta_{j_i-1} < |\beta_{j_i}|$, then $|\beta_{j_i}| < \beta_{j_i+1}$;
- (ii) if $|\beta_{j_i}| > \beta_{j_i+1}$, then $\beta_{j_i-1} > |\beta_{j_i}|$;
- (iii) finally, $\beta_{j_i-1} > |\beta_{j_i}| < \beta_{j_i+1}$.

Let us suppose that $\beta_{j_i+1} \neq \beta_{j_i+1-1}$ for all j_i . Then

$$\begin{aligned} \beta_{j_k} = & -\left((\beta_{j_1-1}^3 + \beta_{j_1}^3 + \beta_{j_1+1}^3) + (\beta_{j_2-1}^3 + \beta_{j_2}^3 + \beta_{j_2+1}^3) + \dots \right. \\ & \left. + (\beta_{j_{k-1}-1}^3 + \beta_{j_{k-1}}^3 + \beta_{j_{k-1}+1}^3) + \sum' \beta_i^3\right)^{1/3} \end{aligned}$$

where \sum' contains the remaining terms (including $\beta_{j_{k-1}}$ and $\beta_{j_{k+1}}$). Since $\beta_{j_l-1}^3 + \beta_{j_l}^3 + \beta_{j_l+1}^3 > 0$ for $l = 1, \dots, k-1$, then clearly

$$|\beta_{j_k}| > \beta_{j_{k-1}} \quad \text{and} \quad |\beta_{j_k}| > \beta_{j_{k+1}}$$

in contradiction with hypothesis (b.2). Therefore $\beta_{j_i+1} = \beta_{j_i+1-1}$ for some j_i . Let us suppose, without loss of generality, that they correspond to the

first $l + 1$ coefficients β_{j_i} . Then, the only possible sequence (different from those considered before) has to be as follows:

$$\beta_{j_i-1} < |\beta_{j_i}| < \beta_{j_i+1} > |\beta_{j_i+2}| < \beta_{j_i+3} > \cdots < \beta_{j_i+2l-1} > |\beta_{j_i+2l}| > \beta_{j_i+2l+1}$$

with $\beta_{j_i+k} \equiv \beta_{j_i+2k}$, $k = 0, 1, \dots, l$ are the negative coefficients. Then

$$\alpha_{j_i} = \beta_{j_i} + \beta_{j_i-1} < 0, \quad \alpha_{j_i+2l+1} = \beta_{j_i+2l+1} + \beta_{j_i+2l} < 0.$$

Also in this case at least one a_i and one b_i are negative because j_i and $j_i + 2l + 1$ differ in an odd number.

Notice that this is the only situation where two negative α_i coefficients in a given method do not stay in consecutive places. We have checked several composition methods published in the literature having observed that this occurrence is in fact quite rare: it is very much frequent that at least two consecutive α_i coefficients are negative, and this discussion provides an explanation of the phenomenon.

3 Compositions of effective order $p \geq 3$

To prove the non-existence of composition methods of the form (4) of effective order $p = 3$ with all the coefficients being positive, let us consider the most general post-processor π_h in the universal enveloping Lie algebra of the free Lie algebra generated by A and B up to order two,

$$\pi_h = \exp\left(h(c_1A + c_2B) + h^2c_3X_2\right) + \mathcal{O}(h^3), \quad (16)$$

with c_i free parameters. Since $c_1A + c_2B = (c_2 - c_1)B + c_1X_1 = (c_1 - c_2)A + c_2X_1$ we can rewrite (16) as follows

$$\pi_h = e^{hc_1X_1} e^{h^2d_1X_2} e^{hcB} + \mathcal{O}(h^3) \quad (17)$$

$$= e^{hc_2X_1} e^{h^2d_2X_2} e^{-hcA} + \mathcal{O}(h^3) \quad (18)$$

where $c = c_2 - c_1$ and d_1, d_2 are parameters depending on c_1, c_2, c_3 . Since e^{X_1} commutes with $\hat{\psi}_h$ up to order p in (5) then $\exp(hc_iX_1)$ in (17) and (18) can safely be removed without loss of generality, and thus we take $c_1 = 0$ or $c_2 = 0$.

If, in particular, we substitute (17) in (5) we find

$$\begin{aligned} \hat{\psi}_h = \exp\left(hX_1 + h^2(f_{2,1} - 2c)X_2 + h^3\left((f_{3,1} - 3c(f_{2,1} - c)X_3 \right. \right. \\ \left. \left. + (f_{3,2} + \frac{1}{2}c(f_{2,1} - c) - d_1))[X_1, X_2]\right)\right) + \mathcal{O}(h^4). \end{aligned} \quad (19)$$

A second order method is obtained if $c = \frac{1}{2}f_{2,1}$. If we substitute this value in (19) and take d_1 such that the coefficient of $[X_1, X_2]$ vanishes, then

$$\hat{\psi}_h = \exp\left(hX_1 + h^3\left(f_{3,1} - \frac{3}{4}f_{2,1}^2\right)X_3\right) + \mathcal{O}(h^4), \quad (20)$$

and exactly the same result is obtained if considering (18) instead of (17). As a result, we find that

$$f_{3,1} - \frac{3}{4}f_{2,1}^2 = 0 \quad (21)$$

is the only condition to be satisfied for a composition to be of effective order three. This condition is equivalent to the kernel condition at order three presented in [1] using a different basis of the Lie algebra.

Theorem 3 *At least one of the a_i as well as one of the b_i have to be negative in the composition (4) if ψ_h is of effective order $p \geq 3$.*

Proof. Suppose that ψ_h is a composition of effective order three with, for example, all a_i positive. This means that it exists a map π_h which can be written as in (17) (with $c_1 = 0$ without loss of generality) such that

$$e^{h^2d_1X_2} e^{hcB} \psi_h e^{-hcB} e^{-h^2d_1X_2} = e^{h(A+B)} + \mathcal{O}(h^4), \quad (22)$$

or equivalently

$$\begin{aligned} \bar{\psi}_h &\equiv e^{hcB} \psi_h e^{-hcB} = e^{-h^2d_1X_2} e^{h(A+B)} e^{h^2d_1X_2} + \mathcal{O}(h^4) \\ &= e^{h(A+B)+h^3d_2[X_1, X_2]} + \mathcal{O}(h^4). \end{aligned} \quad (23)$$

Notice that $\bar{\psi}_h$ have all coefficients a_i positive. On the other hand, $\bar{\psi}_h$ is a method which can be written as a composition of a first order and its adjoint

$$\bar{\psi}_h = \chi^*(\bar{\beta}_0h)\chi(\bar{\beta}_1h)\chi^*(\bar{\beta}_2h) \cdots \chi(\bar{\beta}_{2k-1}h)\chi^*(\bar{\beta}_{2k}h) \quad (24)$$

and from (23) we find that

$$\bar{f}_{3,1} := \sum_{i=0}^{2k} \bar{\beta}_i^3 = 0,$$

with $\bar{\beta}_0 = \bar{\beta}_{2k} = 0$. But, as we know, this condition can not be satisfied with all coefficients a_i positive.

Similarly, if we suppose that all b_i are positive then we can repeat the same procedure taking the processor (18), arriving at the same contradiction. ■

Example. Let us consider the following composition

$$\begin{aligned}\psi_h &= e^{haA} e^{hbB} e^{h(1-a)A} e^{h(1-b)B} \\ &= \chi^*(\beta_0 h) \chi(\beta_1 h) \chi^*(\beta_2 h) \chi(\beta_3 h) \chi^*(\beta_4 h),\end{aligned}\tag{25}$$

with $\beta_0 = 0$, $\beta_1 = a$, $\beta_2 = b - a$, $\beta_3 = 1 - b$, $\beta_4 = 0$, and the consistency conditions already imposed. This composition can not be of order three (there are not enough parameters to solve the order conditions), but it can be of effective order three if condition (21) is satisfied, or equivalently if

$$1 - 12ab(1 - a)(1 - b) = 0,$$

which has no real solution if $a \in (0, 1)$ as well as if $b \in (0, 1)$. In consequence, at least one of the coefficients of A as well as one of the coefficients of B have to be negative.

4 Other classes of composition methods

The previous results can be generalized in several important contexts. For instance, let us consider the method

$$\psi_h = \sum_{k=1}^K \gamma_k \psi_{h,k}, \quad \text{with} \quad \psi_{h,k} = \prod_{i=1}^n e^{ha_{k,i}A} e^{hb_{k,i}B},\tag{26}$$

where it is assumed that $\sum_k \gamma_k = 1$ and $\sum_i a_{k,i} = \sum_i b_{k,i}$, $k = 1, \dots, K$.

Theorem 4 [4]. *If p and K are positive integers such that $p \geq 3$ and $K \geq 2$, and $\gamma_k > 0$ for $k = 1, \dots, K$, then at least one of the coefficients $a_{k,i}$ as well as one of the $b_{k,i}$ have to be negative in the composition (26) if ψ_h has order, or effective order, $p \geq 3$.*

Proof. Since $\sum_i a_{k,i} = \sum_i b_{k,i}$ we can write

$$\psi_{h,k} = \prod_{i=1}^n \chi(\beta_{k,i} h) \chi^*(\beta_{k,i} h),\tag{27}$$

and following the same procedure as previously we find that instead of (10) the necessary condition to be a method of order three or higher is now

$$\sum_{k=1}^K \gamma_k \sum_{i=1}^{2n} \beta_{k,i}^3 = 0.\tag{28}$$

Since $\gamma_k > 0$, $k = 1, \dots, K$, then there exists $1 \leq j \leq K$ such that

$$\sum_{i=1}^{2n} \beta_{j,i}^3 \leq 0,\tag{29}$$

and the previous proof can be applied. ■

Consider now A_1, \dots, A_k non commutative operators and the composition

$$\psi_h = \prod_{i=1}^m e^{ha_{i,1}A_1} \dots e^{ha_{i,k}A_k}, \quad (30)$$

with $\psi_h = \exp(h(A_1 + \dots + A_k)) + \mathcal{O}(h^{p+1})$. Then we have

Theorem 5 *If the composition (30) has order, or effective order, $p \geq 3$, then*

$$\min_{1 \leq i \leq m} a_{i,l} < 0 \quad l = 1, \dots, k.$$

Proof. If there exists $1 \leq j \leq k$ such that all coefficients of A_j are positive we could take $A_j = A$ and $A_i = B/(k-1)$, $i \neq j$, recovering the composition (4) so, at least one of the coefficients of A_j has to be negative. ■

Composition of a symmetric method

We consider now compositions of the symmetric (or self-adjoint) second order scheme $S(h)$ given by (3) (or with the roles of A and B interchanged), i.e.,

$$\psi_h = S(\beta_0 h) S(\beta_1 h) \dots S(\beta_{m-1} h) S(\beta_m h). \quad (31)$$

The basic scheme $S(h)$ can formally be written as the flow of a vector field in the form [5]

$$S(h) = \exp(hX_1 + h^3X_3 + h^5X_5 + \dots). \quad (32)$$

Analogously to the composition (4), repeated application of the BCH formula leads now to

$$\psi_h = \exp(hf_{1,1}X_1 + h^3f_{3,1}X_3 + \mathcal{O}(h^4)),$$

where

$$f_{1,1} = \sum_{i=0}^m \beta_i, \quad f_{3,1} = \sum_{i=0}^m \beta_i^3.$$

Therefore $f_{1,1} = 1$, $f_{3,1} = 0$ are necessary conditions for ψ_h to have order $p \geq 3$. In fact, since $f_{2,1} = 0$ in this case, they are also the conditions to be satisfied by ψ_h to have effective order $p = 3$ and the following result is immediate.

Theorem 6 *If p is any positive integer such that $p \geq 3$ and $S(h)$ is given by (3), then at least two consecutive coefficients a_i, b_i have to be negative in the composition (31) if ψ_h has order, or effective order, p .*

Proof. By substituting in (31) the expression of the basic method $S(\beta_i h) = \exp(\frac{\beta_i}{2} h A) \exp(\beta_i h B) \exp(\frac{\beta_i}{2} h A)$, we obtain a composition of the type (4) with

$$b_i = \beta_i, \quad a_i = \frac{1}{2}(\beta_i + \beta_{i-1}) \quad i = 1, \dots, m$$

if $\beta_0 = \beta_m = 0$. It is immediate to check that if there exists one negative coefficient, say $\beta_j < 0$, $1 \leq j \leq m - 1$ and

$$|\beta_j| > \beta_{j-1} \quad \text{then} \quad a_j < 0, \quad b_j < 0, \quad (33)$$

whereas if

$$|\beta_j| > \beta_{j+1} \quad \text{then} \quad b_j < 0, \quad a_{j+1} < 0. \quad (34)$$

In other words, as soon as one of the β_i is negative and its absolute value is higher than the previous one or the next one then the corresponding composition (4) has, at least, two consecutive coefficients which are negative.

Suppose now that there are k negative coefficients, $\beta_{j_1}, \beta_{j_2}, \dots, \beta_{j_k} < 0$ such that they do not satisfy conditions (33) and (34), and they are not consecutive (otherwise the theorem is obviously satisfied). Then we can write condition $f_{3,1} = 0$ as

$$\beta_{j_k} = - \left((\beta_{j_1-1}^3 + \beta_{j_1}^3) + \dots + (\beta_{j_{k-1}-1}^3 + \beta_{j_{k-1}}^3) + \sum' \beta_i^3 \right)^{1/3},$$

where \sum' contains the remaining terms, including β_{j_k-1} and β_{j_k+1} . Since $\beta_{j_i-1}^3 + \beta_{j_i}^3 > 0$, $i = 1, \dots, k - 1$, then $\beta_{j_k-1} < |\beta_{j_k}| > \beta_{j_k+1}$, and conditions (33) and (34) are satisfied. ■

This result, together with the discussion of section 2 justifies why it is so frequent that at least two consecutive coefficients are negative.

5 Composition methods with positive coefficients

Let us consider now the second order differential equation

$$\ddot{y} = f(y), \quad (35)$$

which can be written in the form (1) by taking $x \equiv (x_1, x_2) = (y, \dot{y})$ and $A(x) = (x_2, 0)$, $B(x) = (0, f(x_1))$, or equivalently

$$A \equiv x_2 \frac{\partial}{\partial x_1}, \quad B \equiv f(x_1) \frac{\partial}{\partial x_2}.$$

This equation frequently appears in relevant problems arising in classical and quantum mechanics, where the operator A is related to the kinetic energy

(quadratic in momenta) and B is associated with the potential energy. Then the commutator $C \equiv [B, [A, B]]$ is explicitly and exactly solvable and, in addition, $[B, C] = 0$ so that B and C can be incorporated in the same exponential. It makes sense, then, to consider the composition

$$\psi_h = \prod_{i=1}^m \exp(ha_i A) \exp(hb_i B + h^3 c_i C) \quad (36)$$

instead of (4). In this case ψ_h can not always be written as the composition of a first order method and its adjoint, and Theorem 2 does not apply. For instance [3]

$$\psi_h = \exp\left(\frac{h}{6}B\right) \exp\left(\frac{h}{2}A\right) \exp\left(\frac{2h}{3}B + \frac{h^3}{72}C\right) \exp\left(\frac{h}{2}A\right) \exp\left(\frac{h}{6}B\right) \quad (37)$$

is a method of order four and [9]

$$\psi_h = \exp\left(\frac{h}{2}A\right) \exp\left(hB + \frac{h^3}{24}C\right) \exp\left(\frac{h}{2}A\right) \quad (38)$$

is a method of effective order four. In the last case we can write $\psi_h = \chi(h/2)\chi^*(h/2)$ with $\chi(h) = \exp(hA) \exp(hB + h^3/6C)$. However, if we analyse the elements of the exponential of $\chi(h)$ in (6) we find that $X_3 = [X_1, X_2]/6$. Then X_3 is not an independent element and can be cancelled with a proper choice of the map π_h , giving a fourth-order method.

Numerical experiments suggest that this is the highest order one can get with the composition (36) with positive coefficients and a rigorous proof is at present under investigation. However, methods of effective order six as well as of order six are known to exist with all coefficients b_i positive.

On the other hand, if we consider a Hamiltonian system of the form

$$H = T(p) + V(q)$$

with T quadratic in p and $V(q)$ a polynomial function up to degree four in \mathbf{q} (or, in general, if $f(y)$ is a polynomial function up to degree three), then $E \equiv [A, [A, [A, [A, B]]]]$ vanish or depends only on the momenta, i.e. $[A, E] = 0$, and can be computed with the kinetic energy, similarly to C . In addition $D \equiv [B, [B, [A, [A, B]]]]$ depends only on the coordinates and thus $[B, D] = 0$. It turns out that the generalised leapfrog splitting scheme

$$\psi_h = \exp\left(\frac{h}{2}A + h^5 e E\right) \exp\left(hB + h^3 c C + h^5 d D\right) \exp\left(\frac{h}{2}A + h^5 e E\right) \quad (39)$$

with $c = \frac{1}{24}$, $d = \frac{1}{1440}$, $e = \frac{1}{2880}$ is a method of effective order six, since these coefficients satisfy the kernel conditions collected in [1] up to this order.

We should recall that methods (37), (38) and (39) are particular examples of composition schemes involving only positive coefficients. The possible existence of other families of composition methods of order $p \geq 3$ with positive coefficients is, at the time being, an open question of great interest, for instance, in the numerical integration of non-reversible systems.

References

- [1] S. Blanes, F. Casas and J. Ros, Symplectic integrators with processing: a general study, *SIAM J. Sci. Comput.* **21** (1999), 711–727.
- [2] S. Blanes, F. Casas, and A. Murua, ‘On the numerical integration of ordinary differential equations by processed methods’. Preprint available at: <http://www.focm.net/gi/gips/>
- [3] P.-V. Koseleff, *Formal Calculus for Lie Methods in Hamiltonian Mechanics*, Ph.D. thesis, Lawrence Berkeley Laboratory LBID-2030, Tech. report UC-405, Berkeley, CA, 1994.
- [4] D. Goldman and T.J. Kaper, ‘ N th-order operator splitting schemes and nonreversible systems’, *SIAM J. Numer. Anal.* **33** (1996), 349–367.
- [5] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer Ser. Comput. Math. 31, Springer-Verlag, 2002.
- [6] 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), 223–238.
- [7] R.I. McLachlan, ‘On the numerical integration of ordinary differential equations by symmetric composition methods’, *SIAM J. Sci. Comput.* **16** (1995), 151–168.
- [8] R.I. McLachlan and R. Quispel, ‘Splitting methods’, *Acta Numerica* **11** (2002), 341–434.
- [9] G. Rowlands, ‘A numerical algorithm for Hamiltonian systems’, *J. Comput. Phys.* **97** (1991), 235–239.
- [10] Q. Sheng, ‘Solving linear partial differential equations by exponential splitting’, *IMA J. Numer. Anal.* **9** (1989), 199–212.
- [11] M. Suzuki, ‘General theory of fractal path integrals with applications to many-body theories and statistical physics’, *J. Math. Phys.* **32** (1991), 400–407.