Comment on "Structure of positive decompositions of exponential operators"

Sergio Blanes*
Instituto de Matemática Multidisciplinar, Universitat Politécnica de Valencia, 46022-Valencia, Spain

Fernando Casas†
Departament de Matemàtiques, Universitat Jaume I, 12071-Castellón, Spain

(Received 27 October 2005; published 18 April 2006)

An elementary proof is shown on the necessary existence of negative coefficients in splitting methods of order $p \geqslant 3$.

DOI: 10.1103/PhysRevE.73.048701 PACS number(s): 02.70.Ss, 02.70.Ns, 95.10.Ce

The evolution of many physical systems can be modeled by a differential equation of the form $\dot{x}=(A+B)x$, where, in general, A and B are noncommuting operators. Writing the flow (i.e., the exact solution) formally as $x(t)=\exp[t(A+B)]\times[x(0)]$, it is well known that the composition

$$\varphi(\tau) = \exp(\tau A)\exp(\tau B) \tag{1}$$

provides a first-order accurate approximation, that is, $\varphi(\tau)=\exp[\tau(A+B)]+\mathcal{O}(\tau^2)$.

The order can be increased by including more exponentials in a time step τ. Thus, for instance, the scheme

$$\psi(\tau) = e^{a_1\tau A}e^{b_1\tau B}\ldots e^{a_m\tau A}e^{b_m\tau B} \tag{2}$$

has order p, i.e., $\psi(\tau)=\exp[\tau(A+B)]+\mathcal{O}(\tau^{p+1})$ if the coefficients a_i and b_i are appropriately chosen. These are examples of *splitting methods* [1], widely used in molecular dynamics, particle accelerators, statistical mechanics, and particularly in the context of *geometric integration* [2], since they preserve structural features of the exact solution. Examples of such features include symplecticity, volume preservation, unitarity, etc. [3].

It has been known for a long time that splitting methods of order $p \geqslant 3$ necessarily contain some negative coefficients a_i, b_i [4,5]. In other words, the methods always involve stepping backwards in time. This constitutes a problem when the differential equation is defined in a semigroup, as occurs, for instance, in time-irreversible systems.

Actually, it has been shown that at least one of the a_i *and* also one of the b_i coefficients have to be strictly negative [6]. The proofs published in [4–6] are based on the fact that a scheme of the form (2) with m any finite positive integer and all the coefficients a_i, b_i being positive cannot satisfy the order conditions up to order 3, which are explicitly

$$\text{Order 1:}\quad \sum_{i=1}^{m} a_i = 1; \quad \sum_{i=1}^{m} b_i = 1, \tag{3}$$

$$\text{Order 2:}\quad \sum_{i=1}^{m} b_i\left(\sum_{j=1}^{i} a_j\right) = \frac{1}{2}, \tag{4}$$

$$\text{Order 3:}\quad \sum_{i=1}^{m-1} b_i\left(\sum_{j=i+1}^{m} a_j\right)^2 = \frac{1}{3}; \tag{5}$$

$$\sum_{i=1}^{m} a_i\left(\sum_{j=i}^{m} b_j\right)^2 = \frac{1}{3}. \tag{6}$$

This impossibility is linked in [5] to a geometrical construction: when all the coefficients $a_i>0$ (respectively, $b_i>0$) are fixed, the equations at order 3 represent a hypersphere and a hyperplane which cannot intersect if all the coefficients b_i (respectively, a_i) are positive. The procedure is rather technical and nontrivial indeed.

In Sec. II of [7], Chin has gone one step further by providing an alternative, constructive proof of this property, which is claimed to be simpler and more illuminating than the original in [4–6]. Assuming that Eq. (5) is satisfied and that all the coefficients a_i are positive, he explicitly obtains an estimate of the amount by which Eq. (6) (the second condition at order 3) is not fulfilled. More specifically, it is shown that

$$2e_{VTV} \equiv \frac{1}{3} - \sum_{i=1}^{m} a_i\left(\sum_{j=i}^{m} b_j\right)^2 \leqslant -\frac{\delta g}{12(1-\delta g)}, \tag{7}$$

with $\delta g \equiv \Sigma_{i=1}^{m} a_i^3$. Furthermore, the coefficients a_i and b_i in (2) are related through [Eq. (2.27) in [7]]

$$b_i = \frac{a_i + a_{i+1}}{2(1-\delta g)}. \tag{8}$$

Thus, he concludes, if one insists that e_{VTV} be zero, then δg can be zero only if at least one a_i is negative such that (a_i+a_{i+1}) or (a_i+a_{i-1}) remains negative. Equation (8) then implies that its adjacent values of b_i or b_{i-1} must also be negative. In particular, the result first proved by Goldman and Kaper in [6] on the existence of negative b_i follows readily.

A close examination of the analysis carried out in Sec. II of [7] shows, however, that Eq. (8) has been deduced under the assumption that all a_i are positive and by requiring, in addition, that the error coefficient e_{VTV} be a stationary func-

*Electronic address: serblaza@imm.upv.es
†Electronic address: Fernando.Casas@mat.uji.es

tion of the $\{b_j\}$ coefficients. In consequence, the above conclusion on the existence of adjacent negative values of b_i or b_{i-1} only applies in this more restrictive setting. As a matter of fact, if e_{VTV} is no longer a stationary function of $\{b_j\}$ then it is not difficult to design schemes where the above pattern does not take place. For instance, the sixth-order Runge-Kutta-Nyström method 14 designed by Okunbor and Skeel [8] [which can be reformulated also as (2)] belongs to this class.

In order to establish that at least one b_i also has to be negative, as in the proof of Goldman and Kaper [6], the complete reasoning has to be carried out again, this time starting from the hypothesis that all b_i coefficients are positive. In this sense, the proof given by Chin is only slightly simpler than that contained in [4–6].

It turns out, however, that a truly elementary proof of this property can be easily constructed by adopting a different approach [9]. It follows readily from the close connection existing between the splitting method (2) and the composition of the first-order method $\varphi(\tau)=e^{\tau A}e^{\tau B}$ with its adjoint $\varphi^*(\tau)=e^{\tau B}e^{\tau A}$ with different time steps [10], i.e.,

$$\psi(\tau) = \varphi^*(\beta_0\tau)\varphi(\alpha_1\tau)\varphi^*(\beta_1\tau)\varphi(\alpha_2\tau)\cdots \times\varphi^*(\beta_{m-1}\tau)\varphi(\alpha_m\tau)\varphi^*(\beta_m\tau). \quad (9)$$

If we put $\beta_0=\beta_m=0$ in (9) we recover the scheme (2) as soon as

$$a_i = \alpha_i + \beta_{i-1}, \quad b_i = \alpha_i + \beta_i, \quad i = 1, \ldots, m. \quad (10)$$

Then

$$\sum_{i=1}^{m} a_i = \beta_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) = \sum_{i=1}^{m} b_i. \quad (11)$$

In consequence, composition (2) can be rewritten as (9) only if (11) holds. When the Baker-Campbell-Hausdorff formula [2] is applied to (9) one gets

$$\psi(\tau) = \exp(\tau f_{1,1}X_1 + \tau^2 f_{2,1}X_2 + \tau^3\{f_{3,1}X_3 + f_{3,2}[X_1,X_2]\} + \mathcal{O}(\tau^4)),$$

with $X_1=A+B$, $X_2=\frac{1}{2}[A,B]$, $X_3=\frac{1}{12}([B,[A,B]]+[A,[A,B]])$, and, in particular,

$$f_{1,1} = \beta_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i),$$

$$f_{3,1} = \beta_0^3 + \sum_{i=1}^{m} (\alpha_i^3 + \beta_i^3).$$

It is clear that a necessary condition to be satisfied by a method of order $p\geqslant 3$ is $f_{3,1}=0$. But this is obviously equivalent to

$$\sum_{i=1}^{m} (\alpha_i^3 + \beta_i^3) = 0 \quad (12)$$

since $\beta_0=0$. Now the proof reads as follows. For all $x,y\in\mathbb{R}$, if $x^3+y^3<0$ then $x+y<0$. Therefore, in the sum of (12) there must exist some $i\in\{1,\ldots,m\}$ such that

$$\alpha_i^3 + \beta_i^3 < 0 \quad \text{and thus} \quad \alpha_i + \beta_i = b_i < 0.$$

But one can also write

$$f_{3,1} = \sum_{i=1}^{m} (\alpha_i^3 + \beta_{i-1}^3) + \beta_m^3$$

just by grouping terms in a different way. Then, since $\beta_m=0$, the same order condition $f_{3,1}=0$ leads to

$$\sum_{i=1}^{m} (\alpha_i^3 + \beta_{i-1}^3) = 0. \quad (13)$$

Given that $\alpha_i^3+\beta_{i-1}^3$ from (12), the regrouped sum (13) must remain zero also for some $\alpha_j^3+\beta_{j-1}^3<0$ and thus finally

$$\alpha_j + \beta_{j-1} = a_j < 0.$$

In consequence, at least one a_i *and* one b_i coefficient have to be negative in *any* splitting method of order $p\geqslant 3$. Observe that the only requirement in our proof is just the necessary condition $f_{31}=0$ at order 3. The remaining order conditions do not alter this basic restriction.

One might argue that this simple demonstration only establishes the necessity of negative coefficients a_i and b_i in the operator splitting (2), but that it does not give any hint about their distribution in the composition, to what extent the order conditions are not satisfied and what is the specific error term which cannot be made to vanish when all the coefficients are positive.

Concerning the first aspect, such an analysis has been done in [9] by using, essentially, the same techniques and a slightly more involved analysis. In particular, a discussion is provided which explains why it is much frequent that if $a_i<0$ then its adjacent coefficients b_i or b_{i-1} are, in fact, negative in a given method.

With respect to the other arguments, the analysis carried out in [7] sheds light precisely on those points, allowing us to consider, in addition, special higher-order methods with positive coefficients by including nested commutators of A and B in the composition.

In any case, the straightforward proof we provide here has the additional advantage of identifying clearly the origin of the problem: the equation $f_{31}=0$ can be satisfied only if at least one a_i and one b_i are negative. According to this conclusion, any splitting method (2) verifying the order condition $f_{31}=0$ has necessarily some a_i and also some b_i coefficients being negative, even if the composition does not satisfy, say, the consistency requirements (3).

ACKNOWLEDGEMENT

This work has been supported by Ministerio de Educación y Ciencia (Spain) under Project No. MTM2004-00535 (also by the ERDF of the European Union). We also thank Professor S.A. Chin for his comments and remarks.

[1] R. I. McLachlan and R. Quispel, Acta Numerica **11**, 341 (2002).
[2] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations* (Springer, Berlin, 2002).
[3] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics* (Cambridge University Press, Cambridge, 2004).
[4] Q. Sheng, IMA J. Numer. Anal. **9**, 199 (1989).
[5] M. Suzuki, J. Math. Phys. **32**, 400 (1991).
[6] D. Goldman and T. J. Kaper, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **33**, 349 (1996).
[7] S. A. Chin, Phys. Rev. E **71**, 016703 (2005).
[8] D. I. Okunbor and R. D. Skeel, J. Comput. Appl. Math. **51**, 375 (1994).
[9] S. Blanes and F. Casas, Appl. Numer. Math. **54**, 23 (2005).
[10] R. I. McLachlan, SIAM (Soc. Ind. Appl. Math) J. Sci. Comput. **16**, 151 (1995).